The when and how of leaving school:
The policy implications of new evidence on secondary schooling in South Africa

MARTIN GUSTAFSSON

Stellenbosch Economic Working Papers: 09/11

KEYWORDS: HUMAN CAPITAL, UNEMPLOYMENT, EARNINGS FUNCTION, SOUTH AFRICA, SECONDARY SCHOOLS, EXAMINATIONS, EDUCATION POLICY
JEL: E24, I28, J31

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

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

The when and how of leaving school:
The policy implications of new evidence on secondary schooling in South Africa[1]

MARTIN GUSTAFSSON

ABSTRACT

South African and international household and education datasets are analysed to characterise patterns of dropping out, grade repetition, academic under-performance and under-preparedness for post-school life in South African secondary schools. A number of measurement error problems are moreover discussed and in some cases remedied. The proportion of South African youths entering upper secondary schooling is above the trend found in comparable middle income countries, the proportion entering the last grade (Grade 12) is about average, but the proportion successfully completing secondary schooling (40%) is below average. The data suggest improving quality should be a greater planning priority than increasing enrolments. A what-if subject choice analysis using examination data moreover suggests that successful completion could be greatly enhanced by guiding students to more appropriate subject choices, possibly through a more standardised set of assessments in Grade 9. Any attempt to reduce dropping out must pay close attention to financial constraints experienced by students with respect to relatively low-cost inputs such as books. Teenage pregnancies must be reduced as these explain half of female dropping out. The quality problem in schools underlined by the fact that income returns and test score gains associated with each additional year of secondary schooling are well below those associated with a year of post-school education.

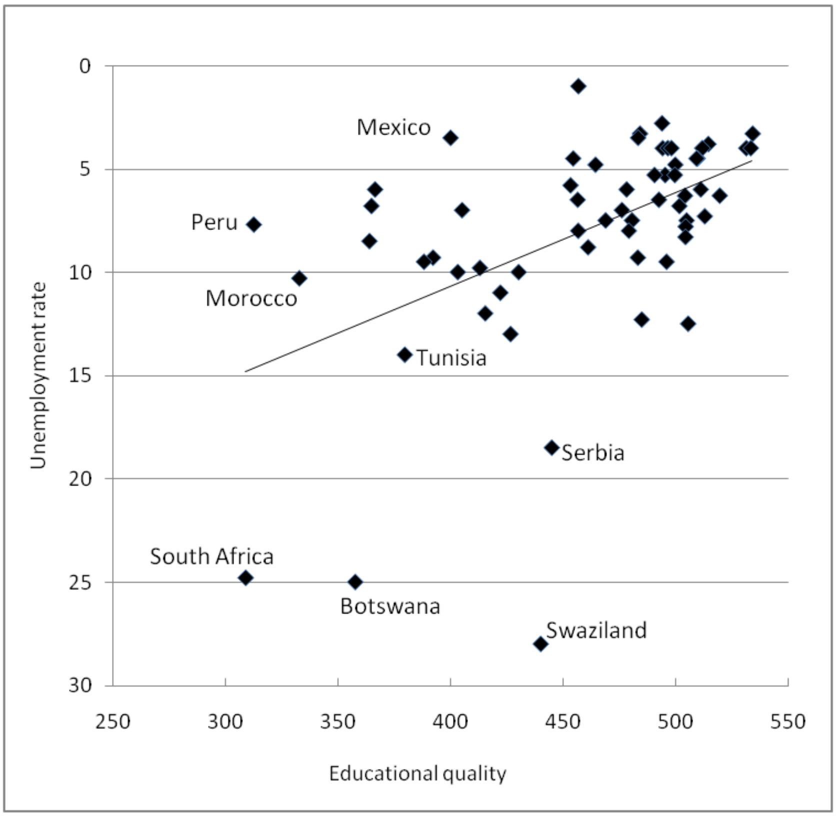
Keywords: Human capital, Unemployment, Earnings function, South Africa, Secondary schools, Examinations, Education policy
JEL codes: E24, I28, J31

[1] This document was produced within the Social Policy Research Group in the Department of Economics at Stellenbosch University with the financial assistance of the PSPPD (Programme to Support Pro-Poor Policy Development in South Africa), a partnership programme between The Presidency, Republic of South Africa, and the European Union (EU). The contents do not reflect the position of The Presidency or the EU.

1 Introduction

If one compares the development indicators of South Africa to those of other middle income countries, it is striking how poorly South Africa (and Botswana) perform against two key indicators: learning outcomes in schools and unemployment. This is illustrated in the following graph, which for a measure of educational quality uses normalised scores based on a variety of international testing programmes focussing on primary and secondary schools. A very similar picture would have emerged had scores from just the secondary schooling level been used.

Figure 1: Educational quality and unemployment[2]



Sources: World dataBank; Hanushek and Woessman, 2009.

The relationship between secondary schooling (and to some extent education generally) and the post-school opportunities of youths, in particular access to employment, is a central concern in this paper. Specifically, the paper makes use of newly available household data to attempt to uncover new facts and patterns, and in some cases confirm existing evidence. Policy solutions relating to the education sector, and in particular secondary schooling, are discussed. There is no shortage of ideas in the public discourse in South Africa on how to tackle the critical problems of youth unemployment. However, arriving at firm policy commitments is often made difficult by insufficient evidence and insufficient analysis of the full range of consequences, intended and unintended, of taking particular policy paths. Whilst the analysis in this paper can potentially inform a wide range of policy debates, it is only education sector policies that are explicitly discussed.

A key data source for the paper is the National Income Dynamics Study (NIDS) 2008 dataset (SALDRU, 2009). This dataset collected data from a nationally representative sample of around 7,300 households (with around 28,000 individuals). Whilst a small sample, the

[2] The correlation coefficient for the two variables in the graph is 0.50.

advantage with NIDS is that it includes questions about issues such as education, employment and the general 'state of mind' of South Africa's youths not included in other surveys. Another key data source in this paper is the 2009 General Household Survey (GHS) dataset, which includes a number of new and interesting variables not included in earlier runs of the GHS. These two datasets comprise the core data source, but in order to verify patterns, provide historical trends and explore cross-country comparisons, a number of other data sources are used too.

In section 2 four policy imperatives which form a backdrop to the data analysis are discussed. Section 3 provides a descriptive analysis of youth engagement in education and the labour market by age, of grade attainment and successful secondary school completion and of gross enrolments. Section 4 positions South Africa's attainment and enrolment figures in a global context.

Section 5 focuses on the reasons why learners leave school before completing Grade 12.

In section 6 the attention turns to learning outcomes and the relationship of this to post-school phenomena, in particular labour market earnings and employment. Grade repetition is also analysed. The 2009 Grade 12 examinations database is examined mainly in order to assess the appropriateness of the subject choices made by learners.

Section 7 offers a conclusion with a special focus on policy implications.

2 The menu of secondary school policy options

Secondary school policy solutions are of course just a part of the broader package of policies one requires to tackle youth unemployment. Yet existing evidence suggests that in South Africa education policy reform should play a particularly important role given that youth unemployment co-exists with an under-supply of skills in certain areas of the labour market. Below, four strategies for improving the readiness of youths for life after school are described. At least three of them can be described as high level strategies. This paper does not deal in any depth with important education policy questions at a more operational level, partly because the paper depends largely on household data and not data from the education system itself. Is there a need to revisit the higher level policy questions with regard to secondary schooling? Arguably there is. Though substantial analysis and debate have occurred, there has not been enough of it and there is still considerable disagreement in the policy debates not only around strategic directions, but sometimes around the basic facts of secondary schooling.

Improving the quality of basic learning outcomes across the board

Improving the quality of basic education offered at the primary and secondary levels of school, with respect to general levels of literacy, numeracy and life skills, prepares youths better for the challenges of post-school education and the world of work. The lack of preparedness for post-school life of a large proportion of South Africa's youths is not just strongly reflected in data such as that of TIMSS and the Grade 12 examinations (Van der Berg, 2007), but is also a problem that is often recognised by employers and universities. The term 'skills crisis' is frequently used, though whether this an appropriate term might be debatable (Kraak, 2008: 22). Solutions to this problem are often sought in the universities and vocational training institutions. Arguably, the role played by insufficient learning outcomes in schools in areas such as literacy and numeracy as a cause behind the 'skills crisis' does not enjoy the attention it should. If there is under-performance in schools, then not only will an insufficient number of people be qualified to enter university streams that should grow, those who do enter-post school studies are more likely to under-perform, which in turn results in under-performance in the economy as a whole.

The pervasive influence of the level of learning outcomes for a country is starkly illustrated in the cross-country analysis of economists such as Hanushek and Woessman (2009). It has been demonstrated that differences in the quality of learning in schools explain, more than any other development indicator, why certain countries perform better economically than others. Specifically, a country that succeeds in improving its learning outcomes from the typical level found in middle income countries such as Mexico, to the level found in rich OECD countries, experiences 2 additional percentage points of economic growth. More generally, the evidence suggests that if a country wants to develop and rid itself of poverty, and it does not enjoy an exceptional endowment of scarce natural resources such as oil, then the logical action to take is to improve the learning outcomes in one's schools[3].

The need to undo the structural and public expenditure legacy of apartheid in the education system meant that the policy focus on learning outcomes tended to be crowded out of the policy agenda for many years. The situation has changed, however, and policy commitments towards improving learning outcomes in schools is now strong. This can be seen, for instance, in the 2009 Medium Term Policy Statement (MTSF) of government, where the links between quality education and economic growth and development are recognised. Improving South Africa's performance within cross-country school testing programmes is explicitly mentioned.

Increasing enrolments at the secondary level

The importance attached to obtaining one's 'Matric', or one's Grade 12 National Senior Certificate (NSC), is deeply rooted in the South African psyche. Not obtaining this qualification is often associated with failure in a general sense. It is thus not surprising that getting everyone to successfully complete secondary schooling, in other words Grade 12, is often put forward as a national development goal. Achieving this would almost inevitably require a substantial expansion in the supply of teachers and school infrastructure, though in the long run one could free up spaces in schools through the reduction of repeater rates. Increasing the output of secondary schools in terms of the number of learners might indeed improve the availability of skills in the labour market, especially if the current distribution of performance does not change much. If 20% of Grade 12 graduates can consistently be considered outstanding, then 20% of a million learners is better than 20% of half a million learners purely from the perspective of the supply of skilled youths. However, constraints relating to the home disadvantages of learners and the supply of teachers would make it extremely difficult to maintain anything like a constant distribution of performance. A separate argument in favour of increasing secondary level enrolments is that this reduces youth unemployment insofar as youths are kept back from entering the labour market.

The argument that secondary enrolments ought to increase often assumes an *a priori* status because this need is seen as so obvious (see for instance DoE (2008: xxiv) and Schindler (2008: 251)). Yet there is evidence that not just in South Africa, but in the Southern Africa region in general, there has been an imbalance between access to secondary schooling and learning outcomes. In a sense, access has been prioritised at the expense of maintaining sufficient levels of learning outcomes (Crouch and Vinjevold, 2006). Much of the international literature promotes expanding secondary schooling, especially if unmet demand for this from the youth is evident (see for instance World Bank (2005)). However, often such advice is aimed at countries with much lower levels of secondary level enrolment than South Africa.

On the topic of grade repetition, which obviously influences enrolment numbers, an important finding is that of Lam, Ardington and Leibbrandt (2008: 22), who conclude that at least in the Cape Town area poor assessment techniques in historically African schools result in a

[3] Gustafsson, Van der Berg, Shepherd and Burger (2010) explore the education-growth relationship from a South African perspective.

situation where one's actual abilities are three times less likely to predict whether one passes one's grade in these schools than in other schools. In other words, learners in historically African schools being made to repeat unnecessarily and may be unjustly discouraged from continuing with their secondary school studies.

The emphasis in the 2009 MTSF is on 'increasing enrolment rates to 95 per cent by 2014' in secondary schooling and 'ensuring that as many young people as possible are able to access and complete secondary education'. What the 95% target refers to is not made explicit and depending on one's interpretation South Africa has either almost reached this target (if one is referring to the gross enrolment ratio, for instance) or is still far from achieving this (if one considers the completion ratio, for instance).

The secondary enrolments question is necessarily a complex one that requires consideration of demand amongst youths for schooling, of capacity to expand the supply of this service, and of trade-offs between the quantity and quality of schooling, between general schooling and vocational education and training outside of schools, and between grade repetition and increased access.

Better 'signalling systems' for secondary school graduates

Employers and post-school education institutions use a person's qualifications to decide whom to employ in what position, or whom to accept in what courses. One's qualification thus functions as one's 'currency' representing past educational achievements. The better the design of the system of education qualifications, the more efficient the transactions between youths and prospective employers or education institutions. Gustafsson and Bartlett (2008) argue that an important gap in the South African schooling system is the lack of a national qualification below Grade 12, resulting in a situation where over half of youths have no widely recognised qualification to demonstrate what they have achieved after (usually) more than ten years of schooling. The international literature on the role of school qualifications in the labour market is under-developed. What evidence exists points to the fact that schooling is worth more in the labour market when it comes with a qualification (Dearden, 1999). This is to be expected, given the labour market signalling role of qualifications. Government's position on a qualification below Grade 12 has been ambivalent. A commitment towards the Grade 9 General Education Certificate (GEC) was expressed as early as 1995, in Education White Paper 1. Since then, it has moved on and off the education policy agenda. Arguments against the GEC have included its cost and suspicions that it could undermine efforts to get more learners, in particular learners from disadvantaged backgrounds, to obtain the Grade 12 NSC. The proposed introduction of standardised and (to some extent) externally monitored assessments in Grade 9 in 2011, following the introduction of such assessments in Grades 3 and 6 in 2008, may provide new arguments for and against a Grade 9 qualification (see the President's 2010 State of the Nation Address).

A greater emphasis on vocational education and training at the secondary level

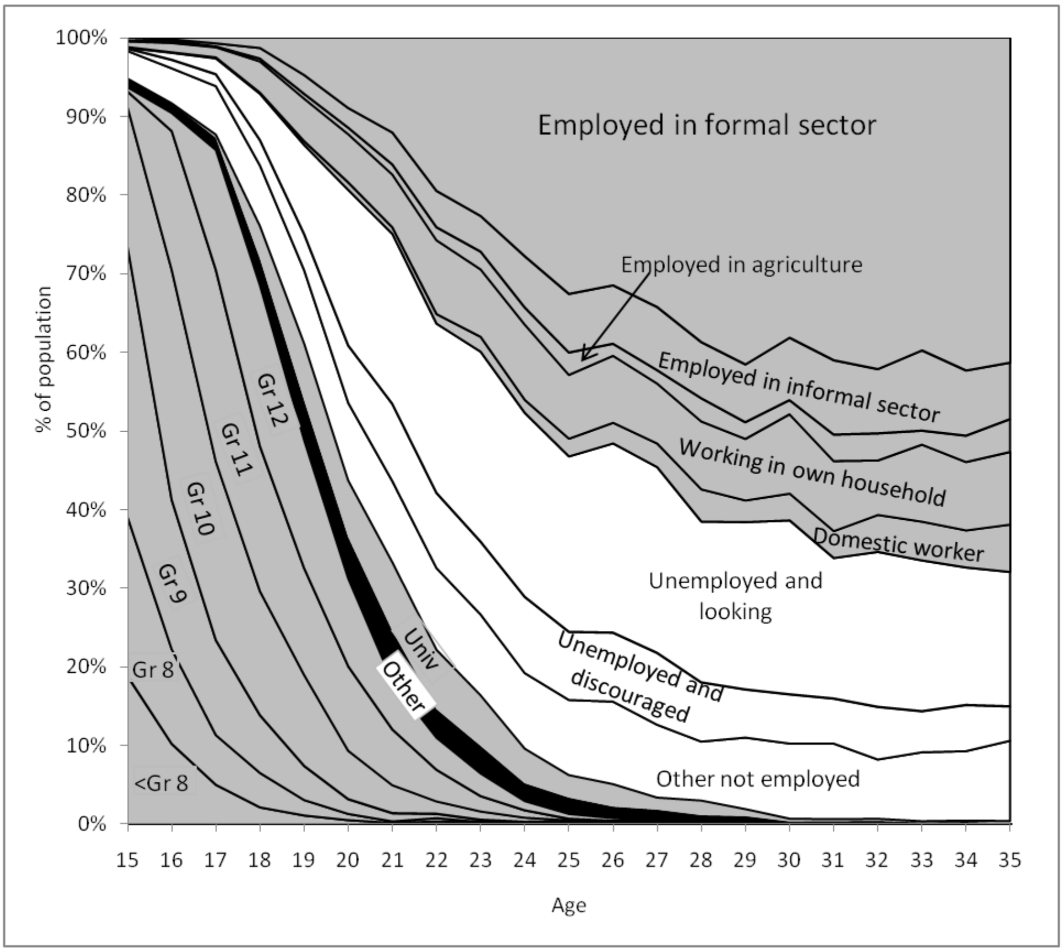
The so-called skills crisis in the labour market often leads to calls for more vocational training for youths. In many ways, such calls compete with the very strong attachment, in the public and amongst certain opinion makers, to the Grade 12 Matric. The benefits of more investments in vocational training, relative to the cost of such training (which is mostly higher than the cost of general schooling), is a matter over which there is still not much agreement, though reviews of the literature by analysts such as Ziderman (1997) suggest that well designed training programmes do indeed present a cost effective option for countries such as South Africa. This paper is somewhat limited in its ability to deal with this policy question due to data limitations. It is worth noting that the school curriculum allows for considerable vocational training within secondary schools, as opposed to FET colleges. Of the 24 non-language subjects of the Grades 10 to 12 curriculum introduced in 2005, as many as 16 can be considered vocationally oriented.

3 The basic youth education and employment statistics

3.1 Youth economic activities by age

Figure 2 provides a picture of the economic roles played by youths, by age, in 2009. The paper follows the common South African practice of considering those aged 15 to 35 as youths. A table with the values underlying the graph is provided in Appendix A. Overall, 30% of youths aged 15 to 35 are, in a sense, preparing themselves for their economic role insofar as they are currently studying, 39% are contributing to the economy of the country insofar as they are working, and the remaining 31% are in some way or another unemployed (this is using a broad definition of unemployment, as explained below). This 31%, represented by the white segments in Figure 2, translates into around 5.7 million youths.

Figure 2: The activities of youths by age



Sources: Quarterly Labour Force Survey 2009 (third quarter); General Household Survey 2009.
Note: The QLFS was used for all breakdowns except for the breakdown of the student population by grade and education institution (for the latter the GHS was used).

The breakdown of school learners by grade and by age has been greatly facilitated by the introduction in the 2009 General Household Survey of a question asking respondents which school grade they are currently enrolled in (previously the only grade-specific question was what one’s highest completed school grade was). What stands out in Figure 2 is how many over-aged learners there are in the secondary grades. If one considers anyone who has repeated a grade as being over-aged, and assumes that everyone enters Grade 1 in the year they turn 7 and thus that someone aged 7 should be at least in Grade 1 (and that someone aged 14 should be at least in Grade 8, and so on), then 60% of the Grades 8 to 12 learners captured in Figure 2 are over-aged. This situation has in fact been improving over the years. It is difficult to obtain earlier estimates of the 60% statistic mentioned here due to limitations in the availability of data, but one statistic that can be tracked across time very easily is school learners, of any grade, aged 19 and above as a percentage of all enrolled learners aged 15 and above. This statistic has moved from 36% in 1999, to 28% in 2004, to 20% in 2009 (according to the GHS). Ironically, this improvement in the education system can result in problems in the labour market and in higher education. If the age of learners leaving school drops, then over some years there is a bulge in the number of youths leaving school, in other words youths who need jobs or places in post-school institutions. If this demand is not met, the percentage of unemployed youths can rise. Section 3.2 below argues that the shifts in enrolment were truly efficiency gains in the sense that there was less repetition resulting in the average age of learners leaving secondary school. What did *not* happen was more dropping out, in other words learners leaving school at a lower *grade* than before. In fact, the grade attainment of youths improved.

There are three white segments in Figure 2 representing different modes of unemployment. The 'Unemployed and looking' category of youths is a fairly straightforward one to define. Generally, Stats SA (2009c: xvi) considers anyone who is not employed but has actively sought work in the previous four weeks as belonging to this category. The category 'Unemployed and discouraged' is also based on a Stats SA category. People in this category have not sought work in the previous four weeks because they were certain there were no jobs available in the geographical vicinity, or in their area of specialisation. People in the category 'Other not employed' in Figure 2 would not be considered unemployed by Stats SA because they do not fulfil the formal requirements for this status (these requirements are based on international practice), yet it would be common to refer to such people as unemployed. For 31% of the 'Other not employed' poor health is given as the reason for not working. A further 10% simply say they have no desire to work. A few regard themselves as too young to work, whilst for the remainder the reason for neither studying nor working is not clear from the data. Of the approximately 3 million who say they are actively looking for a job, around 1.7 million have been looking for over a year. This helps to explain why many youths would be inclined to give up.

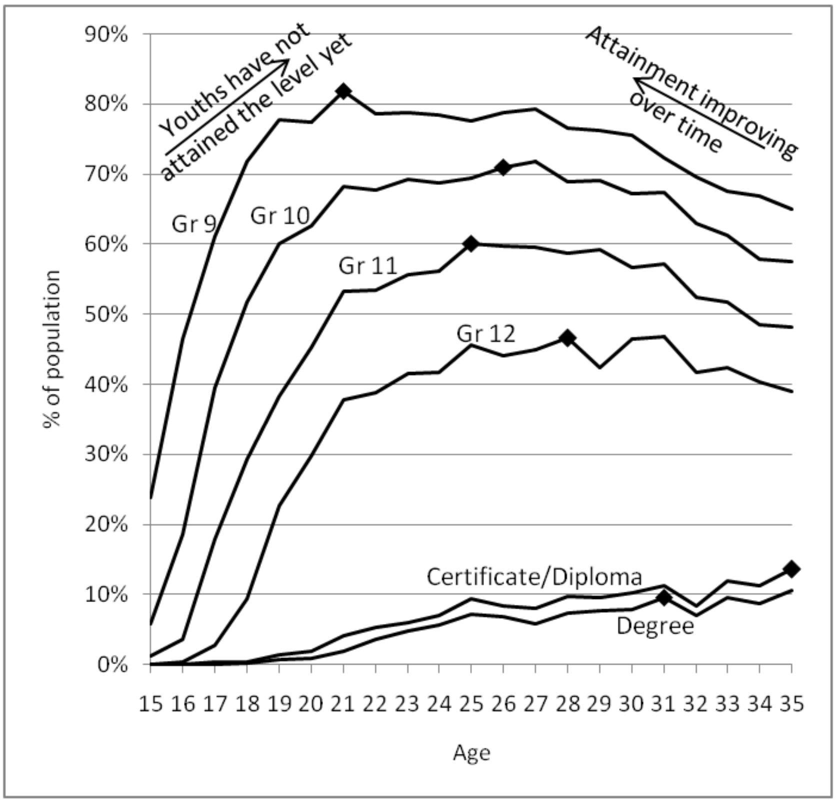
In Figure 2 those who worked in their own household, for instance as housewives or looking after children, were considered to be economically active. These individuals are, naturally, making an important social contribution, even if they are not paid for their work. Of the economically active youths, around 60% of them are employed in the non-agricultural formal sector. The Quarterly Labour Force Survey data used for Figure 2 distinguishes between the employed who are fully employed and those who are under-employed in the sense that they do not work full-time and would like to work more time. The extent of under-employment as defined by Stats SA is relatively low. Of the youths who are in formal employment, only 2% are considered to be under-employed. As one might expect, the figure is higher for those in informal employment, at around 9%.

If the Figure 2 patterns are broken down by gender important differences emerge. The percentage of males aged 15 to 35 enrolled in an education institution is higher than for females – 32% against 28%. As will be seen below, the gender differences when it comes to *attainment* of certain education levels are smaller, suggesting that the differences seen here have to do with different enrolment patterns by gender, in particular more grade repetition amongst males. The percentage of youths who are not economically active (the white segments in Figure 2) is not that different for the two genders – 30% for males and 32% for females. 27% of male and only 18% of female youths are employed in the formal sector. However, the total numbers of males and females who would be considered economically active in Figure 2 are not that different if one includes those working in their own household. This last category is 90% female.

3.2 Educational attainment of youths

The question is often asked what percentage of South Africa’s youth obtains the Matric, or Grade 12 certificate. Clearly the answer cannot be the widely publicised pass rate, as there are learners who do not even reach Grade 12. The Grade 12 curve in the following graph illustrates the percentage of youths of each age who have successfully completed Grade 12 or anything above Grade 12 (using 2004 data, which are compared to later data in the graph that follows). The curve is an inverted U, with the left-hand dip indicating how young learners had not reached Grade 12 yet and the right-hand dip indicating older youths who were less likely to complete Grade 12 because grade attainment has been improving over time. The peak of the curve is at age 28, suggesting that by this age everyone who was, in a sense, meant to attain Grade 12 had attained this grade. Put differently, if one takes into account all those who complete Grade 12 late, perhaps through adult education centres, then age 28 is the age beyond which one would not expect many more individuals to attain Grade 12. Age 28 for this statistic seems like an extraordinarily high age, which raises the question of whether the situation has improved since 2004, in other words whether youths have been attaining Grade 12 earlier in life. As will be seen in a subsequent graph, the situation did indeed improve beyond 2004. The main purpose of Figure 2 is to illustrate how one should answer the question of what percentage of youths obtain a Grade 12 certificate. The best answer is probably that represented by the peak, which corresponds to 47% of an age cohort. Similarly, we can say that 60% of youths had attained at least Grade 11 in 2004, 71% had attained at least Grade 10, and 82% had attained at least Grade 9.

Figure 3: Educational attainment in 2004[4]



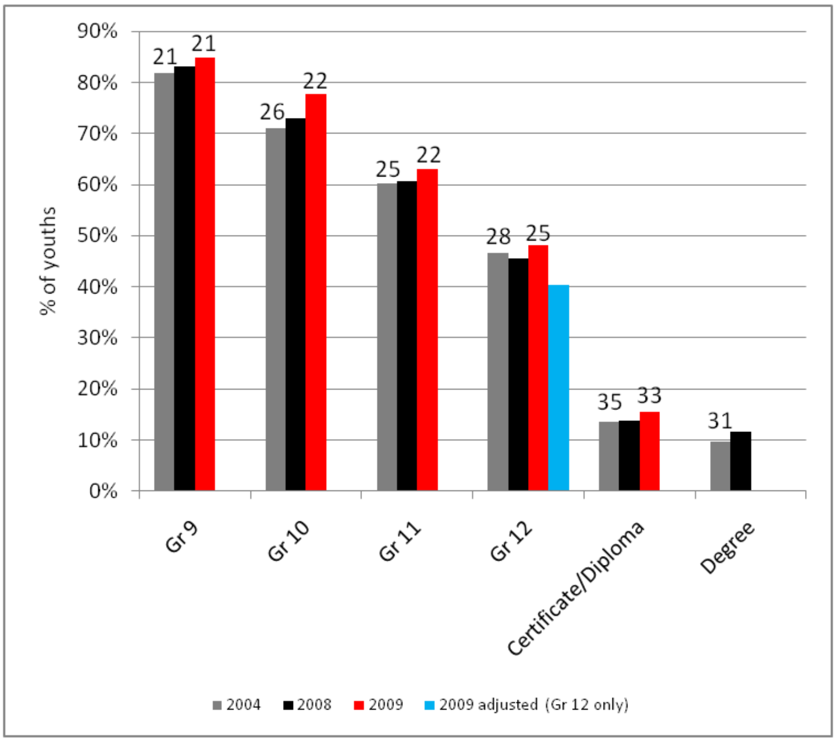
Source: General Household Survey 2004.

Figure 4 provides the 2004 values obtained from the previous graph, plus GHS values for 2008 and 2009 (the 2009 adjusted value for Grade 12 is discussed further down). There has been a slight improvement in attainment over the years in two respects. Firstly, more youths have completed Grades 9 to 11 (though a clear improvement is not evident for Grade 12). For The percentage completing Grade 9 successfully rose from 82% to 85%, whilst the Grade 10 improvement was from 71% to 78% and the Grade 11 improvement from 60% to 63%. Moreover, the age by which all youths had reached each level dropped, for instance from age 28 to 25 in the case of Grade 12. In terms of the previous graph, the peak of the curve shifted leftwards. These shifts occurred with respect to almost all types of educational attainment illustrated here (the category 'Degree' was left blank for 2009 as the value, at 4%, was so completely at variance with the historical trend, for instance as seen in 2008, suggesting a problem with these data in the 2009 dataset). Youths were thus completing slightly more years of schooling and were doing so at a slightly younger age. It is important to emphasise that youths were leaving at an earlier *age*, but not at an earlier *grade*. In analyses of unemployment trends it has been suggested that changes in school enrolment patterns were an important reason for a rise in the unemployment rate (see for instance Burger and Von Fintel, 2009). Youths entered the labour market earlier, causing a bulge in the number of job-seekers. Figure 4 does indeed support such a scenario. However, the youths entering the labour market were not *less* educated than before, they were in fact *more* educated. It is extremely likely that this trend is the result of decreases in the grade repetition rate over time, although as indicated in Appendix D, confirming this through an analysis of grade repetition itself is made difficult

[4] There is in fact some bias in the graph towards identifying a point on each curve that is as far leftwards as seemed justifiable. For instance, for Grade 12 the peak is in fact at age 31, with a value of 46.7%. However, the point furthest left that deviated from the 46.7% by no more than 1 percentage point was selected. This is how the age 28 point was selected. In the 2004, 2008 and 2009 figures discussed in this section, 1 percentage point was regarded as the justifiable cut-off.

due to data limitations. What Appendix D also demonstrates is that the phenomenon of grade attainment at specific grade levels at a progressively younger age is evident even when a large range of years than those in Figure 4 are studied.

Figure 4: Progress in educational attainment 2004-2009



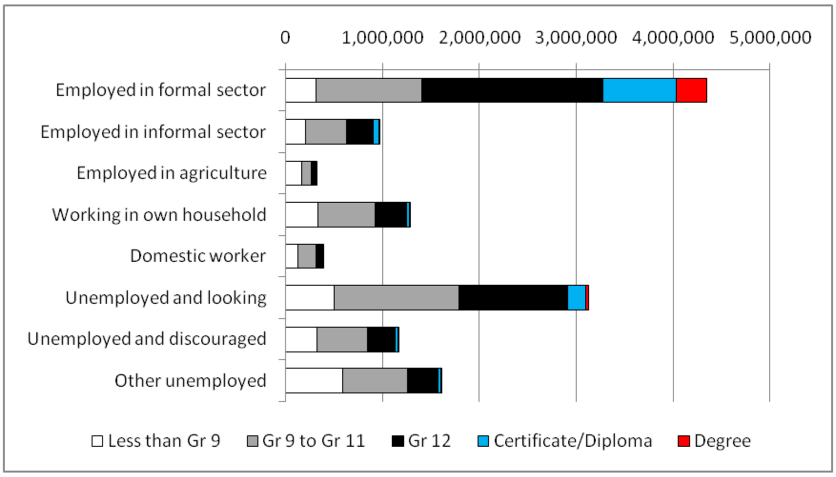
Source: General Household Survey 2004, 2008, 2009 for first three data series.
Note: The values at the tops of columns indicate the earliest age at which this level of enrolment is reached.

However, there is unfortunately a problem with the Grade 12 level of 48% in 2009 illustrated in Figure 4. In a way, it is logical to expect a reporting problem resulting in an over-estimate. Respondents in the GHS are asked what their highest successfully completed grade is. Respondents who attended Grade 12 but did not pass the examinations may give Grade 12 as a response, as opposed to Grade 11, either because they misunderstand the question or because they are embarrassed about not having obtained their Matric, which is a high status qualification. In Appendix B analysis of a variety of data sources suggests strongly that this indeed occurs and that the true percentage of youths obtaining a Matric is likely to be around 40%, with 39% corresponding to the public NSC examinations and just around 1% to the non-public IEB examinations (the 40% is reflected in the '2009 adjusted' column in the above graph). This adjustment is obviously important to bear mind when considering policy options and making cross-country comparisons.

What do the magnitudes from Figure 4 mean in terms of the holding of qualifications? In particular, what widely recognised qualifications do the 60% of youths who do not obtain a Matric hold? Analysis of the NSC results suggests that around 22% of youths who have left school have no Matric Certificate but do have an NSC statement of results indicating what their results were per subject (these results would reflect an overall fail in the examinations). Whilst not a qualification, this document is at least widely recognised proof that the holder at least reached Grade 12 and attempted the examinations. Conceivably, this could be of some value when the person seeks employment or admission into a non-school education and training institution. Only around 1% of youths hold no Matric but do hold some other non-school certificate or diploma issued by, for instance, an FET college. This leaves around 37% of youths with no widely recognised proof of their educational status. They would have report cards issued by their school, but these would contain results that are not standardised and therefore carry little currency in the labour market or in further education and training institutions.

If we super-impose grade attainment with the activities of youths as illustrated in Figure 2, we obtain the picture appearing below. More years of schooling does appear to assist youths to some degree in the labour market. For instance, having completed Grade 12 appears to improve one chances of being formally employed – those without at least Grade 12 comprise 65% of youths working in the informal sector but only 32% in the formal sector. Having a degree appears to almost guarantee a place in the formal sector. However, it is also striking how similar the compositions of the different bars are. Many youths (1.1 million) entered the formal sector having passed Grade 9 but not Grade 12 and, conversely, many youths with Grade 12 (also 1.1 million) are unemployed and looking. When considering policy solutions to deal with youth unemployment, these patterns are obviously crucial to keep in mind. Above all, the 2009 patterns seen in Figure 5 would seem to be at odds with the notion that giving youths more years of schooling (for instance getting more youths to complete Grade 12) translates into more employment in a simple or mechanistic fashion. The dynamics are clearly more complex than that.

Figure 5: Activities of youths against educational attainment in 2009



Source: Quarterly Labour Force Survey 2009 (third quarter).

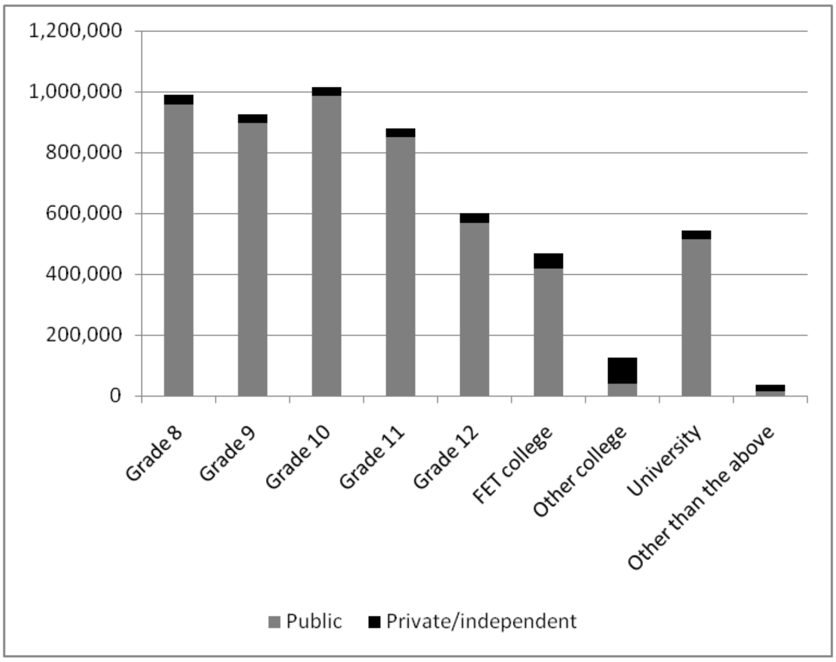
What the above graph does not reflect is the breakdown by occupation within each bar, which would explain some of the patterns we see. Moreover, income, and in particular strong positive wage effects associated with having Grade 12 (Keswell and Poswell, 2004), are beyond the scope of this paper. This is obviously a limitation, yet it in many ways employment is a more serious problem for youths than low wages, justifying to some degree an approach which focuses exclusively on whether one is employed or not.

3.3 The youth enrolment numbers

Due to grade repetition, enrolment figures are high relative to grade attainment. For example, whilst only 78% of youths successfully complete Grade 10 (see Figure 4), Grade 10 enrolments divided by the population aged 16 is 110%. This is mainly due to grade repetition, but also due to youths enrolling in Grade 10 but not successfully completing the grade.

Enrolment figures seem relatively straightforward for Grades 8 to 11. However, in the case of Grade 12, many different household surveys suggest that the official enrolment numbers are too low by about 100,000. The raw numbers for all grades differ when household data and official enrolment statistics based on school surveys are compared. This is due to the way data are weighted in the household surveys. This is not the issue here, however. The question raised here is based on a comparison of enrolments across grades *within the same data source.* The matter is discussed in Appendix C, where the conclusion is drawn that the available data do not make it possible to explain the Grade 12 discrepancies in a satisfactory way. In the graph that follows and in the paper as a whole it is assumed that official grade enrolment statistics are correct, even for Grade 12. It is very possible that this is indeed the case and that the discrepancies mentioned here are the result of problems with the household data collections.

Figure 6: Enrolment patterns of youths in 2009



Sources: DoE, 2009; General Household Survey 2009; CHE, 2009.
Note: In the four right-hand columns, values based on the GHS have been adjusted downwards by 5% to cater for the estimated over-weighting of observations (5% was arrived at through comparison of the official enrolment data and household data).

In schools, the pattern of an enrolment dip in Grade 9, relative to enrolments in the adjacent grades, is a pattern that is found in many years (see Appendix B). Reportedly, this is due to high repetition in Grade 8 related to adjustment problems of learners experiencing their first year in the secondary level (usually in a new school) and high repetition again in Grade 10 as learners grapple with the new curriculum phase starting in Grade 10 and involving, for instance, a few new learning areas. The pattern of a steady decline in enrolments from Grade 10 to Grade 12 is also a persistent one. Even if there was indeed an under-count of 100,000 Grade 12 learners, this pattern would still hold true.

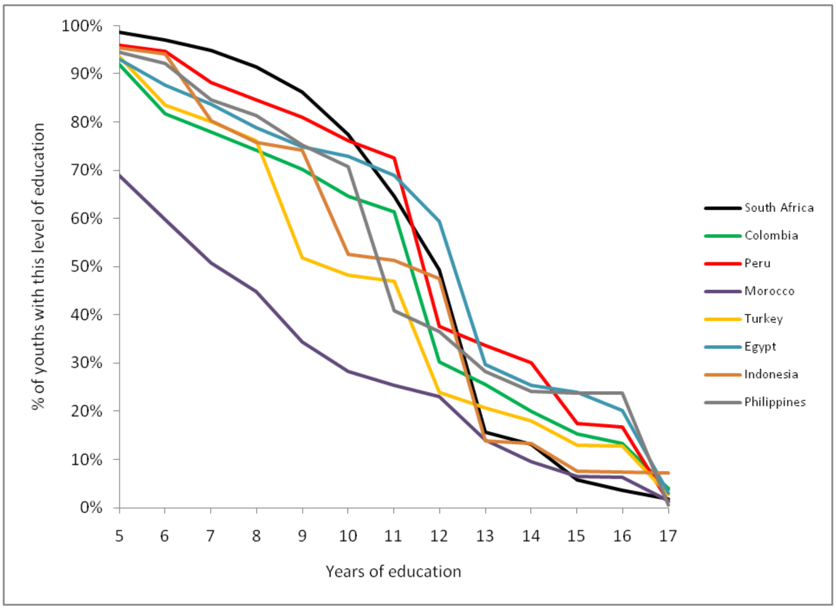
The largest post-school enrolment column in Figure 6 is the one referring to universities (this would include universities of technology). This is followed by FET college enrolments. Here enrolments in public colleges include many part-time students and are based on what the DoE has published (DoE, 2010: 23). If one uses only the 2009 GHS, then enrolments in public FET colleges drop from around 418,000 (the public segment of the relevant column in Figure 6) to around 115,000. The GHS counts only enrolment on the date of the survey and only if this enrolment lasted at least six months in the year. The private FET college enrolment value used in the graph uses the GHS definition only and this value would clearly be higher if all part-time enrolments were also counted (data on the latter for private colleges seem not to exist). The column 'other colleges' also uses just the GHS definition of enrolment. Enrolment at public universities is derived from the GHS (only youths aged 35 or younger were counted for this and all the columns in the graph)[5]. Enrolment at private universities is from a Council on Higher Education report (CHE, 2009).

4 South Africa's youth schooling statistics in a global context

Whilst cross-country comparisons of enrolment using the UNESCO Institute for Statistics (UIS) online database is straightforward, obtaining data for cross-country comparisons of enrolment or educational attainment by age is more challenging[6]. For the latter, use is made here of the standardised Demographic and Health Survey (DHS) and World dataBank datasets. For the period 2000 and later DHS data on seven middle income countries (other than South Africa) were obtainable.

Figure 7 illustrates the percentage of youths successfully completing different years of education, where these years are counted starting at Grade 1 (in other words pre-school is not included). Up to and including Grade 12, South Africa is near the best or in fact the best of the eight countries analysed. However, beyond Grade 12, in other words with respect to what would mostly be post-school education, South Africa, together with Morocco and Indonesia, display relatively poor educational attainment. Fewer than 10% of youths in these countries attain 15 years of education (this reflects, for instance, the completion of a three-year degree course in South Africa). In contrast, the figure is at least 15% in Colombia and Peru and 24% in Philippines and Egypt.

Figure 7: Cross-country comparison of educational attainment



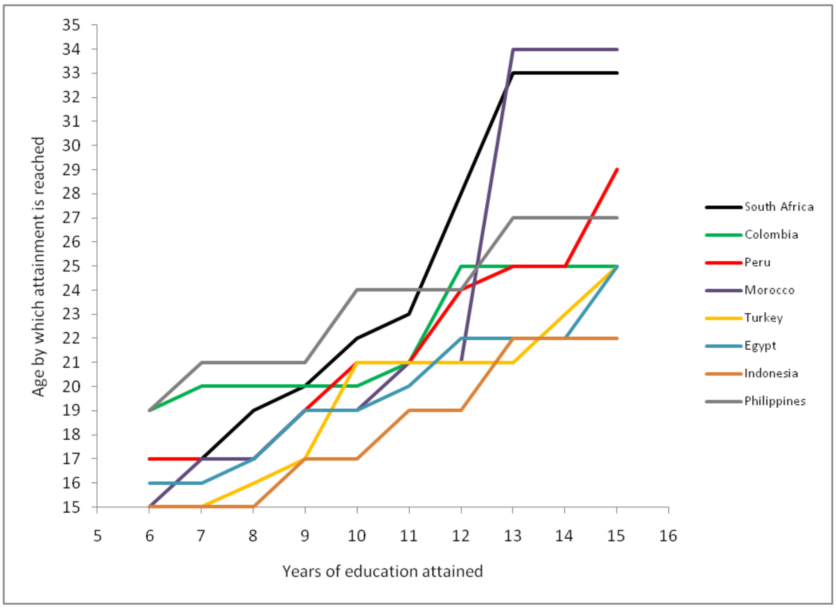
Sources: Quarterly Labour Force Survey 2009 (third quarter) for South Africa; Demographic Household Surveys for other countries.
Note: The DHS data are all from collections later than 2000.

[5] In the GHS a third of university enrolments amongst youths are reported as being in private institutions. This must largely be a matter of respondents considering the institution private, perhaps due to what they perceive as the high fees being charged, whilst the institution is in fact public.

[6] The World Bank has a facility at http://iresearch.worldbank.org/edattain/, which uses DHS data. However, this facility is better suited for examining primary than secondary schooling.

The next two graphs examine the ages by which those who will attain a particular level of education, get to attain that level. In other words, the measures introduced in Figure 3 for South Africa are compared internationally. In Figure 8 a 1% leeway is used. Thus, for instance, in South Africa 49% of youths get to obtain twelve years of education but the earliest age by which 48% of youths obtain this level of education is 28 years[7]. The age 28 value is thus plotted on the vertical axis for 12 years of education. The same method is used for all countries. In Figure 9 a wider leeway of 5% is used, which lowers many of the ages of attainment. The two graphs illustrate how the patterns seen in South Africa are not altogether unusual, even though they may appear counter-intuitive (for instance an age of attainment of 28 for just 12 years of education). Yet the South African curves in both graphs stand out as being particularly high, suggesting that grade repetition (and other factors such as long waiting periods between the completion of secondary schooling and entry into higher education) that delay the entry of youths into the labour market, and hence reduce lifetime returns to education investments, warrant special policy attention. At the same time, the graphs suggest that there are substantial obstacles in the way of very dramatic efficiency improvements. Many countries, for instance, do not succeed in achieving a situation where 12 years of schooling is generally completed by age 20.

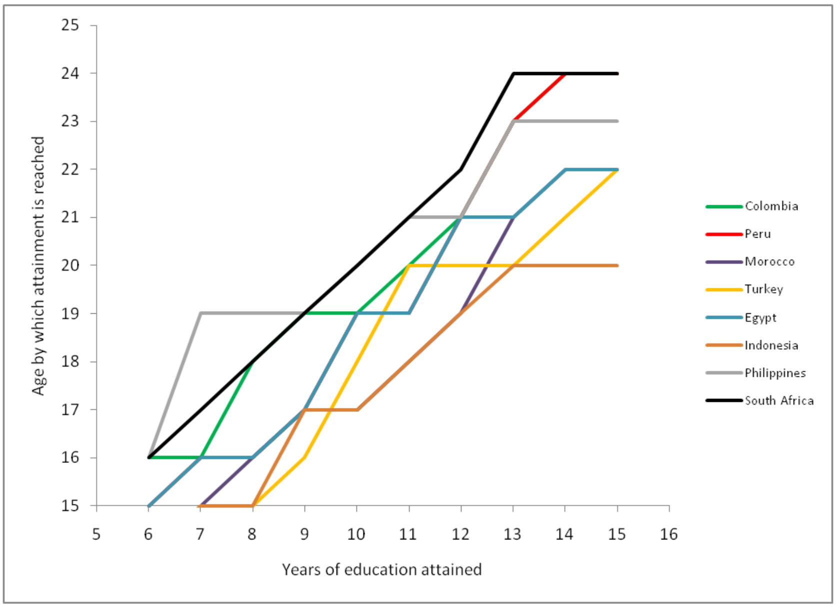
Figure 8: Cross-country comparison of age of attainment (1% leeway)



Sources: As for Figure 7.
Note: In this graph and the next one some smoothing of curves occurred to avoid clearly anomalous bumps.

[7] The reason why the slightly higher value of 49% appears here whilst 48% was indicated in relation to Figure 4 is mainly that the 49% value includes a few youths who do not have the Matric but have achieved twelve years of education through an FET college.

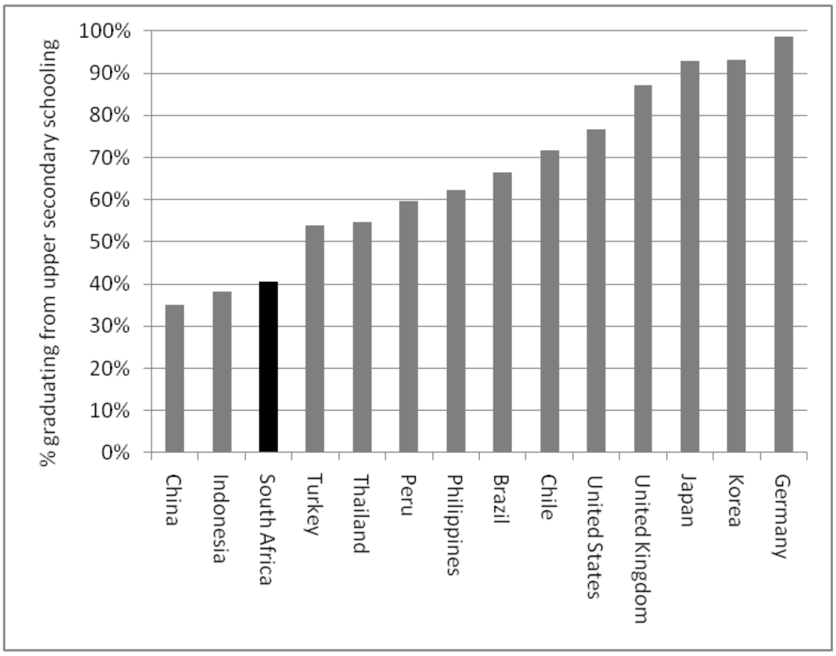
Figure 9: Cross-country comparison of age of attainment (5% leeway)



Sources: As for Figure 7.

The following graph is based not on household datasets (as was the case with the previous three graphs), but on official population and upper secondary school graduation figures published by the OECD. South Africa's position appears less favourable here than it did in Figure 7, partly because the lower and more accurate value of 40% is used here for South Africa (see discussion in section 3.2) and partly because a few other developing countries (such as Turkey) appear in a more favourable light here (compared to Figure 7) due to the use of a different data source. From a policy perspective, what seems especially noteworthy is that developed countries such as the United States and United Kingdom (which one can assume would report highly reliable data to the OECD) are not very close to achieving successful upper secondary graduation for all. In the USA, 23% of youths do not complete upper secondary schooling, whilst in the UK the figure is 13%. It is only Germany in the graph that is close to achieving a 100% graduation rate. Of course it is possible that, for instance, many of the 23% of youths in the United States who do not complete formal upper secondary schooling, complete vocational training at an equivalent level. What the data used for Figure 10 underline, above all, is that whilst aiming for higher upper secondary graduation rates in developing countries seems sensible, aiming to have everyone complete this kind of education is not necessary optimal.

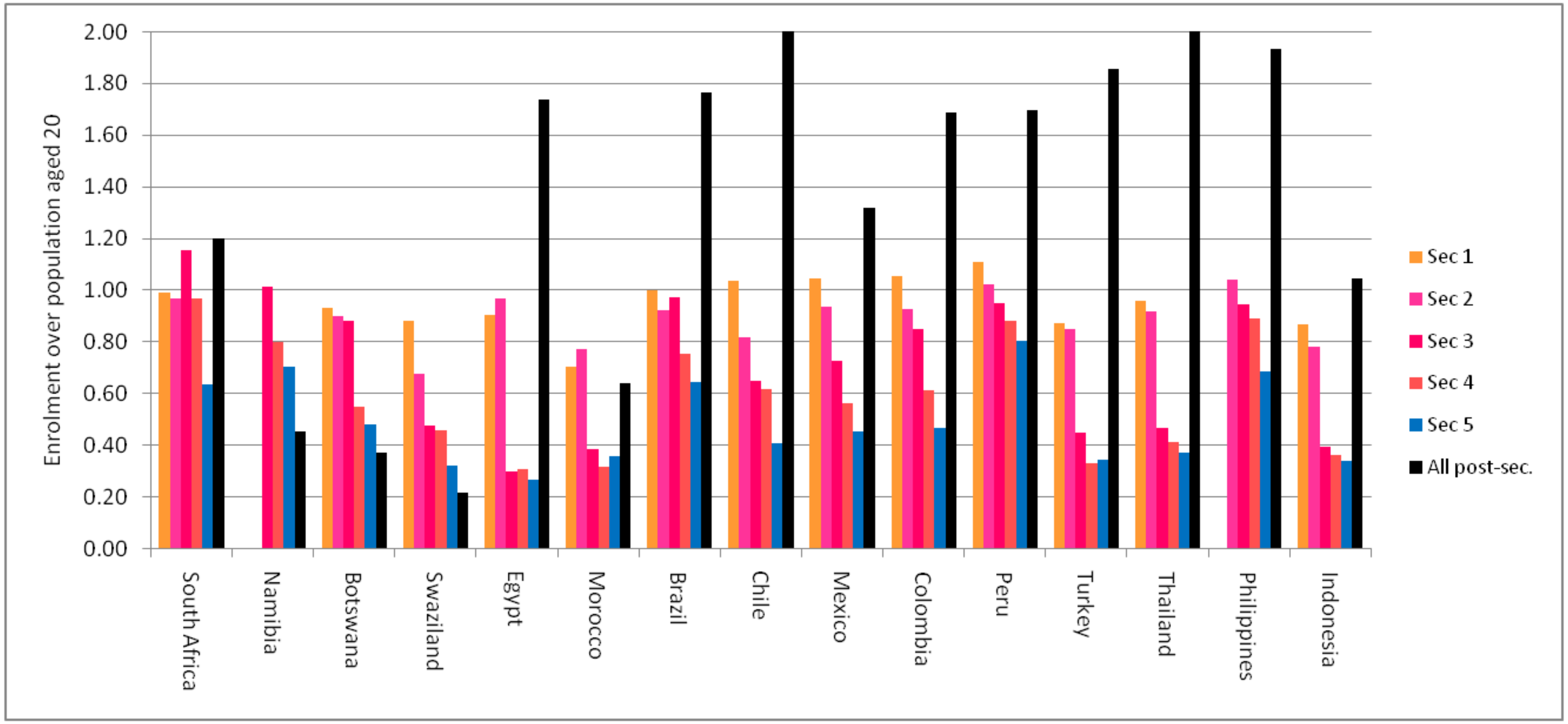
Figure 10: Cross-country comparison of recent secondary school graduation rates



Source: OECD.StatExtracts. For South Africa see discussion in section 3.2. Note: Developing countries other than South Africa have 2003-2004 averages, whilst developed countries have 2005-2007 averages.

Figure 11 illustrates secondary (Grades 8 to 12, even if more grades existed) and post-school enrolment values across 15 middle income countries. The post-school values are the sum of what UNESCO refers to as 'post-secondary non-tertiary' and 'tertiary'. UNESCO counts both full-time and part-time post-school students, but only those enrolled at a single point in time (the survey date). For this reason the GHS and not DoE (2010) was used as the data source for FET college enrolment. The graph shows that in Grades 8 to 11, enrolment is almost equal to the size of a population cohort in South Africa. In comparison to other countries, South Africa does exceptionally well in this regard. With regard to Grade 12 enrolment, South Africa is less exceptional, but nevertheless above the average – only three other countries have a better enrolment ratio in this grade. When it comes to post-school enrolment, however, South Africa's ratio is considerably lower than that in eight other countries (these eight countries all have enrolment levels that are at least 40% higher than the South African one), fairly similar to that in Morocco and Indonesia, and considerably better than that in the other Southern African countries. Indeed, there appears to be a pattern of low post-school enrolments in the Southern African region. It seems debatable whether South Africa succeeds in breaking this pattern. The UNESCO data thus confirm what was seen in Figure 7. If there is an under-enrolment problem in South Africa, and if one uses other countries of a similar level of economic development as one's yardstick, then the problem is not in secondary schooling but in education and training occurring after school.

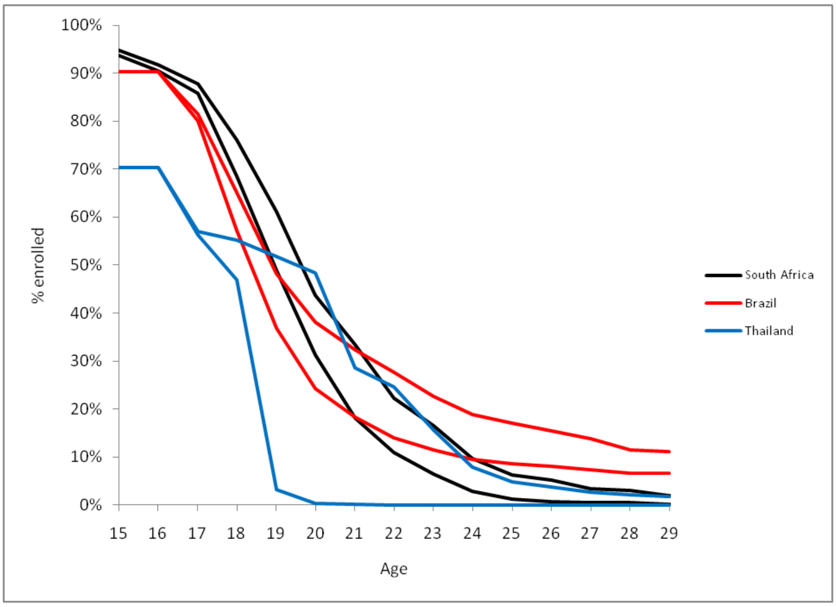
Figure 11: Investments in enrolments in middle income countries



Source: UIS education statistics; in the case of South Africa the same sources as those used for Figure 6.
Note: Values for countries other than South Africa were the most recent ones available for the years 2004 to 2008.

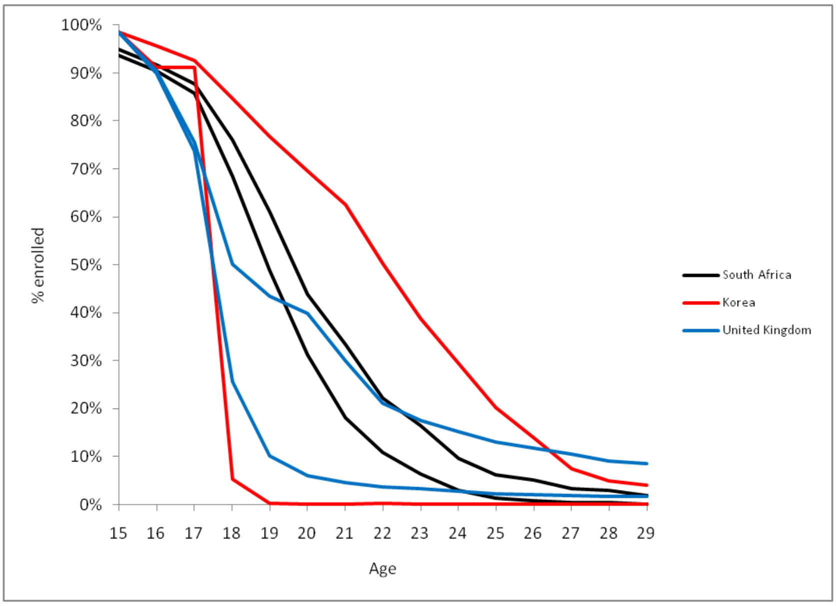
The cross-country comparison of enrolment levels by age of the following two graphs suggests that an optimal development strategy would include reducing over-aged enrolments in schools and directing resources towards more enrolments in post-school education. For each country, the bottom curve indicates the percentage of youths, by age, enrolled in school, whilst the distance between the bottom and top curves indicates the percentage of youths enrolled in some kind of post-school education. South Africa's current situation is not unique. Both South Africa and Brazil share a similar situation of high levels of over-aged enrolment at the secondary level (the situation here is somewhat worse in Brazil) and low post-school enrolments relative to school enrolments (see the narrowness of the gap between the two curves in the case of both countries). Thailand, despite having much lower secondary enrolment values before age 18, displays fewer over-aged learners and a post-school enrolment area that is over 70% larger than that in Brazil or South Africa. Figure 12, which compares South Africa to two fairly typical developed countries, the United Kingdom and Korea, confirms that with development should come less over-aged enrolment and a substantially greater level of enrolment in post-school education (though the difference between the UK and Korea with respect to the latter reflects how diverse the possibilities are.)

Figure 12: Cross-country comparison of enrolment by age and level



Source: OECD.StatExtracts. The South Africa source is as for Figure 2.
Note: For each country, the bottom curve indicates the percentage of youths, by age, enrolled in school, whilst the distance between the bottom and top curves indicates the percentage of youths enrolled in some kind of post-school education.

Figure 13: Cross-country comparison of enrolment by age and level



Source: See previous graph.

What are the salient points from the foregoing cross-country analysis? When it comes to attainment of the secondary grades up to Grade 11, South Africa fares better than seven comparator countries. However, the attainment of a Grade 12 pass, or successful completion of upper secondary schooling, appears to be on the low side for South Africa in a cross-country comparison. When it comes to post-school education, however, South Africa's enrolment situation (and that of our neighbouring countries) is amongst the lowest in the world.

5 When and why learners leave secondary school

In this section and the sections that follow, when the 2008 National Income Dynamics (NIDS) and 2009 General Household Survey (GHS) data are used to reflect what is occurring in secondary schools, non-public schools are not excluded from the analysis. This is not possible in NIDS because the differentiation between public and independent schools is not made. In the GHS the differentiation is made, but there are good reasons to heed the warning that many South African respondents tend to consider public schools charging relatively high fees as private schools, when in fact they are public. In the GHS 7% of secondary school learners appear to be in independent schools, when official enrolment figures suggest this figure should be around 3%. Importantly, the values reported here for all schools can be regarded as very close to the values for public schools only, given how small the independent school sector is. To give an example, the 2009 GHS dataset indicates that 26.9% of secondary learners are served a school lunch every day if all observations are used. If only public school observations are used (remembering that the GHS considers some public schooling independent) the statistic changes to 28.2%. Such differences are not large enough to influence the policy conclusions.

Figure 4 reflected what percentage of youths attain certain grades, even if this is through leaving the schooling system and then returning to it later. NIDS allows for a stricter view of dropping out insofar as it permits an analysis of youths who are in school one year and not in school the next year. In fact, the NIDS dataset allows for what are arguably the most reliable drop-out rates ever for South African schools. Table 1 reflects learners enrolled in 2007 (see row headings) and where they were in 2008 (see column headings). The '% dropped out' values follow the UNESCO methodology for the drop-out rate, in other words those who left school between 2007 and 2008 are divided by the total enrolment in their 2007 grade. Clearly, dropping out is already a problem at the Grade 8 level, with 6% of learners enrolled in this grade in one year being outside the schooling system the next year. Dropping out is even a problem in Grades 5, 6 and 7, where, according to NIDS, the drop-out rate is 1%, 2% and 3% respectively, though below Grade 5 the statistic is virtually 0%. To a large extent those who drop out before Grade 9 are over-aged for their grade. The fact that the cumulative value of 40% in Table 1 is close to the proportion of youths not attaining Grade 11 in Figure 4 suggests that once learners drop out the first time, they do not return to school and they are in a sense 'lost'. The dropping out situation seems to be slightly worse for girls. The Grade 11 cumulative drop-out value of 40% is 41% if only girls are considered and 38% for only boys.

Table 1: Dropping out of secondary schooling 2007-2008

2008 grade► *2007 grade▼*	*8*	*9*	*10*	*11*	*12*	*Dropped out*	*% dropped out*	*Cumul-ative dropped out*
8	63,217	451,537				32,032	6%	6%
9		90,423	703,378			76,160	9%	15%
10			197,332	673,043		120,728	12%	27%
11				192,567	658,671	128,889	13%	40%
12					161,403	547,846		

In response to the NIDS question on whether secondary school learners intend successfully completing Matric, 99% of learners say they do. This underlines the high social value attached to the Matric. The fact that around 60% of youths currently do not obtain the Matric provides a sense of how large the sense of educational failure must be amongst youths.

What do youths themselves say are their reasons for dropping out of secondary school? NIDS allows for an analysis not only of those dropping out, but also those who dropped out *mid-way* through the school year. In fact, dropping out mid-way through the year is much less common than dropping out after the end of the year. Only 1% of Grades 8 to 10 learners drop out within the year but the statistic rises to around 4% of enrolled learners in Grades 11 and 12 (so, for instance, around one-third of those who drop out at Grade 11 do so mid-way through Grade 11, whilst the other two-thirds do so after the end of the year). Amongst those who drop out mid-way through the year, the largest reason given for dropping out is lack of financing (37%), with the second largest reason being pregnancy (27%). For all Grades 8 to 11 drop-outs (regardless of when in the year dropping out occurred), the largest reasons for dropping out are lack of financing, pregnancy and wanting to look for a job (around 20% each).

The 2009 GHS provides a similar picture. Of those aged 20 or below who are not enrolled anywhere and have not completed Grade 12, 28% give financial problems as the reason for their non-enrolment whilst the reasons currently working, having failed the examinations and pregnancy each account for around 10%.

Clearly, pregnancy is a critical matter for female drop-outs. In NIDS, pregnancy is given as the reason for dropping out in the case of 42% of female drop-outs. The relationship between pregnancy and age provides an important indication of where interventions might be needed. As seen in Table 2, which uses NIDS data, only 2 of every 100 girls aged 15 years have given birth in the past or are currently pregnant, but this figure climbs steeply to 13% (9% + 4%) at age 16 and thereafter increases steadily at each age, reaching 47% at age 20 and 66% by age 22. Pregnancy is by no means always associated with dropping out. For instance, of the 31% of females aged 18 who have given birth or are currently pregnant, around 50% are still at school and this figure increases the younger the person. Yet becoming pregnant certainly appears to increase one's chances of dropping out, especially beyond age 18. In dealing with this issue, it appears important not to confine intervention programmes strictly to the 'teenage pregnancy' scope. As revealed by the GHS, the percentage of female learners aged 20 and above (and therefore no longer teenagers) in Grades 10, 11 and 12 is 9%, 19% and 27% respectively.

Table 2: Female schooling and pregnancy status in 2008[8]

	In school		Not in school		
Age	Has **not** been pregnant	Has been pregnant	Has **not** been pregnant	Has been pregnant	Total
15	92	1	6	1	100
16	83	9	4	4	100
17	65	10	17	8	100
18	38	15	31	16	100
19	17	6	43	34	100
20	12	10	41	37	100
21	7	7	37	50	100
22	4	3	31	63	100

Note: 'Has been pregnant' values in fact reflect females who are currently pregnant or who have ever given birth. They are therefore under-estimates in the sense that they do not reflect past pregnancies that did not result in childbirth.

There has been considerable public concern and some research around whether the South African 'teenage pregnancy problem' has become more serious over the years (see for instance Panday, Makiwane *et al*, 2009). There is indeed a shortage of relevant data to draw reliable conclusions on whether teenage pregnancy and adolescent fertility rates have been declining. One imperfect indicator is the percentage of young females who have a child living with them in the same household. This is available in the GHS and the 2003 to 2009 trend is illustrated in the next table. The indicator is imperfect mainly because infant mortality is ignored, as are children of the mother living in a different household. The figures in Table 2 suggest that there has been no or almost no change over the years. The steady decline from 21% of females aged 15 to 22 having a child at home in 2004 to a figure of 18% in 2009 could be indicative of a very gradual decline, but the size of the sample makes it impossible to be certain about this.

Table 3: Young mothers with children at home 2003-2009

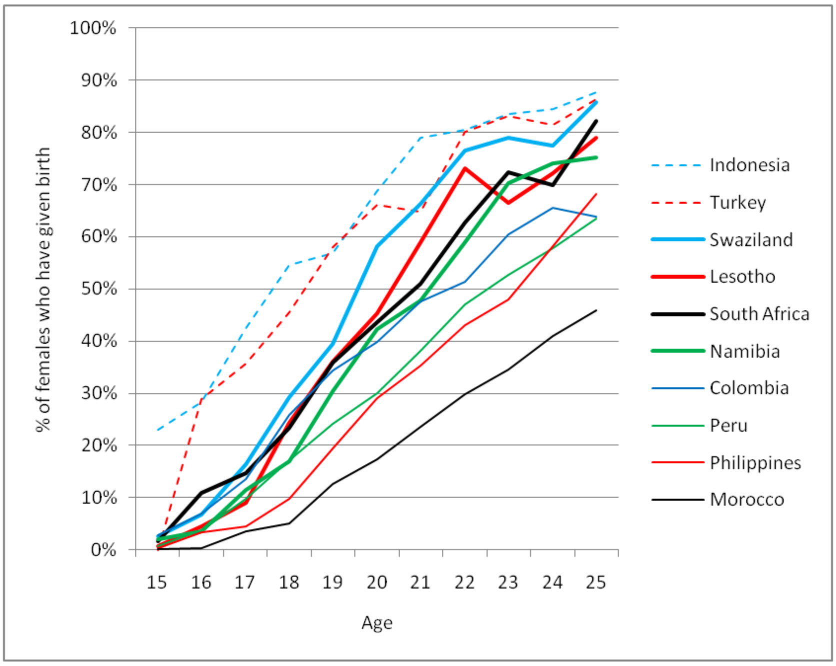
Age	2003	2004	2005	2006	2007	2008	2009
15	1%	2%	2%	2%	1%	1%	1%
16	3%	3%	4%	4%	4%	4%	4%
17	9%	9%	8%	9%	7%	9%	9%
18	10%	15%	16%	14%	13%	13%	13%
19	19%	24%	22%	18%	19%	21%	21%
20	26%	30%	32%	31%	24%	26%	26%
21	30%	38%	37%	32%	37%	31%	31%
22	40%	39%	38%	39%	40%	39%	39%
15 to 22	17%	21%	20%	19%	18%	18%	18%

In a global context, the situation in South Africa with regard to early motherhood is not unique, as illustrated in the next graph. South Africa more or less follows the pattern seen in three neighbouring countries (Namibia, Swaziland and Lesotho), especially below age 20. In certain countries beyond the region, specifically Indonesia and Turkey, having a child at a

[8] The levels reported on here agree broadly with what was reported in Panday, Makiwane *et al* (2009: 16). That source indicated that in 2003 around half of 20 year old women had given birth already. In Table 2 the corresponding figure is 53% (12% + 41%).

young age is more common than in South Africa, whilst for other countries such as Morocco and Philippines, this is less common than in South Africa.

Figure 14: Cross-country comparison of young mothers



Sources: 2008 National Income Dynamics Study for South Africa; Demographic Household Surveys for other countries, all from years after 2000.

The statistics on childbirth and pregnancy seem to suggest two policy responses. One is to try and reduce pregnancies amongst learners through better advocacy campaigns directed at youths. The other is to ensure that policies aimed at preventing unfair discrimination against pregnant female learners are correctly implemented across all schools, and not just some schools.

The fact that financial constraints should be a key factor behind dropping out raises the question of what the costs of schooling are. NIDS provides various details in this regard. 75% of secondary level learners paid during 2007 less than R650 in annual school fees, less than R350 in uniforms, less than R100 on books and stationery and less than R50 on other school items. These costs relate to one learner and the entire 2007 year. 43% of learners paid no school fees at all, and of these learners 95% were not asked to pay fees, implying that they were in no fee schools. The costs mentioned here may not seem high, yet clearly they constituted barriers for many youths from poor households.

The 2009 GHS provides details on transport to schools, a factor which could contribute towards the high cost of continuing in school. According to this source, 73% of secondary level learners walk to school, with the second largest category being travel by minibus taxi (9% of learners). 75% of learners take less 30 minutes to get to school, 4% take more than an hour and only 1% take more than 1.5 hours. In 2009, the GHS began asking whether youths were attending the closest available school. 83% said they attended the closest school. The figure was similar for all secondary grades. The most common reason for not attending the closest school was that it did not offer education of a sufficient quality (25% of the 17% who attended some other school). The second largest reason was that the right subjects were not offered. These figures suggest that transport could pose a financial barrier for a minority of learners, for instance the 9% making use of minibus taxis.

The 2009 GHS moreover asked what problems were experienced at school by individual learners in the household. Any number of the complaints listed in Table 4 below could be ticked. For 18% of secondary learners there was a complaint. Of these, 50% marked only one problem. Table 4 indicates that lack of books at the school stands out as the most serious problem. This suggests that the financial constraints problem could relate to an inability to afford books to compensate for inadequate provision by the school. The values for the complaints indicators in Table 4 have dropped since the question was introduced in the 2004 GHS. The percentage of secondary learners for which there were complaints has dropped steadily from 33% in 2004 to the 18% mentioned for 2009. Complaints relating to problematically high school fees have dropped most, from 14% in 2004 to 6% in 2009, a reflection of the introduction of no fee schools. In all the years 2004 to 2009 lack of books has been the largest complaint. This should be of great concern to policymakers given that a shortage of books is something relatively easy to remedy (for instance in comparison to the upgrading of teachers).

Table 4: Complaints about secondary schooling

	Grades 8-9	*Grades 10-12*
Lack of books	8%	9%
Poor quality of teaching	2%	4%
Lack of teachers	3%	4%
Facilities in bad condition	4%	5%
Fees too high	6%	6%
Classes too large/too many learners	6%	5%
Teachers are often absent from school	3%	3%
Teachers were involved in a strike	2%	2%
Other	1%	1%

Dropping out of school can occur gradually, with the learner attending a decreasing number of days as the year proceeds. According to the 2009 GHS, around 8% of secondary learners say they were absent from school in the week before the survey (in cases where the school was not closed for a holiday). For these 8% of learners the average number of days missed in the week was 2.0. Both figures are very similar across the five grades. For 52% of the learners who were absent, the reason was illness. Reasons typically associated with poverty, from not having transport money to the need to perform chores at home, accounted for 11% of absences in total. The overall level of absence seems high and the fact that only half of this was associated with illness points to a need for policy interventions such as further financial assistance for the poor and stronger incentives to ensure that learners attend school.

Lunches provided at school are often considered a way of encouraging learners to stay at school. In recent years, the education departments have tried to roll school nutrition programmes out into secondary schools. The 2009 GHS suggest that this has been relatively successful. 27% of secondary learners say they receive a school lunch every day and 30% say they receive one at least some of the time. At the primary level the two values are 66% and 75%. The lunch every day figure within secondary schools is highest in Grade 8, at 34%. The figure becomes lower as one moves up the grades, reaching 19% in Grade 12.

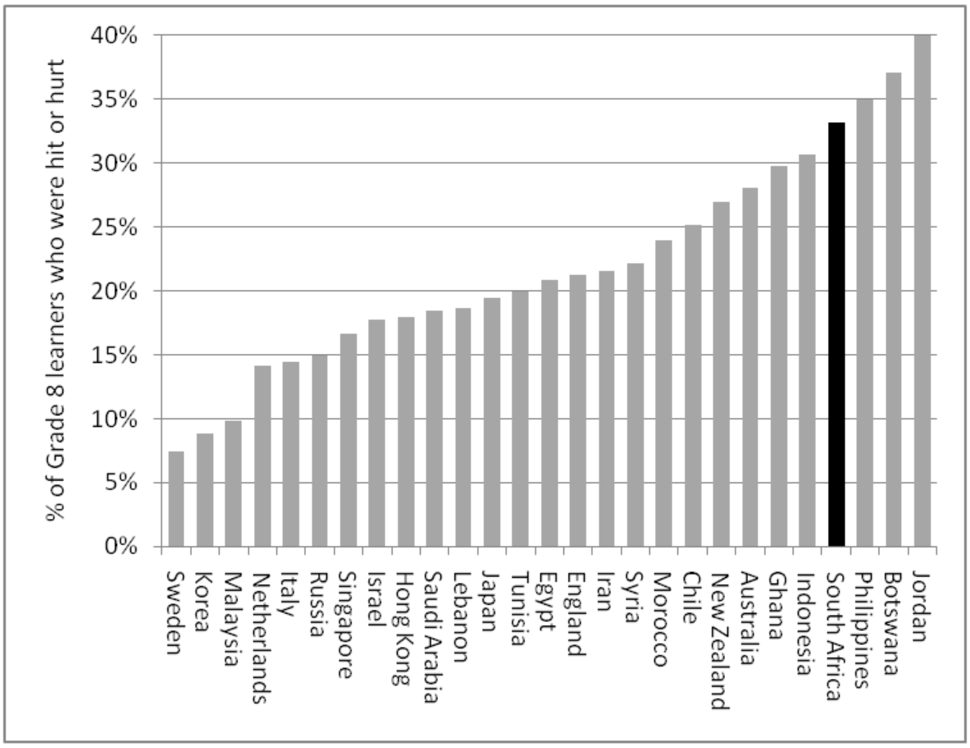
Learners are likely to miss school if they feel unsafe at school. This factor accounted for only 1% of absences in the 2009 GHS. The GHS moreover provides school violence statistics that are below what media attention in recent years would suggest is true. 19% of secondary learners say they have experienced some physical or verbal abuse at school in the months preceding the survey, but as Table 5 indicates, the bulk of this abuse was in the form of corporal punishment by teachers (more than one item in the list could be ticked). Values for male and female learners are very close to each other.

Table 5: Types of violence experienced

	% of all sec. learners
Corporal punishment by teacher	17%
Physical violence by teacher	1%
Verbal abuse by teacher	1%
Verbal abuse by learners	2%
Physical abuse	1%
Other	0%

The TIMSS 2003 dataset for Grade 8 learners provides a very different picture. TIMSS includes five questions relating to theft, physical violence, teasing or exclusion experienced by learners. Figure 10 illustrates responses to just one of the questions. For all five questions South Africa is in position 2, 3 or 4 amongst the Figure 10 countries (where position 1 is the worst) and the percentage of learners complaining ranges from 33% (seen in the graph in relation to being hit or hurt) to 52% (learners who were made fun of). It is inconceivable that differences between the 2003 TIMSS and 2009 GHS sources are attributable to the years being different. These discrepancies underscore the importance of probing different sources, in particular where survey questions deal with more subjective issues.

Figure 15: Cross-country comparison of school violence



Source: TIMSS 2003.

NIDS includes a number of emotional health questions. 32% of secondary learners say they are hopeful about the future all the time (68% are at least some of the time). 3% say they are depressed all the time (34% at least some of the time). There is a slightly larger percentage of learners depressed all the time in Grade 12 than in Grade 8 (5% against 2%), but at the same time the percentage of learners who are hopeful all the time is also higher in Grade 12 (37% against 17%). These figures confirm what is already widely recognised, namely that Grade 12 learners require rather strong psycho-social support.

Alcohol and drug abuse is often put forward as evidence of emotional ill health and as a factor leading to poor learning. NIDS indicates that the percentage of male learners who say they have tried alcohol increases steadily from a low 5% in Grade 8 to 35% in Grade 12. The corresponding figures for female learners are lower: 2% and 12%. However, only 1% of secondary learners say they drink alcohol every week. Where they do the number of drinks may be very high as reflected by an average of 18 drinks a week (but this is based on just 24 responses in the dataset). The 2009 GHS indicates that 1.4% of secondary learners suffer from alcohol or drug abuse, with the value varying little by gender or across grades.

A multivariate logit model was constructed to examine what factors appear prominently associated with having dropped out between 2007 and 2008 in the NIDS dataset. Explanatory factors tested included gender, level of income, level of poverty, having had a child (in the case of females), whether one's geographical area was rural or urban and formal or informal and, finally, one's score in the NIDS numeracy test. The poor explanatory power of the model (pseudo R^2 of just 0.054) makes it not worth reproducing in the report. The only statistically significant explanatory variable was having had a child. That this variable should appear more significant than academic performance (in the numeracy test) and household income is telling and underlines the importance of dealing adequately with the question of learner pregnancy in the education policies.

6 Subjects, standards, grade repetition and qualifications

6.1 Grade-on-grade performance improvements

The 2008 National Income Dynamics Study (NIDS) dataset includes results from a mathematics test taken by respondents who were present during the fieldworker visit and who were willing to take the test. As explained by Van Broekhuizen and Von Fintel (2010), the taking of the NIDS mathematics test was far from random and involved several selection effects. What is important for the analysis that follows is that students were far more likely to take the test than non-students. The NIDS test was in fact four mathematics tests at different levels of difficulty. The test taken depended on the highest grade at which one had studied mathematics. The cut-offs were set at Grades 4, 7 and 10 meaning, for instance, that if someone's highest or present grade was Grade 4 or lower, the first (easiest) test would be taken. If someone's highest or present grade was in the range Grade 5 to 7, the second test would be taken. The range Grade 8 to 10 meant the third test would be taken and the range Grade 11 and above meant the fourth test would be taken. It is important to bear in mind that the grade considered was the highest grade at which mathematics was taken. During the years immediately prior to the 2008 NIDS survey, specifically during 2005 to 2007, a rule in the national curriculum was phased in whereby all learners in Grades 10 to 12 would have to take either mathematics or mathematical literacy. Before this, it was possible to do neither of these subjects in these three grades. Mathematics has always been compulsory in Grades 1 to 9, however. The NIDS test did not clarify whether mathematical literacy should be counted as mathematics in determining which test to take. The data suggest that many respondents did in fact not consider mathematical literacy to be mathematics in the NIDS process.

The presence of the mathematics test in NIDS provides opportunities for probing the degree to which schooling makes a difference to mathematical knowledge and the benefits of this knowledge, which can to some degree be considered a proxy for cognitive skills in general, in the labour market. Both the relationship between schooling and mathematical knowledge and the relationship between mathematical knowledge and success in the labour market are examined in this section. The overall objective, however, is to evaluate what the evidence suggests to policymakers about qualitative changes needed in schooling (keeping in mind that the previous sections have dealt largely with the more quantitative matters of, for instance, grade attainment).

The youngest age at which a substantial proportion of household members took the test was 12. Those aged 17 had the highest test participation rate, at 46%. Of those aged 12 to 20, 41% took the test, of those aged 21 to 30, 28% took the test and of those aged 31 to 40, 19% took the test. Clearly, participation diminished with age. The number of respondents for whom test results and other data needed for the various analyses that follow was 4,351. Results must obviously be interpreted cautiously, both due to the smallness of the sample (different analyses require different sub-sets of the 4,351 test-takers) and selection effects. Moreover, the shortness of the test means that the ability of results to discriminate between respondents is not great. Each tests, regardless of level, consisted of 15 multiple-choice questions, with each question carrying four options. This means that random guessing will tend to result in a percentage correct score of 25% (or in being correct in an average of 3.75 of the 15 questions). As will be seen below, in many secondary school learners obtained scores only slightly above the 25% threshold, suggesting that they in fact were able to master only one or two of the 15 questions.

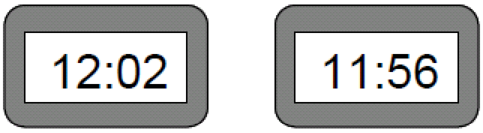
The scores from the four tests were used by the NIDS team to produce one normalised score that took into account the different levels of difficultly of the tests and thus should make it possible, at least in theory, to compare the mathematics knowledge of respondents who wrote different tests. How the four sets of test scores relate to the normalised score is examined below. First, however, some basic analysis of the scores in just the third test are presented to paint an initial picture. The third test, which would be written by learners whose highest mathematics grade was in the range Grade 8 to 10, provides the largest number of observations if one’s focus is on understanding the performance of learners in the secondary grades. By focussing on just one test the question of the reliability of the normalised score based on all four tests is avoided.

What is immediately striking when the scores from the third test are examined against the current enrolment of respondents is, firstly, how low the scores are and, secondly, how little they improve from one grade to the next. The average number of correct questions for learners currently enrolled in Grades 8, 9, 10, 11 and 12 was 4.1, 4.1, 4.3, 4.9 and 4.8. This is indeed only slightly above the 3.75 random guessing threshold, for all five grades. These averages are based on the data of 743 learners of whom at least 116 are in each of Grades 8 to 11 (there were only 52 test-takers in Grade 12). The sample is small and in Grades 10 to 12 would cover mainly those learners who were not good at mathematics. This is why they did not proceed beyond Grade 10 (and in all likelihood beyond Grade 9 given that mathematics proper is only compulsory to the end of Grade 9). Yet even if the average scores referred to above are considered a rough indication of the situation, the problem of insufficient progress in cognitive skills from one grade to the next seems confirmed. One should bear in mind that Grades 8 and 9 learners are all required to take mathematics, so grade-on-grade improvements can be expected from Grades 8 to 10 (NIDS 2008 occurred in the first half of 2008 so even Grade 10 learners who dropped mathematics should have had the benefit of additional Grade 9 schooling in mathematics).

Table 6 below provides a view of a selection of questions from the third test. The two questions learners did best in, questions 4 and 7, and the two they fared worst in, questions 1 and 9, are shown. Clearly, for the latter two questions the percentage correct statistics are below the random guessing threshold of 25% for all grades. Only in the case of question 4 and one more question not shown in Table 6 was the percentage correct statistic consistently better the higher the grade in the range Grades 8 to 11. For nine of the 15 questions the percentage correct statistics per grade yielded the expected positive least squares trendline. Yet even where improvements by grade existed, they tended to be weak. In an analysis which considered the top quintile of performers in each grade separately, the grade-on-grade improvement was virtually the same as for the sample as a whole. This rather surprising result suggests that even better performing learners (who tend to be from better performing schools) have difficulty improving their mathematics from one grade to the next.

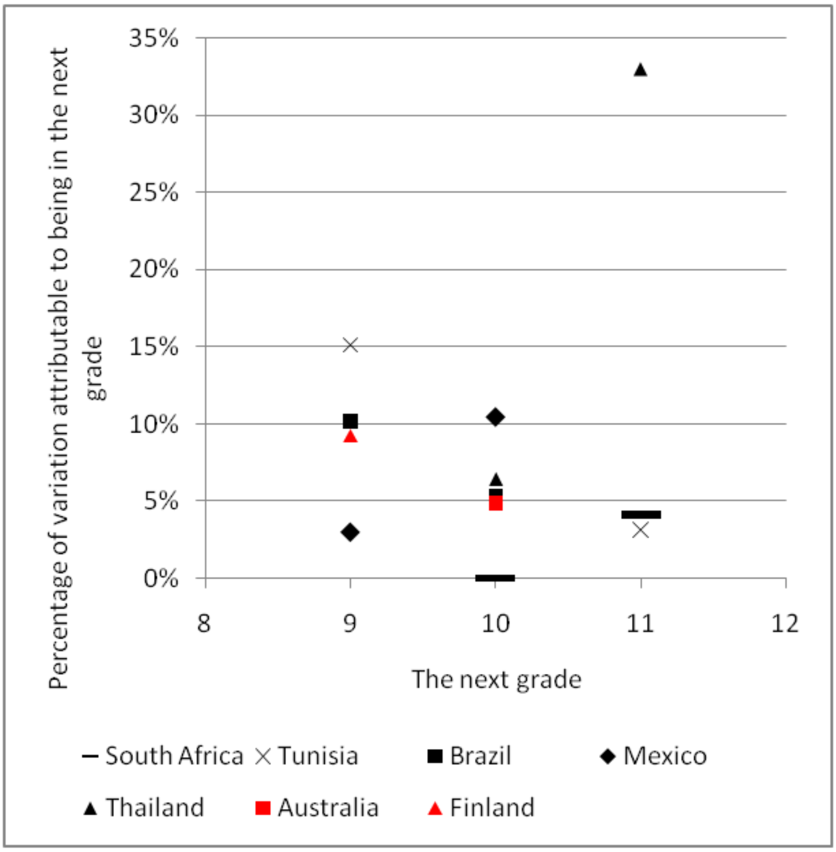
Table 6 suggests that not only do learners have problems in mathematics, they experience problems taking tests of any kind and specifically they are unable to understand written test questions as one might expect. In this regard, question 7, which only 50% of learners got correct, is telling. If one can assume that over half of learners are able to tell the time on a digital interface (this assumption seems to be a reasonable one), then many learners failed to select the correct option because they were unfamiliar with tests generally or multiple-choice tests specifically, or had problems understanding the question.

Table 6: Percentage correct responses to individual test items by grade

Test question	*8*	*9*	*10*	*11*	*12*	*8 to 12*
4. If 3+ ? =16, which number should replace? a) y = 10 b) y = 11 c) y = 13 d) y = 24	72	75	83	89	91	81
7. What's the time difference between these two watches?  a) 4 min b) 1 h 4 min c) 6 min d) 1 h 6 min	49	45	47	69	45	50
1. $\left(\frac{1}{2}-\frac{1}{3}\right)-\left(\frac{1}{4}-\frac{1}{3}\right)=$ a) $\frac{1}{12}$ b) $\frac{1}{6}$ c) $\frac{1}{4}$ d) $\frac{1}{3}$	21	20	19	19	17	19
9. Convert 2 and ¾ days into minutes a) 3600 b) 3660 c) 3900 d) 3960	8	10	9	15	14	11

There is generally little data that is comparable across countries and that reflects performance of learners in different grades. Such data can potentially measure the schooling system's ability to add value to the learning process in each grade. Data from the OECD's Programme for International Student Assessment (PISA) do however provide an opportunity to gain a picture of grade-on-grade improvement in mathematics at the secondary school level in a few other countries. Instead of focussing on a specific grade, PISA focuses on 15 year old youths, regardless of their grade. The performance of test-takers in different grades can therefore be compared. Figure 16 illustrates the comparison using a statistic that ought not to be relatively insensitive to whether the PISA test or South Africa's NIDS test was used. In the case of PISA, 2003 test scores were used. The vertical axis refers to the intraclass correlation coefficient where grade is the class. The statistic indicates the percentage of the variation between learner scores attributable to being in the next as opposed to some other factor such as the learner's innate ability and the social disadvantage experienced by the learner. To illustrate, if we consider the South African test-takers from Grades 10 and 11 as a set and calculate the intraclass correlation coefficient, specifying grade as the differentiator (or class), then it is found that being in Grade 11, as opposed to Grade 10, accounts for 5% of the variation of the scores within the set of learners. A similar analysis of the set of Grades 9 and 10 learners reveals that 0% of the variation is explained by being in the next grade. The set of Grades 8 to 9 learners is excluded from the analysis because the average test score in Grade 9 was lower than that in Grade 8. Turning to the PISA countries, ten sets of learners, where each set spans two grades, are illustrated. Where a specific country is excluded at a specific grade, this is always due to an insufficient number of observations (as PISA tests only 15 year olds the distribution within one country will be concentrated across two or at most three grades). As explained in Appendix E, in order to minimise bias associated with grade repetition, sets of PISA learners with similar grade repetition histories were analysed.

Figure 16: Cross-country comparison of grade-on-grade progress

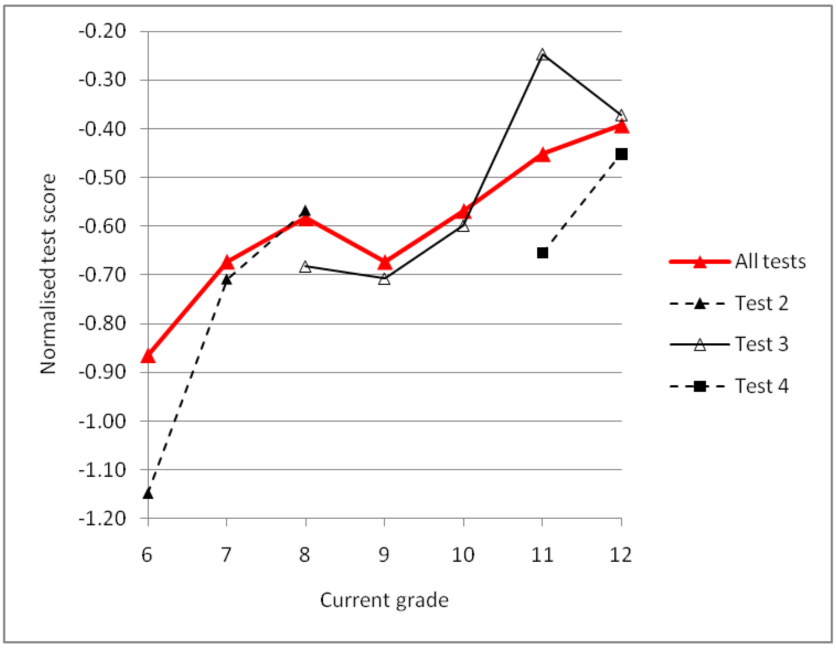


Source: National Income Dynamics Study 2008 for South Africa. PISA 2003 dataset for other countries (available at http://www.oecd.org/document/0/0,3343,en_32252351_32236173_33694144_1_1_1_1,00.html).

Figure 16 suggests that by international standards grade-on-grade progress in mathematics in South Africa is below what it should be. It is significant that even Brazil, a country with levels of social and income inequality similar to those in South Africa, should display higher statistics than South Africa (the greater the levels of social inequality in a country, the more the variation in scores would be explained by these inequalities, and the less the variation would be explained by being in the next grade).

Figure 17 illustrates the test score picture with respect to learners enrolled in Grades 6 to 12. The values along the vertical axis are the normalised test scores calculated by the NIDS team. A point was only included in the graph if at least 50 learners had results. This explains why only tests 2, 3 and 4 are reflected, and not test 1 (the easiest test, which was written mainly by younger learners). This also explains why no point beyond Grade 12 is reflected. Students enrolled in higher education who also took a test came to fewer than 50. This is unfortunate, as it would have been useful to compare the value added by being in higher education, to the value added to being in a higher grade within school. The total number of Grades 6 to 12 learners who took a test was 606 for test 2, 763 for test 3 and 319 for test 4. The number of learners in these grades with a normalised test score was 1,818 (this figure includes a few takers of test 1). The general pattern displayed in the graph is of an increase in test score by grade, with an abnormal dip appearing for Grade 9. The sample is so small, and selection effects strong, that no meaningful policy information can be read into the Grade 9 dip. What is telling, however, is that the overall picture of poor grade-on-grade improvements is confirmed. One should bear in mind the small differences referred to above. As mentioned earlier, the test 3 curve in Figure 17 in fact ranges from 4.1 questions correct to 4.9 questions correct. Considering the test consisted of 15 questions, this range is remarkably narrow. The hypothesis that accountability pressures linked to the Grade 12 national examinations would cause a surge in performance between, say, Grade 11 and Grade 12, is clearly not supported by the evidence presented here, as compelling as this hypothesis may seem[9].

Figure 17: Mathematics scores amongst the enrolled



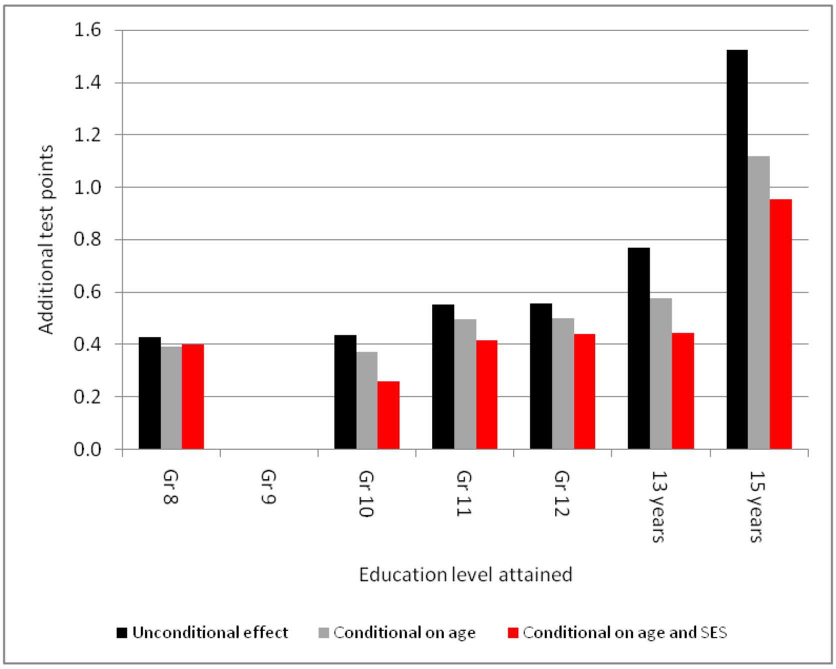
Source: National Income Dynamics Study 2008.

[9] See for instance Hanushek and Woessman (2009: 13).

6.2 School performance and the labour market

Even though the data do not allow us to compare the test scores of school learners with post-school students, they do allow us to compare the scores of those with and without post-school education amongst *non*-students. In the following graph the unconditional columns reveal a picture for the schooling system that is similar to what was seen in the previous graph. The zero point on the vertical axis in Figure 18 is the average test score for non-students with less than Grade 6. The difference between Grade 8 and Grade 12 is small, around 0.2 normalised score points in both graphs, or a grade-on-grade increase of around 0.05 points. In contrast, the difference between the scores of non-students with Grade 12 and those with around one year of post-school education is relatively large, around 0.2 points. The difference of almost 1.0 between Grade 12 and fifteen years of education, or three years of post-school education, is even larger. This difference translates into about 0.3 points per year. As with most of these kinds of analyses in South Africa, there is a fair degree of guessing in the translation of response values to years of education. Specifically, 13 years of education is mainly represented by Grade 12 plus a certificate, whilst 15 years of education is mainly represented by a bachelors degree. Even if it were possible to tighten up this aspect of the analysis, it seems unlikely that the overriding conclusion would change, namely that a year of schooling adds less to one's level of mathematical skills than a year of post-school education. Whether this pattern in South Africa is unusual could be not be established as similar analyses from other countries were not found.

Figure 18: Mathematics scores amongst the non-enrolled



Source: National Income Dynamics Study 2008.
Note: Grade 9 values are excluded due to insufficient statistical significance. Where age is used as a conditioning variable, age 30 was assumed when the height of the column was calculated.

We now turn to the conditional columns in Figure 18. The underlying regression analysis is described in Appendix F. Age is often considered a proxy for work experience, which could translate into better test results. Indeed, when the analysis is made conditional on age, the differences across the columns are smaller, though the very large score increase associated with having a degree remains evident. However, it is the analysis that is conditional on age, race and parent's level of education that provides what is in many senses the truest picture.

When all these factors are controlled for, it is only having a degree which makes a difference to one's mathematical abilities.

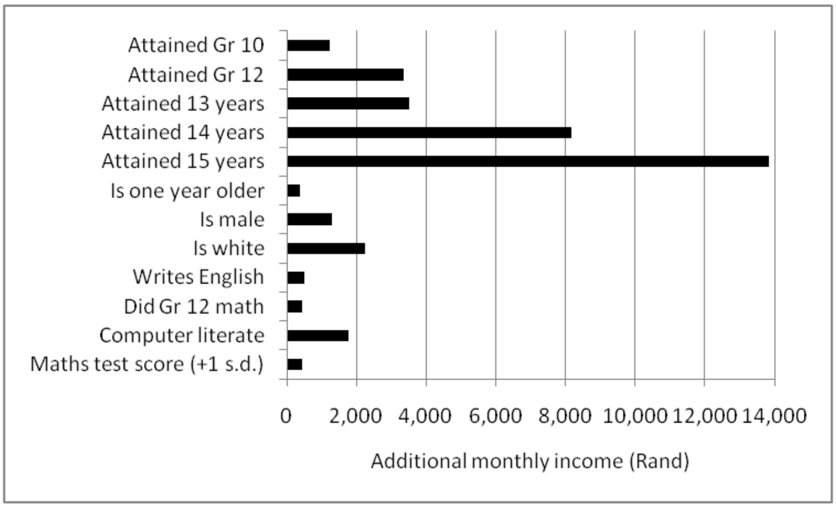
Appendix F, Table 20, provides the results of an earnings function analysis in which the impact of various educational and non-educational factors on labour market earnings are explored. The analysis uses the NIDS dataset. Figure 19 below summarises the findings. Each bar illustrates the additional monthly 'take-home' earnings associated with the factor in question. The point of departure is a monthly wage of R3,041, which is the mean value across the wage earners in the analysis[10]. Clearly, education beyond Grade 12 is associated with considerable earnings benefits. In particular, 14 years of education (Grade 12 plus a diploma) and 15 years of education are associated with additions to one's monthly income of around R8,000 and R14,000 respectively (this is relative to having less than a Grade 8 level of attainment). However, even having completed Grade 12 is associated with noticeable benefits, especially if one considers that this factor is larger than either being white or being male, factors that are widely considered to still play an important role in the labour market. To a considerable degree the Grade 12 labour market benefits would be linked to the fact that the Grade 12 examinations lead to the only widely recognised qualification in the schooling system (and a qualification which, according to previous analysis in this report, 60% of youths do not obtain). The additional earnings associated with a score in the mathematics test that is one standard deviation higher are small, at R432, though one standard deviation translates into an improvement of just two additional questions correct (out of the 15). It is thus conceivable that an improvement from five questions correct (the average for a learner enrolled in Grade 12) to 15 correct would be associated with five times this monetary benefit, or R2,160 (even this, however, would not represent a remarkable improvement when compared to the other bars in Figure 19). The magnitude of the impact of the mathematics test score on earnings is in fact similar to what has been found in comparable models run on data from other countries[11].

Around half of income earners responded that they were able to write 'very well' in English. This characteristic is associated with an increase of R513 in monthly earnings. A similar increase is associated with having taken mathematics in Grade 12, something which 17% of income earners did. What is especially noteworthy is the relatively strong association between being computer literate, specifically saying one is 'highly literate' or masters 'basic use', and income (one-third of income earners considered themselves computer literate according to these criteria). The monthly increase associated with this characteristic is R1,758. These associations are not necessarily a reflection of pre-employment education influencing employment earnings. Skills such as computer literacy are acquired during employment too. Yet the statistics are telling for education policymakers even insofar as they point towards the skills valued in the labour market. In particular, the fact that only half of income earners considered themselves able to write 'very well' in English points to a clear need for improvements in this language in schools. If one focuses on all adults aged 20 to 29, regardless of employment status, then 50% of respondents considered they were able to write 'very well' in English but 10% said they did this 'not well' and 5% responded 'not at all'. Moreover, promoting computer skills in schools carries clear labour market advantages. Of all adults aged 20 to 29, only 31% considered themselves computer literate using the criteria mentioned previously.

[10] Because the model used the natural log of income, and not the raw income value, as the dependent variable, it is necessary to have an explicit point of departure when calculating the difference made by the various explanatory factors.

[11] See for instance Hanushek and Zhang (2009: 131).

Figure 19: Effects of educational variables on monthly income



Source: National Income Dynamics Study 2008.
Note: See Appendix F, Table 20, for details on the model underlying the values used for this graph. Model D in Table 20 was used for all values except for the maths test score value, for which Model C was used.

A model predicting one's probability of being employed was also constructed (see Table 21 in Appendix F). This model resulted in findings similar to the ones discussed above in relation to the earnings function. The employment model confirms the importance of computer literacy, which is associated with a 7 percentage point improvement one's probability of being employed. A one standard deviation improvement in one's mathematics test score is associated with a 3 percentage point improvement in one's probability of being employed. However, writing English well and having done mathematics in Grade 12 were not found to be associated with being employed (even if these characteristics are associated with better earnings once one has a job).

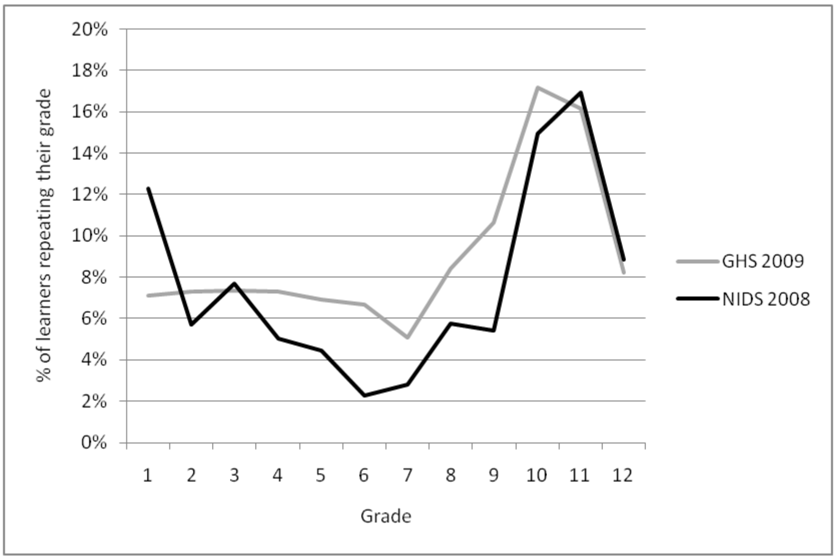
6.3 Grade repetition

The NIDS 2008 data in combination with the 2009 General Household Survey (GHS) data make it possible to take a major step forward when it comes to information on grade repetition in South African schools. Both data sources include information on what grade is being repeated in the current year. Such information has not been available in any previous nationally representative South African household dataset[12]. It is widely recognised that household data offer distinct advantages over school census data with respect to grade repetition information given that the latter source is highly susceptible to measurement problems and deliberate under-statement of grade repetition. Figure 20 reflects the percentage of learners repeating their current grade according to NIDS 2008 and the 2009 GHS. Both curves include both public and non-public schools. Table 16 in Appendix D provides the underlying values. The most noteworthy deviations between the two sources are, firstly, that Grade 1 repetition appears to be considerably higher in the NIDS dataset and, secondly, that between Grades 4 and 9 the NIDS values are 2 to 4 percentage points lower than the GHS values. From the secondary school policy perspective, what stands out is the exceptionally high level of grade repetition in Grades 10 and 11. Both data sources point to the existence of this phenomenon. A key explanation is probably that as learners approach the Grade 12 examination, which is the first point at which high stakes accountability pressure is brought to bear on schools, teachers raise standards above what learners have been accustomed to,

[12] One household dataset that does include this information but has not been widely available is a the 2004 HSRC Client Survey dataset, which is based on relatively small sample of around 5,700 households.

resulting in a holding back of learners in Grades 10 and 11. This would have the dual effect of reinforcing preparation for the Grade 12 examinations, through repetition, and discouraging learners with low scores from continuing with their schooling. One solution to this problem would obviously be to strengthen teaching and learning practices, as well as accountability mechanisms, in earlier grades, so that the pressures seen in Grades 10 and 11 could be spread out more evenly across the grades.

Figure 20: Grade repetition in NIDS and GHS



Source: National Income Dynamics Study 2008; General Household Survey 2009.

An attempt was made to test, at the national level using the NIDS dataset, the Lam, Ardington and Leibbrandt (2008) findings regarding grade repetition. These findings point to weaknesses in historically African schools, in the Cape Town area, when it comes to promoting and repeating the right learners. Specifically, it is suggested that schools-based assessments are not sufficiently precise in these schools to ensure that learners who really need to repeat are in fact the learners who are made to repeat. A logit model (see Table 17 in Appendix D) finds weak support for this hypothesis, though the results are not conclusive enough to draw firm policy conclusions.

6.4 Grade 12 subject combinations

Learner records from the 2009 National Senior Certificate database were made available in order to investigate specific policy issues, in particular the extent to which Grade 12 learners could be failing to obtain the NSC simply due to inappropriate subject choices. Before an analysis in this regard is discussed, some general information on subject choices is presented. The following two tables (and several tables within Appendix G) are based on a set of 528,986 learners. These are learners in public schools who sat for seven or (rarely) eight subjects. Generally two subjects were language subjects and five subjects were non-language subjects. Life orientation was a compulsory non-language subject for all learners. One of the non-language subjects had to be mathematics or mathematical literacy. If one considers just those learners who had four non-language subjects other than life orientation there were altogether 3,490 different combinations of these four non-language subjects, though as the following table indicates, 58% of learners took one of the ten most common combinations. A quarter of learners took the two most common combinations, which are (1) mathematics, geography, physical science, life sciences and (2) mathematical literacy, accounting, business studies, economics (all the four-letter subject codes are explained in Table 25 in Appendix G).

What is noteworthy is how subject variety decreases the poorer the quintile. In quintile 5, 28% of learners took the ten nationally most common combinations. The figure rises steadily to 70% in the case of quintile 1. An alternative way of demonstrating this inequality is to count the quantity of the overall 3,490 combinations appearing within each quintile. The figures from quintile 1 to quintile 5 are: 788, 597, 1135, 1371, 2838. Within quintile 5 there are three times as many subject combinations as within quintile 1.

Table 7: Cumulative participation in most common non-language combinations

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Overall
MATH + GEOG + PHSC + LFSC	15	17	15	14	8	14
MLIT + ACCN + BSTD + ECON	27	30	28	25	13	25
MATH + ACCN + BSTD + ECON	38	41	39	34	17	34
MATH + PHSC + LFSC + AGRS	50	51	47	37	18	42
MLIT + HIST + GEOG + LFSC	57	57	52	42	20	47
MLIT + GEOG + LFSC + AGRS	61	61	55	43	21	49
MATH + PHSC + LFSC + ACCN	62	63	58	46	25	52
MLIT + HIST + GEOG + TRSM	64	65	60	49	25	54
MATH + PHSC + LFSC + CATN	66	67	62	51	28	56
MLIT + HIST + GEOG + AGRS	70	69	64	52	28	58
Other combinations	100	100	100	100	100	100

Note: The combinations are sorted from most to less common (at the overall level). The raw numbers behind the percentages appear in Table 22 in Appendix G.

What are the less usual subjects that more advantaged learners have better access to? Table 23 provides a picture of the situation with reference to all the non-language subjects listed in the official curriculum for the Further Education and Training band (Grades 10 to 12)[13]. Clearly subjects requiring specialised equipment and possibly smaller classes are much less accessible to the poor. For instance, only 9% of examination candidates in visual arts are from quintiles 1 and 2, which together comprise 40% of all candidates. On the other hand, 58% of learners in this subject are from quintile 5, which comprises less than 20% of all candidates[14]. Apart from visual arts, the following subjects are all characterised by over 40% of learners being from quintile 1 and fewer than 20% being from quintiles 1 and 2 combined: design, information technology, computer applications technology, engineering graphics and design, hospitality studies. The situation is still inequitable, though slightly less so, in the following subjects: dance studies, music, mechanical technology and civil technology. Given how low the overall enrolment is in nearly all of these subjects, the notable exception being computer applications technology, the solution appears to lie in expanding the presence of all these subjects above all across quintiles 1 to 4. The largest part of the challenge would arguably be to increase the number of teachers able to teach these subjects. It is noteworthy that many of the FET schools subjects have a vocational orientation. In fact, the list of subjects in Table 8 appears to negate the commonly held view that equates FET in schools with general education and FET in the colleges (which follows a different curriculum) with vocational education. The curriculum in schools allows for a considerable vocational focus, in particular if the smaller subjects can be expanded across the full range of schools.

[13] Government Notice 1407 of 2003.

[14] There are many reasons why each quintile does not comprise exactly 20% of all learners. For instance, the method for determining quintiles is based on household data and not enrolment data and the figures in Table 8 exclude independent school enrolment (this is high among the relatively rich).

Table 8: Individual subject participation across quintiles

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Overall	Learners
Dance Studies	6	14	18	22	39	100	330
Music	11	11	16	18	44	100	1,291
Visual Arts	5	4	15	17	58	100	6,116
Dramatic Arts	13	7	23	31	26	100	5,460
Design	6	3	11	19	62	100	2,250
History	28	17	25	18	12	100	84,030
Geography	26	18	26	17	12	100	199,725
Mathematics	24	16	24	18	17	100	269,697
Mathematical Literacy	24	16	25	18	17	100	259,260
Physical Science	24	16	25	18	17	100	205,368
Life Sciences	26	17	25	17	15	100	277,569
Information Technology	9	4	13	20	54	100	5,528
Computer Applications Technology	10	10	15	19	46	100	46,847
Agricultural Sciences	40	21	27	8	4	100	84,818
Accounting	23	16	26	19	16	100	162,426
Business Studies	23	15	25	18	19	100	190,601
Economics	26	18	27	18	10	100	142,446
Mechanical Technology	7	7	19	32	35	100	6,842
Electrical Technology	12	9	23	26	29	100	6,197
Engineering Graphics and Design	9	7	19	25	41	100	24,715
Tourism	21	12	24	22	21	100	67,865
Hospitality Studies	10	10	19	21	40	100	10,065
Consumer Studies	15	14	21	20	30	100	31,083
Civil Technology	9	8	19	27	37	100	9,334
Religion Studies	25	12	27	28	8	100	1,621
Overall	24	16	25	18	17	100	528,986

Note: The first six data columns refer to percentages of learners.

A table like the above but with a breakdown by province (see Table 25 in Appendix G) reveals a few notable patterns. Dance studies and music are particularly concentrated in Western Cape. Some 63% of dance studies learners are in this province, whilst the figure for music is 30%. Clearly there is a need to develop these two subjects in the rest of the country. Design is highly concentrated in Gauteng and Western Cape, which together account for 79% of all examination candidates. The tendency to take mathematical literacy as opposed to mathematics is particularly strong in Northern Cape, which could be an indication of the weakness of mathematics in this province, or possibly more appropriate subject choices being made (the analysis discussed below will show that many learners who should take mathematical literacy take mathematics instead). Computer applications technology has a strong presence in Gauteng and Western Cape, but also, perhaps surprisingly, in Northern Cape and Free State. The other five provinces appear to have too few enrolments in this subject. As one might expect, the general pattern is for agricultural sciences to be strong in more rural provinces, such as Limpopo and Mpumalanga, and weak in more urban provinces, such as Gauteng and Western Cape. Breaking this pattern however, is a relatively weak presence of agricultural sciences in Free State, despite the large agriculture sector in this province. Mechanical technology and electrical technology stand out as being particularly poorly represented in Eastern Cape, Mpumalanga and Limpopo. Tourism and hospitality studies seem poorly represented in Limpopo and North West, especially given the strength of the tourism industry in these provinces.

Education analysts have long suspected that inappropriate subject selection in Grade 10, the grade in which the FET curriculum begins for learners, lies behind some of the poor Matric results, in particular within schools situated in less advantaged areas. The analysis that follows seems to confirm this suspicion.

Table 9 below provides the results of the analysis. A random sample of 5,000 examination candidates from 2009, all of whom were from public schools, were analysed through a computer program that tested, in the case of learners who did not obtain the NSC, what the outcome may have been had different Grade 12 non-language subjects been chosen[15]. Of the 5,000 learners, 59% passed the examinations and obtained the NSC. This is close to the official pass percentage for all candidates of 60.6% (this figure includes independent school learners, who tend to perform better than public school learners). Of the 5,000 learners, 2% failed to obtain the NSC because they did not obtain 40% in their first language. Remedying the situation for these kinds of learners was not a consideration in the analysis. Rather, the analysis focussed on those learners who did not obtain the NSC, but might have had they switched from mathematics to the easier mathematical literacy, or if they had switched one of the three other non-language subjects (other than life orientation) to another non-language subject. Of those learners who passed their first language but did not obtain the NSC (38% of 5,000 learners) the great majority were in schools where other non-language subjects existed that they could in theory switch to. The data did not permit for an explicit analysis of timetabling barriers, though implicitly this was taken into account through the avoidance of unusual combinations. In other words, it was assumed that if another non-language subject was taken by another examination candidate in the school, then a switch to that subject was possible and a hypothetical score for the subject was estimated if the data permitted this (meaning if the outcome would not be a highly unusual combination). Only 1% of learners (or 50 learners in the 5,000 sample) found themselves in a situation where there were no subjects to which they could switch in the school. It was found that 1,894 of the 5,000 learners (or 15% of the learners) could have obtained the NSC they failed to obtain if they had switched from one non-language subject to another one available within the school. This implies that the pass rate of 59% could have risen by 15 percentage points, to 74%, had different non-language subject combinations been taken. This is obviously a very large difference. The analysis clearly has a diverse range of important policy implications, though important limitations to the analysis described below need to be taken into account.

[15] The program, in Stata, can be obtained from the author of the report (see e-mail address on the first page). The program can be used for any sample size or for all learners in the database. The advantage with a sample is that the time that the program needs to run is reduced. A random sample of 5,000 learners (not clustered by school) is large enough to yield results that for the purposes of this study will be virtually identical to the results one would obtain if all learners were analysed.

Table 9: Results of a subject switching analysis on a random sample of learners

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Overall
Randomly selected learners analysed	1,144	801	1,267	944	844	5,000
Learners not passing (%)	53	47	43	39	16	41
Learners who did not pass due to home language problems (%)	1	2	2	3	2	2
Non-passing learners without home language problems who had subjects within the school they could switch to (%)	51	43	41	36	14	38
Non-passing learners without home language problems who had NO subjects within the school they could switch to (%)	1	1	0	0	0	1
Non-passing learners without home language problems who had subjects within the school they could switch to and for which estimates of alternative scores were possible (%)	51	43	40	36	14	38
Non-passing learners for whom alternative scores could be calculated who were found to pass	17	17	16	15	6	15

Note: All percentages refer to a part of the randomly selected learners referred to in the first row.

Table 10: Most common subject switches resulting in a new overall Grade 12 pass

Switch from...	Switch to...	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Overall
Mathematics	Mathematical Literacy	24	24	24	24	20	24
Physical Science	Business Studies	9	11	13	14	16	12
Physical Science	Accounting	7	9	8	11	13	9
Physical Science	History	7	9	9	9	6	8
Physical Science	Geography	7	7	7	4	4	6
Physical Science	Agricultural Sciences	5	5	3	2	0	4
Accounting	History	3	1	4	1	1	2
Accounting	Geography	2	2	3	1	2	2
Economics	Geography	2	2	2	1	1	2
Business Studies	Tourism	2	1	2	1	3	2
Economics	History	2	1	1	1	0	1
Economics	Tourism	1	1	1	2	2	1
Agricultural Sciences	Business Studies	2	1	1	1	0	1
Accounting	Tourism	1	1	2	1	3	1
Agricultural Sciences	History	2	1	0	1	0	1
Life Sciences	Economics	1	1	1	1	2	1
Agricultural Sciences	Economics	2	1	1	1	0	1
Geography	History	1	1	1	1	2	1
Economics	Consumer Studies	1	0	1	1	2	1
Economics	Life Sciences	1	2	1	0	0	1
Other switches		17	20	15	21	25	19
Total of the above		100	100	100	100	100	100
Total		486	360	507	392	126	1,871

Note: Total values refer to the number of subject switches which were found in the analysis to assure an overall Grade 12 pass for a student who had not passed. All other values are percentages of the total value within a particular column. The table breaks down the 20 most common switches, sorted from most to less common (in the overall column). The total of 1,871 subject switches are distributed across 725 learners, meaning that on average there were 2.6 possible beneficial subject switches per learner.

The above table sums up what the most common workable subject-to-subject switches were. The mathematics to mathematical literacy switch accounted for around a quarter of workable switches. Moreover, this was a workable switch for 444 or 61% of the 725 learners who, according to the analysis, moved from having an overall fail to an overall pass. Importantly, this does not necessarily mean that 61% of the learners in question should switch from mathematics to mathematical literacy. Many of these learners could instead make another switch, which would also secure the learner a Grade 12 pass. A full 39% of the successful switches involved a switch from physical science to another subject. Here it is important to bear in mind that the 2009 situation is said to have been exceptional due to a particularly difficult physical science examination paper, which caused the physical science pass rate to decline from 55% to 37% between 2008 and 2009[16]. The need for an exodus from physical science would presumably not appear as strong if data from a year other than 2009 were used. One may ask how it is possible for a switch from life sciences to economics and from economics to life sciences to *both* be beneficial (see Table 10). The explanation lies in the fact that better economics results may appear in conjunction with certain other non-language subjects and that better life sciences results may be associated with *different* combinations of other subjects. Put differently, some learners may be better suited for economics, whilst others may be better suited for life sciences. The reader is referred to the explanation, in Appendix G, of how ordinary least squares regression models were used to provide estimates of scores learners could be expected to obtain in subjects that they did in reality not take.

[16] See DBE (2010a: 109). Importantly, statistics on p. 55 of the report should not be used as these are erroneous (this was confirmed with the DBE analysts concerned).

The findings of the analysis are so significant from a policy perspective that they call for some interrogation. Firstly, what reasons may there be for considering the 15 percentage point improvement in the overall pass rate referred to above as an *over*-estimate?

- **Learners are left with insufficient time to deal with all subjects.** It seems very possible that there is a general pattern whereby learners have insufficient time to study adequately for all subjects and end up neglecting one or two subjects. In that case, replacing a neglected subject with another subject may not solve the problem as without enough time the learner would continue to neglect, and not pass, one or two subjects. The analysis that is described above carries the implicit assumption that learners do the best they can in all the subjects they take and that the principal barrier is the ability of learners, not an insufficient level of effort (and an unequal spread of the effort across subjects).

- **It is not possible in reality to predict with certainty which combinations are optimal.** Clearly it is impossible to know beforehand, in the case of learners who are struggling, the subjects that will optimise the probability of obtaining one's Matric. However, more certainty could be introduced if there were better systems of assessment of subjects before Grade 12, in particular within Grade 9 given that the choice of FET subjects must be made upon entering Grade 10.

- **The 2009 physical science problems distort the picture.** Given the exceptional situation with respect to physical science in 2009, the 15% improvement described here includes a degree of compensation for this situation to deal with learners who failed to obtain the NSC because of their physical science results. For this reason it would be useful to repeat the analysis using examination data from a less problematic year.

- **Timetabling restraints within schools limit switching opportunities.** Whilst the analysis avoids switches towards unusual subject combinations (see Appendix G for the details), it is likely that many timetabling constraints, which to some extent would be specific to individual schools, have not been respected within this analysis.

What reasons may exist for considering the 15% an *under*-estimate?

- **In reality switches to unusual combinations are possible.** Clearly some switches to unusual subject combinations which are not considered by the analysis, would in fact be possible. In particular, the analysis under-estimates the possibility of shifting towards subjects with low levels of overall enrolment such as design, even if these subjects are offered in the learner's school.

- **In reality learners may not be limited to just the subjects in their school.** The analysis assumes that the availability of subjects per school remains as it is. Clearly new opportunities can be created if more subjects are introduced into certain schools.

- **In reality learners may switch more than one non-language subject.** Importantly, the analysis is limited to finding solutions in the form of one subject switch at a time. Thus a learner who needs six subjects altogether with a score of at least 30 according to the examination rules, and has only four subjects at this level, will not even be considered by the computer program as it does not simulate switching two subjects at once. The computer program will, however, consider learners with only five subjects with a score of at least 30 as in such a case switching one subject can result in an overall pass.

- **Language switching may widen the feasible options.** The analysis considers only non-language subject switches. Switches to or from a non-official language such as French are not considered and nor is a switch from, say, isiZulu home language to Sesotho home language.

Without further analysis of the dataset itself and of information from beyond the dataset, it is impossible to assess overall whether the analysis described above results in an inflated scope for improvement or not. However, it seems highly likely that the scope for improving performance in the Grade 12 examinations through better subject choices in the grades preceding Grade 12 is large. The matter is of course not a mechanistic one in the sense that optimising NSC passes is the only objective at play. There would be a strong aspirational element in the matter of subject choice, at the individual level but also at the system level. Individual learners may not wish to 'downgrade' their subject choices to reduce their risks of not obtaining the NSC, especially if changes would exclude certain higher education options. Even where a particular choice of subjects poses a high risk of overall failure a learner may be unwilling to change to a different combination due to high aspirations. To some degree this problem can be addressed through better assessment practices throughout the secondary school grades which provide a clearer picture to the learner of where his or her strengths and weaknesses lie. At the system level, it may seem politically and ideologically unacceptable to encourage learners from poorer areas, whose probability of obtaining the NSC is on average lower due to the negative impacts of socio-economic background, to switch to easier subjects, for instance mathematical literacy instead of mathematics. At the same time, as the analysis above has shown, poorer schools enjoy a narrower set of subject options and are thus forced to limit their choices. This is clearly unjust. To some extent the injustice would lie in the denial of opportunities for poorer learners to take subjects, such as consumer studies, which could improve their overall Grade 12 results (according to Table 10 consumer studies is a subject towards which some switching should occur and yet according to Table 8 this subject is disproportionally unavailable to poorer learners).

Two policy solutions seem to stand out to deal with the problem of poor subject choice. As has been mentioned above, better assessment and, linked to this, subject and career counselling, would assist learners to make appropriate and realistic subject choices. Another solution is to use analyses such as the one described here to identify where exactly the subject choice problems lie and for provincial departments to advise schools on how to advise their learners. One extension of the analysis that could be undertaken fairly easily is to identify individual schools that are particularly poor at guiding learners towards the right subject combinations.

7 Conclusion

This conclusion focuses on the policy implications of the preceding analyses. The four policy rubrics introduced in section 2 are used.

Firstly, what does the evidence presented in this paper suggest about ways of improving the general quality of education at the secondary level? To a fairly large extent the evidence adds weight to policy priorities that are already strong, for instance in the Minister of Basic Education's delivery agreement with the President. The evidence thus confirms the correctness of these priorities and the importance of allocating sufficient effort and resources to them. The fact that grade-on-grade academic improvements at the secondary level are found to be below those in other similar countries and below what is achieved through a year of post-school education confirms the need and the feasibility of improving learning and teaching across all secondary grades. One way of achieving this is through better access to learning materials amongst learners. The finding that households consider a lack of books to surpass other problems in schools, is very telling. This finding both confirms that access to learning materials such as textbooks has indeed been under-prioritised in the past and that the current shift in policy towards solving this gap is correct. The data also confirm the fundamental importance of getting learners to become proficient in reading and writing in English, especially from an employment and earnings perspective. The fact that only around half of learner absenteeism should be attributable to poor health, with a further 10% attributable to a shortage of money, points towards considerable scope for increasing learning time in schools through, for instance, advocacy campaigns aimed at parents. It is at the same time important to gain a clearer picture of why learners absent themselves from school, beyond what the current data allow us.

Secondly, how and at what speed should secondary level school enrolments be increased? Previous findings that secondary level enrolments in South Africa are relatively high by international standards, are confirmed. This pattern holds even more true in 2009 than it did in 2003 insofar as there has been substantial growth in enrolments in Grades 9 to 11. However if one focuses specifically on Grade 12 enrolments and the percentage of youths obtaining the Matric, then South Africa does not fare too well. South Africa's level of enrolment in the last year of upper secondary schooling is more or less in line with what is found in similar developing countries. In other words, it is not exceptionally high. Moreover, the percentage of youths *successfully* completing secondary schooling, in other words obtaining the Matric in the case of South Africa, is a bit low by international standards. This statistic has stood at around 40% for several years. It could easily be argued that the figure should reach around 50% in the near future. The interest currently in getting more learners to complete Grade 12 successfully has been strong and this has manifested itself largely in substantial growth in the grades preceding Grade 12. However, it appears as if insufficient readiness amongst pre-Grade 12 learners for the Matric examination has brought about substantial grade repetition in Grades 10 and 11 as well as high levels of dropping out of these grades, but hardly any change to the Matric numbers. The international comparison suggests that a key and immediate focus ought to be on ensuring that of the youths who currently get to Grade 12 (around 60% of youths) a larger proportion get to pass the Grade 12 examinations (currently only around two-thirds of enrolled Grade 12 learners do). An argument can be made that more Grades 10 and 11 learners should make it through to Grade 12, yet the evidence suggests that this goal should be secondary to the primary goal of increasing the numbers of youths obtaining the Matric by around 25% over current levels. Put differently, if increasing enrolments in Grade 12 barely makes a difference to the number of successful Grade 12 graduates, for instance because class size increases make teaching and learning conditions more difficult, then very little would have been gained from the change. An obvious question is what the equity implications are of an approach of largely focussing on improving graduation levels within the existing enrolment numbers for some years. The figures suggest that the approach is a clearly pro-poor one insofar as Grade 12 enrolments are fairly evenly spread across the socio-economic quintiles (the poor are not under-enrolled in Grade 12) yet Grade 12 pass rates differ markedly, from 47% in the poorest quintile to 84% in the least poor quintile (see Table 9). The policy challenge is thus to pay special attention to the factors that inhibit quality secondary schooling amongst the poorest. From a budgetary perspective, one argument against very large enrolment increases in secondary schools in the near future is that the international evidence points rather clearly to the need for substantial growth in *post*-school educational enrolments, both in the vocational training and university sectors. As argued above, the gap in enrolments above the school level could be as high as 300,000. Closing this gap is clearly costly. In such a context a large and simultaneous expansion in the secondary schooling sector may not be justified or affordable. Importantly, the recommendations made here with respect to secondary school enrolments and graduation rates relate largely to the medium term. In the long term, higher levels of upper secondary school completion should be aimed for, though in this regard it is useful to bear in mind that even amongst rich countries completing twelve years of education is far from universal.

Thirdly, what does the evidence suggest about possible changes to the systems that offer youths educational qualifications? It is significant that, as explained above, around 60% of youths end up with no national or widely recognised educational qualification, despite spending a relatively high number of years in education (according to the international comparisons). The centrality of the Matric is also significant. Only around 1% of youths find themselves in the situation where they do not hold a Matric but do hold another widely recognised qualification. The advantage with educational qualifications is that they empower youths when attempting to enter employment or a post-school educational institution. But beyond this benefit, the examinations and assessments underlying educational qualifications assist in promoting accountability and a focus on standards within the education system. An obvious question is what the costs and benefits would be of introducing one or more national qualifications in the schooling system below the Grade 12 level in order to improve the focus on educational quality in earlier grades but also to improve the 'educational currency' of school-leavers who do not obtain the Matric. A renewed focus on the Grade 9 General Education Certificate, an issue which has remained on the policy 'back-burner' for many years, seems justified. In fact, the current introduction of national assessments in Grades 3, 6 and 9 represent an important step towards greater accountability in Grade 9, even if currently it is not envisaged that the new Grade 9 assessments will result in a national qualification. The evidence presented in the paper clearly points to the need for better assessment systems in Grade 9. Grade-on-grade improvements throughout secondary schooling are low and inadequate to prepare learners for the relatively demanding Grade 12 examinations, resulting in high levels of grade repetition in Grades 10 and 11 as it becomes clear that academic progress has been insufficient. Apart from sharpening the focus on educational standards at an earlier point, better assessment systems in Grade 9 could assist learners in making the right subject choices for Grades 10 to 12. The evidence from the analysis of non-language subject combinations and a learner's chances of passing Grade 12 suggest strongly that learners are currently not well informed when it comes to knowing in which subjects they will succeed. Expanding the FET college sector is currently high on the list of government's priorities. A part of this challenge rests with schools, which need to advise learners in Grades 9, 10, 11 and 12 on whether to continue at school, or enrol at a college. Better Grade 9 assessments and a Grade 9 qualification would facilitate this process.

Fourthly, <u>how can schools provide a more vocationally oriented education for youths</u>? The current policy emphasis on improving access to vocational education and training amongst youths has largely focussed on the role of FET colleges. However, schools too can contribute towards expanding this type of education. Many of the subjects of the Grades 10 to 12 curriculum, such as engineering graphics and design, electrical technology, information technology and tourism have a relatively strong vocational focus and promote skills that are widely considered to be scarce in the labour market. Yet access to these kinds of subjects is low overall and moreover concentrated within schools serving the non-poor. Improving access to these subjects in historically disadvantaged schools, through the introduction of the necessary facilities and relevant teacher training interventions, appears to be an effective option for improving the post-school opportunities of youths.

References

Argent, J. (2009). *Household income: Report on NIDS Wave 1*. Cape Town: SALDRU. Available from: <http://www.nids.uct.ac.za> [Accessed November 2010].

Branson, N. & Lam, D. (2009). *Education: Analysis of the NIDS Wave 1 dataset.* Available from: <http://www.nids.uct.ac.za> [Accessed February 2010].

Burger, R. & Von Fintel, D. (2009). *Determining the causes of the rising South African unemployment rate: An age, period and generational analysis*. Stellenbosch: University of Stellenbosch. Available from: <http://ideas.repec.org/p/sza/wpaper/wpapers97.html> [Accessed March 2010].

Council on Higher Education (2009). *Higher education monitor: The state of higher education in South Africa.* Pretoria: CHE. Available from: <http://www.che.org.za/documents/d000201/Higher_Education_Monitor_8.pdf> [Accessed February 2010].

Crouch, L. (1991). A simplified linear programming approach to the estimation of enrollment transition rates: Estimating rates with minimal data availability. *Economics of Education Review*, 10(3): 259-269.

Crouch, L. & Vinjevold, P. (2006). *South Africa: Access before quality, and what to do now?* Granada: Universidad de Granada. Available from: <http://www.ugr.es/~recfpro/rev101ART6ing.pdf> [Accessed June 2010].

Dearden, L. (1999). *Qualifications and earnings in Britain: How reliable are conventional OLS estimates of the returns to education?* London: Institute of Fiscal Studies. Available from: <http://www.ifs.org.uk/wps/wp9907.pdf> [Accessed January 2010].

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].

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

Department of Basic Education (2010b). Annual Survey of Schools 2009 (dataset). Pretoria. [Earlier collections were also consulted.]

Department of Education (2009). *2009 School Realities*. Pretoria. Available from: <http://www.education.gov.za/EMIS/default.htm> [Accessed January 2010]. [Earlier editions were also consulted.]

Department of Education (2010). *Education statistics in South Africa 2008*. Pretoria. Available from: <http://www.education.gov.za/EMIS/default.htm> [Accessed April 2010]. [Earlier editions were also consulted.]

Gustafsson, M. & Bartlett, A. (2008). *The labour market significance of the new Grade 9 certificate*. Pretoria: HSRC. Available from: <http://www.hsrc.ac.za/research/output/outputDocuments/5253_Gustafsson_Thelabourmarketsignificance.pdf> [Accessed September 2008].

Gustafsson, M., Van der Berg, S., Shepherd, D. & Burger, C. (2010). *The costs of illiteracy in South Africa*. Stellenbosch: University of Stellenbosch. Available from: <http://ideas.repec.org/s/sza/wpaper.html> [Accessed July 2010].

Hanushek, E.A. & Woessman, L. (2009). *Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation*. Washington: National Bureau of Economic Research. Available from: <http://www.nber.org/papers/w14633.pdf?new_window=1> [Accessed March 2009].

Hanushek, E.A. & Zhang, L. (2009). Quality-consistent estimates of international schooling and skill gradients. *Journal of Human Capital*, 3(2).

HSRC (2005). HSRC Client Survey October 2004 (dataset). Pretoria.

ICF Macro (2010). Demographic and Health Surveys (multiple datasets). Calverton. Available from: <http://www.measuredhs.com> [Accessed July 2010]. [The following DHS datasets were used: Colombia 2005, Egypt 2008, Indonesia 2007, Morocco 2003-04, Peru 2004-08, Philippines 2008, Turkey 2003.]

IEA (2004). TIMSS 2003 (dataset). Chestnut Hill: Boston College. Available from: <http://timss.bc.edu/timss2003.html> [Accessed July 2007].

Keswell, M. & Poswell, L. (2004). Returns to education in South Africa: A retrospective sensitivity analysis of the available evidence. *South African Journal of Economics*, 72(4): 834-860.

Kraak, A. (2008). The education-economy relationship in South Africa, 2001-2005. In Andre Kraak and Karen Press (ed.), *Human resources development review 2008: Education, employment and skills in South Africa*. Pretoria: HSRC: 1-25. Available from: <http://www.hsrcpress.ac.za> [Accessed March 2008].

Lam, D., Ardington, C. & Leibbrandt, M. (2008). *Schooling as a lottery: Racial differences in school advancement in urban South Africa.* Cape Town: SALDRU. Available from: <http://www.saldru.uct.ac.za/papers/wpapers/2008_18.pdf> [Accessed November 2008].

OECD (2010). OECD.StatExtracts (internet-based data query facility). Paris. Available from: <http://stats.oecd.org/index.aspx?queryid=254> [Accessed July 2010].

Panday, S., Makiwane, M., Ranchod, C. & Letsoalo, T. (2009). *Teenage pregnancy in South Africa with a special focus on school-going learners.* Pretoria: HSRC. Available from: <http://www.hsrc.ac.za/Research_Publication-21277.phtml> [Accessed August 2010].

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

Schiefelbein, E. (1975). Repeating: An overlooked problem of Latin America. *Comparative Education Review*, 19(3): 468-487.

Schindler, J. (2008). Public schooling. In Andre Kraak and Karen Press (ed.), *Human resources development review 2008: Education, employment and skills in South Africa*. Pretoria: HSRC: 228-253. Available from: <http://www.hsrcpress.ac.za> [Accessed March 2008].Statistics South Africa (2008). Community Survey 2007 unit records (dataset). Pretoria.

Statistics South Africa (2001). *The youth of South Africa: Selected findings from the Census '96*. Pretoria. Available from: <http://www.statssa.gov.za> [Accessed March 2008].

Statistics South Africa (2009b). Labour Force Survey 2009 (third quarter) (dataset). Pretoria. [Earlier collections in this series were also used.]

Statistics South Africa (2009c). *Quarterly Labour Force Survey: Quarter 3, 2009* (Statistical release P0211). Pretoria. Available from: <http://www.statssa.gov.za> [Accessed January 2010].

Statistics South Africa (2010). General Household Survey 2009 (dataset). Pretoria. [Earlier collections in this series were also used.]

UNESCO: UIS (2010). Education statistics (internet-based data querying facility). Montreal. Available from: <http://www.uis.unesco.org/ev_en.php?URL_ID=3753&URL_DO=DO_TOPIC&URL_SECTION=201>

Van Broekhuizen, H. and D. Von Vintel (2010). *Who responds to voluntary cognitive tests in household surveys? The role of labour market status, respondent confidence, motivation and a culture of learning in South Africa.* Stellenbosch, Stellenbosch University.

Van der Berg, S. (2007). Apartheid's enduring legacy: Inequality in education. *Journal of African Economies*, 16(5): 849-880.

World Bank (2005). *Expanding opportunities and building competencies for young people: A new agenda for secondary education.* Washington. Available from: <http://go.worldbank.org/TU36ANZYW0> [Accessed June 2010].

World Bank (2010). World databank (internet-based data querying facility). Washington. Available from: <http://databank.worldbank.org>

Ziderman, A. (1997). National programmes in technical and vocational education: Economic and education relationships. *Journal of Vocational Education and Training*, 49(3): 351-366. Available from: <http://www.triangle.co.uk/pdf/viewpdf.asp?j=vae&vol=49&issue=3&year=1997&article=Ziderman&id=163.195.192.74> [Accessed Jan 2005].

Appendix A: Youth economic activity data

The following table provides the percentages underlying Figure 2, as well as a total percentage for each activity incorporating all ages in the 15 to 35 range.

Table 11: The activities of youth by age

Age	*<Gr 8*	*Gr 8*	*Gr 9*	*Gr 10*	*Gr 11*	*Gr 12*	*Other*	*Univ*	*Unemployed other*	*Unemployed and discouraged*	*Unemployed and looking*	*Domestic worker*	*Working in own household*	*Employed in agriculture*	*Employed in informal sector*	*Employed in formal sector*
15	18.8	20.2	34.4	17.8	2.0	0.5	1.1	0.0	3.5	0.3	0.1	0.0	0.8	0.0	0.4	0.0
16	10.2	12.0	19.0	29.3	17.7	2.3	1.3	0.0	4.5	1.1	0.9	0.0	1.2	0.2	0.2	0.2
17	5.0	6.3	12.0	22.7	24.4	15.3	1.6	0.4	6.2	1.5	2.0	0.1	1.3	0.1	0.4	0.7
18	2.1	4.4	7.3	15.8	18.4	20.5	3.4	4.2	7.7	3.2	5.9	0.1	4.0	0.3	1.3	1.3
19	1.1	2.0	4.3	11.6	13.6	16.3	5.1	7.2	9.3	4.6	11.3	0.4	5.4	0.6	2.4	4.7
20	0.5	0.8	1.8	6.2	10.7	11.1	5.3	7.2	9.9	7.3	19.8	0.9	6.2	0.8	2.5	8.9
21	0.2	0.2	1.0	3.6	7.1	6.1	6.5	8.7	10.3	9.8	21.6	0.8	6.7	1.3	4.0	12.1
22	0.3	0.4	0.5	1.6	4.0	4.1	3.7	7.6	10.4	9.5	21.5	1.2	9.4	1.6	4.6	19.5
23	0.2	0.0	0.3	1.1	1.9	2.9	3.5	6.4	10.3	9.2	24.1	1.9	8.6	2.2	4.5	22.7
24	0.1	0.1	0.0	0.6	0.9	1.2	2.3	4.4	9.6	9.7	23.4	1.7	9.6	2.2	6.5	27.8
25	0.2	0.1	0.0	0.1	0.2	0.6	2.1	2.9	9.6	8.6	22.3	2.3	8.1	2.8	7.4	32.6
26	0.3	0.0	0.0	0.1	0.0	0.3	1.5	2.9	10.5	8.8	24.0	2.6	8.5	1.5	7.4	31.5
27	0.0	0.1	0.0	0.1	0.2	0.0	1.4	1.6	9.3	9.1	23.6	3.0	7.6	1.7	8.0	34.3
28	0.0	0.0	0.0	0.1	0.0	0.3	0.7	1.9	7.5	7.5	20.4	4.1	8.6	2.9	7.1	38.7
29	0.0	0.0	0.0	0.0	0.0	0.1	0.9	0.9	9.1	6.1	21.3	2.8	7.8	2.1	7.3	41.6
30	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.4	9.6	6.2	22.1	3.4	10.1	1.8	7.9	38.2
31	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.3	9.6	5.7	17.8	3.4	8.9	3.3	9.4	41.0
32	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.3	7.5	6.7	19.7	4.7	6.9	3.4	8.1	42.2
33	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.2	8.8	5.2	19.1	4.9	9.8	1.8	10.2	39.8
34	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.2	8.8	5.8	17.5	4.7	8.7	3.3	8.3	42.3
35	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.3	10.2	4.4	17.1	6.1	9.2	4.2	7.1	41.3
15-35	2.1	2.5	4.4	6.0	5.5	4.4	2.2	3.0	8.6	6.1	16.5	2.1	6.8	1.7	5.1	23.0

Appendix B: What percentage of youths matriculate?

Four approaches are presented here for estimating this important percentage for 2009 (or in one approach 2008). The results of the analysis appear in Table 12 below.

Table 12: Percentage of youths matriculating

Data source	*Matric-ulants*	*Size of age cohort*	*Percent-age*	*Comment*
GHS 2009	468,237	970,886	48.2	Age 25 yielded maximum.
NIDS 2008	337,940	834,693	40.5	Age 21 yielded maximum.
GHS 2009 + NSC 2009	354,673	1,063,936	33.3	Average for ages 17 to 19 used.
GHS 2009 + NSC 2009 + EMIS 2009	354,673	900,069	39.4	Average for ages 17 to 19 used.

- In the first row, just General Household Survey (GHS) data are used. The 48.2% figure is what was illustrated in Figure 4 for 2009.

- The second row is based on just NIDS 2008 data. The NIDS questionnaire is arguably stronger than the GHS in insisting that only successful completion of Grade 12, and not unsuccessful participation in that grade, should be counted.

- The third row represents an approach that may be tempting, but is undeniably wrong. Here the number of National Senior Certificate passes (*after* successes in the supplementary examinations early in 2010 had been taken into account) is divided by the average age cohort size in the population in age range 17 to 19. The resulting 33.3% can be regarded as inaccurately low given that a number of analyses (including this one) point to the fact that the population of youths is over-estimated in the GHS (and in Stats SA's mid-year population estimates, which are close in magnitude to the weights in the GHS).

- The fourth row presents what can be considered the most precise estimate. Here number of NSC passes is divided by the average age 17 to 19 age cohort size adjusted downward. The downward adjustment is in proportion to the relationship between, firstly, the official EMIS enrolments (public and independent) in 2009 in Grades 8 to 10 and, secondly, the number of Grades 8 to 10 learners estimated in the 2009 GHS (which was the first GHS to ask in which grade youths are *currently* enrolled in). Why should this approach be considered the most accurate? It is based on the most precise values we have. It is based on figures from the NSC dataset, which is undoubtedly the most reliable data source of all given the high stakes operational nature of the dataset. It is based on EMIS values which, whilst subject to problems such as distortions by school principals in the survey form, are widely considered to be relatively reliable. Lastly it is based on just the Grades 8 to 10 enrolment values in the GHS, so the problems associated with Grade 12 enrolment values (see Appendix C) are kept out of the calculation.

Of course the first two approaches are measuring something different to the last two approaches. The first two examine the age cohort at which one obtains the highest percentage of successful Grade 12 graduates. The last two examine graduations in one year divided by an age cohort. However, if all the data were true reflections of reality and the number of NSC passes per year were more or less stable, then all four approaches should yield very similar results. The number of passes has in fact been increasing over the years to some degree. In 2002 the 300,000 mark was passed (Van der Berg, 2007: 853). One would therefore expect *lower* values using the approach in the first (or second row), as opposed to the approach in the fourth row. Yet when the GHS source is used, one obtains a substantially *higher* value. This seems to confirm that respondents in the GHS are over-stating their Grade 12 attainment.

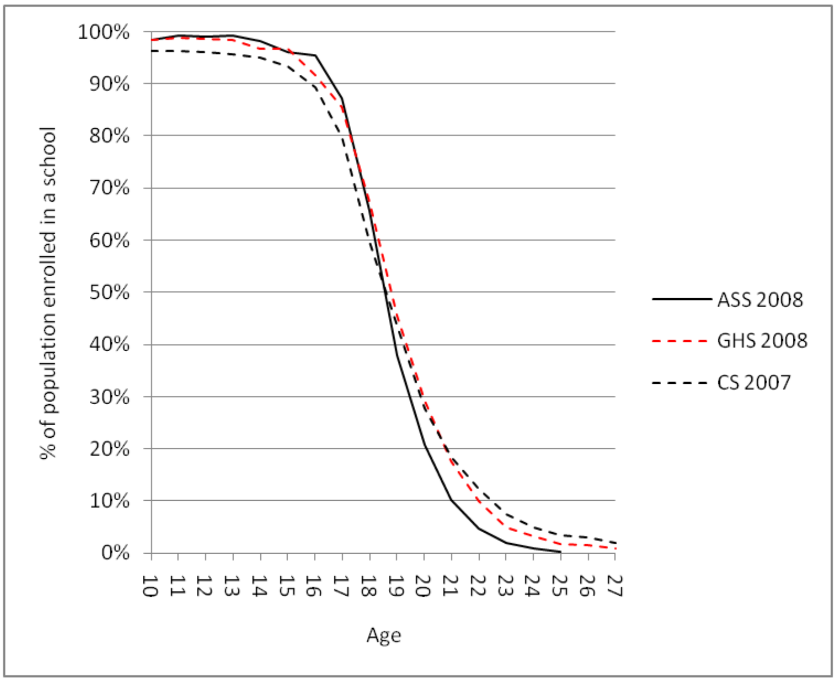
Could the higher GHS values be the result of having passed non-public examinations? This seems unlikely. The non-public Independent Examinations Board (IEB) graduates only around 10,000 learners per year, a figure that can account for only 1.1 percentage points of the difference between the 48.2% and 39.4% values.

Appendix C: The question of Grade 12 enrolments in the household surveys

There are important differences between official secondary school enrolment statistics based on school surveys and what household surveys reveal. The analysis that follows describes these differences and discusses what the likely explanations may be.

The next graph illustrates that for ages 18 and above household surveys suggest there is more enrolment in schools than what is suggested by the Annual Survey of the Department of Basic Education, which collects data from schools. The Annual Survey of Schools (ASS) curve in Figure 21 uses as an anchor the fact that 0.98 of the population aged 10 in 2008 was enrolled in a school. This fact is obtained from the General Household Survey (GHS). Above age 10, population values from the GHS are adjusted downwards in proportion to the age 10 adjustment to produce the ASS curve in the graph. One important matter that this analysis does not deal with is the fact that Statistics South Africa enrolment and population values for youths, which since the 2001 census have been based on estimated weights attached to individuals in sample surveys, appear to be over-estimates if one assumes that official enrolment figures (which unlike Stats SA values, are based on annual census collections) are more or less correct. The extent of the Stats SA over-estimation appears to be around 5%. The GHS and Community Survey curves in the graph are based on the percentage of the population at each age responding that they are enrolled in a school. Stats SA surveys differentiate adult basic education institutions from schools, and only the latter was considered in the two household-based curves.

Figure 21: School enrolment by age in three data sources



Source: Annual Survey of Schools 2008, General Household Survey 2008, Community Survey 2007.

The gap between the GHS 2008 and the ASS 2008 curves in the age range 18 to 27 comes to 0.40 of a population cohort, or around 350,000 people. The data thus seems to suggest that the schools survey leaves out this number of school learners, specifically older learners, and that these learners could be enrolled in non-public and relatively informal institutions that the official schools survey does not reach. One could argue that the discrepancies between the curves in Figure 21 are mostly due to measurement errors in the household-based values. It would be tempting to accept this argument were it not for the fact that the grade-specific analysis that follows seems to confirm that there is indeed a substantial number of learners not covered in the schools surveys.

The next table provides grade-specific enrolment values in the Grades 8 to 12 range from a variety of sources. The table includes household statistics derived from a question on the current grade of learners (the ‘direct approach’) as well as grade-specific enrolment statistics imputed from two questions, the one being whether someone is in school and the other being what the highest grade successfully completed is (the ‘indirect approach’). To illustrate how the indirect approach works, if someone says he is enrolled in school and that the highest grade successfully completed was Grade 10, it is assumed that the learner is currently in Grade 11.

If one examines the values calculated using the indirect approach and compares them to the official enrolment statistics, then what is striking, apart from the fact that household survey values are all higher, is that enrolment in Grade 12, relative to enrolment in the other grades, is considerably higher when the household data are used. Specifically, the ratio of Grade 12 learners to the average in Grades 8 to 10 is consistently 0.61 for the period 2008-2009 in the official statistics, but always at least 0.70 in the household data (from 2006 to 2009). (The last column of the table indicates the ratio.) This is a considerable difference that translates into almost 100,000 more Grade 12 learners in the household data.

Table 13: Grades 8 to 12 enrolments (various sources)

	Gr 8	*Gr 9*	*Gr 10*	*Gr 11*	*Gr 12*	*Total*	*Rat.*
THE DIRECT APPROACH							
2009: Official statistics based on the Snap Survey							
Public schools	957,574	897,345	987,680	851,006	568,995	4,262,600	0.60
Independent schools	32,035	28,072	28,680	29,509	30,631	148,927	1.03
Total	989,609	925,417	1,016,360	880,515	599,626	4,411,527	0.61
% public	97	97	97	97	95	97	
2009: Public NSC candidates							
					620,192		
2009: General Household Survey							
Public schools	1,085,019	993,786	1,075,536	927,537	735,324	4,817,203	0.70
Independent schools	76,247	65,100	64,177	71,667	61,851	339,042	0.90
Other	14,966	14,077	7,923	10,258	5,089	52,312	0.41
Total	1,176,232	1,072,963	1,147,636	1,009,462	802,264	5,208,556	0.71
% public	93	94	94	93	92	93	
2008: Official statistics based on the Snap Survey							
Public schools	899,097	877,143	1,047,874	873,125	566,460	4,263,699	0.60
Independent schools	27,506	25,513	28,653	29,627	28,756	140,055	1.06
Total	926,603	902,656	1,076,527	902,752	595,216	4,403,754	0.61
2008: Totals from the Annual Survey of Schools							
Public schools	893,563	866,815	1,035,600	861,588	558,889	4,216,455	0.60
Independent schools	27,970	25,894	30,467	31,478	29,176	144,985	1.04
Total	921,533	892,709	1,066,067	893,066	588,065	4,361,440	0.61
2008: Public NSC candidates							
					589,759		
2008: 2008 National Income Dynamics Study (NIDS)							
Total	995,865	942,285	1,029,805	910,740	887,828	4,766,523	0.90
2004: Official statistics based on the Snap Survey							
Public schools	985,132	891,930	1,034,145	806,554	480,646	4,198,407	0.50
Independent schools	25,578	22,799	23,790	22,583	24,746	119,496	1.03
Total	1,010,710	914,729	1,057,935	829,137	505,392	4,317,903	0.51
2004: 2004 HSRC household survey							
Total	1,069,585	904,037	1,112,354	945,024	935,761	4,966,761	0.91
THE INDIRECT APPROACH							
2009: General Household Survey							
Total	1,151,712	1,053,579	1,118,400	1,009,226	780,804	5,113,721	0.70
2008: General Household Survey							
Total	1,071,707	1,070,836	1,165,757	1,020,631	852,987	5,181,917	0.77
2007: General Household Survey							
Total	1,038,299	1,063,620	1,182,533	1,004,239	898,253	5,186,943	0.82
2007: Community Survey							
Public schools	997,838	1,249,657	1,212,083	1,100,845	1,224,766	5,785,189	1.06
Independent schools	38,160	42,557	43,781	48,572	81,275	254,344	1.96
Total	1,037,093	1,294,257	1,257,899	1,151,630	1,309,278	6,050,155	1.09
2006: General Household Survey							
Total	1,078,640	1,062,355	1,138,128	1,020,482	825,973	5,125,577	0.76
2005: General Household Survey							
Total	1,189,185	1,055,803	1,114,680	1,019,404	769,028	5,148,101	0.69
2004: General Household Survey							
Total	1,127,295	1,002,186	1,099,106	1,026,946	738,671	4,994,203	0.69

Sources: The various datasets mentioned as well as: DBE, 2010a; DoE, 2009; DoE, 2010.

Up to 2008 a plausible explanation seemed to errors arising from the use of the indirect method when household data were analysed. However, from 2009, the GHS includes a question asking what the *current* grade of enrolment is, making it possible to obtain enrolment statistics using the direct approach. The 2009 GHS values in Table 13 suggest that the direct and indirect approaches do not yield very different results and strengthen the argument that there is an inconsistency in the Grade 12 enrolment figures (between the household and official enrolment sources) which warrants serious attention. The 2009 GHS in fact confirms findings from two previous but smaller household surveys, namely the 2004

HSRC survey and the 2008 NIDS survey, that enrolment in Grade 12 appears higher (compared to other grades) than what the official enrolment statistics suggest.

Figure 21 suggests that the learners captured in the household data and not in the official enrolment statistics are older learners. One may be dealing with learners enrolled in informal institutions not covered by the surveys of the DBE. However, the 2009 GHS seems to contradict such a hypothesis. According to the GHS, 99% of those who reported being in a grade in the Grade 8 to 12 range said they were either in a public or private school. None reported being in an ABET centre (some respondents did say they were in an ABET centre, but none said this was in any of the five grades). Moreover, only 0.2% of learners in Grades 8 to 12 reported doing their secondary schooling via correspondence.

What is noticeable is that the percentage of learners in public schools is considerably lower in the household data than in the official data. However, it is believed that this simply reflects the fact that many believe that public schools with high fees ('ex-Model C schools') are private when in fact they are public. In any event, the 2009 GHS does not reflect an exceptional deviation for Grade 12 when it comes to the proportion of learners in independent schools, which contradicts the possibility that high enrolments in informal private independent schools focussing specifically on Grade 12 explain the enrolment discrepancies (to some extent the 2007 Community Survey data suggest that this might be happening though the magnitude is relatively small). Importantly, if one focuses on just public schools, there is a large difference with respect to Grade 12: the ratio in the final column is 0.70 using the 2009 GHS against 0.60 using the official statistics. One possible explanation is that many Grade 12 learners enter the year after the DBE surveys have been run but before the GHS occurs in around July. However, if there are indeed around 96,000 Grade 12 learners not accounted for in the official statistics (this is the figure one arrives at if one adjusts the official statistics upwards in line with the 2009 GHS values) then this is not reflected in the number of candidates registered for the national examinations. The latter values are closer to the official statistics than a figure inflated by 96,000 learners.

An explanation does not seem possible without more information. Yet the discrepancies discussed above are so large that they warrant further attention as new datasets become available.

Notably, the ratio of Grade 11 enrolments to enrolments in lower grades are not that different in the official enrolment statistics and the household data. The mystery is clearly a Grade 12 one.

Appendix D: In pursuit of reliable grade repetition values

The NIDS dataset allows for a reconstruction of grade- and age-specific enrolments during years preceding the survey using a number of historical questions included in the questionnaires. The adult questionnaire (directed to those aged 15 and above) asks, firstly, in which year the respondent completed his highest grade (in the case of those who are not currently in school) and, secondly, in which year the respondent started Grade 1. Moreover, all adult respondents are asked what grades they repeated, and how many times they repeated those grades, though only those aged 15 to 30 were asked to respond to this question. These questions mean that it is possible to determine the enrolment history, in terms of grade-specific enrolment by year going back to Grade 1, for those aged 15 to 30 on the survey date.

To provide a more complete history of enrolments, data collected through the NIDS child questionnaire were also used. Through this questionnaire, the year in which Grade 1 was started and the history of grade repetition are collected.

Table 14 indicates that of the 12,545 respondents who should ideally have the data needed to calculation their enrolment history, 11,393 of them, or 91%, did. For virtually all respondents with data it was possible to calculate 'top-down' values of grade-specific enrolment per year, meaning values starting from the current year or most recent year of enrolment and working backwards. For around 60% of respondents it was moreover possible to calculate 'bottom-up' values, or values starting from the year in which the respondent was in Grade 1 and working forwards. In the case of 3,777 respondents the Grade 1 starting year, the grade repetition responses and the most recent year of school enrolment (or the current year, if the respondent was currently enrolled) agreed with each other. In other words, the top-down and bottom-up figures agreed with each other. For others, two different enrolment histories were obtained, depending on whether one started with the last year of enrolment or Grade 1. Arguably, the misalignments between the two indicated in Table 14 are not so extensive that they preclude useful analysis.

Table 14: Observations with enrolment and repetition details

	Adults	Children	Total
Expected to have enrolment history	7,069	5,476	12,545
Have some enrolment history	5,992	5,401	11,393
Have top-down figures	5,943	5,401	11,344
Have bottom-up figures	3,706	3,924	7,630
Have repeated	3,431	1,330	4,761
Misalignment			
Top-down >3 years behind	47	47	94
Top-down 3 years behind	34	19	53
Top-down 2 years behind	86	66	152
Top-down 1 year behind	272	245	517
No misalignment	1,415	2,362	3,777
Top-down 1 year ahead	848	890	1,738
Top-down 2 years ahead	406	214	620
Top-down 3 years ahead	241	57	298
Top-down >3 years ahead	308	24	332

One very likely cause for the misalignments is that grade repetition was under-reported. This would explain why it was more common to find the top-down values ahead of the bottom-up values, and not the other way round. For example, if someone said she started Grade 1 in 1994 and completed Grade 12 in 2006, then she was in school for thirteen years and must have repeated a grade once. If the repeated grade was not included in the dataset, then the bottom-up approach would indicate that the respondent attended school during the years 1994 to 2005 whilst the top-down approach would indicate that the respondent attended school during the years 1995 to 2006. Thus there would be a misalignment in which the top-down approach was one year ahead of the bottom-up approach.

Clearly the misalignments problem makes it especially important to interpret the historical enrolment values with care. The next table provides grade repetition statistics using two approaches. In the first approach, the values obtained through the top-down approach are used, plus the values from the bottom-up approach where there were no top-down values available (because the respondent had not indicated in which year he finished schooling). In the second approach, only respondents where the top-down and bottom-up approaches yielded the same values were used. Whilst the first approach allows for more observations to be used, the second approach can be considered more reliable in the sense that one can be highly certain that the reported grade repetition is accurate (otherwise the top-down and bottom-up approaches would not coincide). It is significant that both approaches produce the same overall pattern of a gradual increase in the overall percentage of learners repeating their grade, from 5% in 1994 to around 8% in 2008. At the secondary level, in the years 2001 to 2008, there was a clear peak in grade repetition in Grade 10 and 11. The percentages in Table 15 are calculated using the NIDS household weights. The number of observations provide an idea of the reliability of the statistics. The fact that there is considerable unevenness in the grade-specific trend is understandable if one considers that the median number of repeaters per cell underlying the grade-specific annual repetition percentages is just 25 learners.

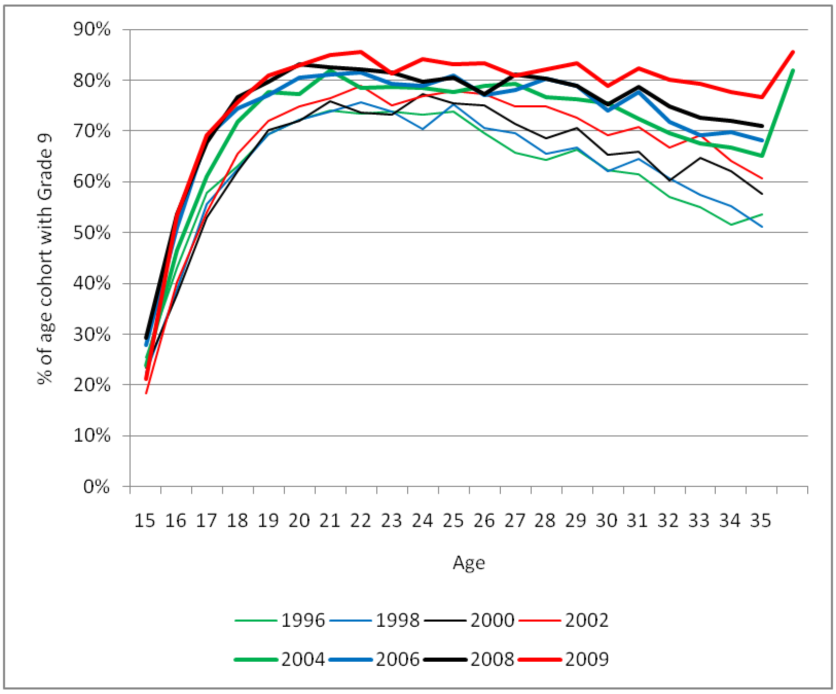
Table 15: Percentage of learners who are repeating 1994-2008

Using both top-down and bottom-up values but preferring the former	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
Gr 1	5.0	6.8	6.8	6.4	9.0	10.3	8.2	7.4	12.6	8.9	12.0	10.2	10.7	10.5	12.3
Gr 2	6.1	1.7	5.0	4.7	3.2	5.1	5.9	9.5	7.6	5.2	5.5	9.6	9.3	9.6	5.7
Gr 3	5.3	8.9	5.3	7.6	7.8	5.3	11.1	8.1	9.7	6.8	9.8	7.2	7.4	7.5	7.7
Gr 4	6.6	4.2	5.6	2.1	4.9	4.4	5.6	6.9	8.5	7.4	5.7	7.6	9.2	8.0	5.0
Gr 5	5.9	5.0	4.1	6.8	5.2	5.0	4.3	3.0	5.5	5.1	7.7	4.2	5.3	6.0	4.5
Gr 6	2.3	1.2	3.6	4.8	2.1	4.6	2.3	3.5	4.9	5.8	5.3	5.9	5.3	3.9	2.3
Gr 7	2.7	4.2	4.1	2.3	5.9	5.1	3.3	3.9	5.2	5.2	4.6	6.1	5.7	4.2	2.8
Gr 8	2.9	5.5	6.7	8.6	13.2	7.3	12.1	8.7	6.8	3.8	5.9	5.5	7.5	7.3	5.7
Gr 9	6.6	3.3	11.7	10.4	5.7	7.5	6.0	8.6	7.5	10.2	7.4	5.7	7.5	5.7	5.4
Gr 10	6.3	7.4	7.8	7.8	6.0	15.4	8.5	10.6	10.4	11.9	19.6	17.7	18.5	19.0	14.9
Gr 11	0.0	4.6	8.7	2.3	10.7	7.7	11.8	14.0	14.7	14.6	10.7	18.0	14.5	18.7	16.9
Gr 12	0.0	0.0	3.3	3.2	13.0	19.4	10.0	6.4	5.3	5.4	3.9	5.6	8.7	11.5	8.8
All grades	4.9	4.8	5.9	5.8	6.5	6.9	6.9	7.0	8.0	7.2	8.2	8.5	9.2	9.2	7.5
Gr 8-12	4.4	5.0	8.2	7.4	9.4	10.2	9.6	9.6	8.8	9.0	9.8	10.6	11.9	12.7	10.4
Obs.	2,880	3,329	3,884	4,417	4,835	5,206	5,536	5,972	6,415	6,831	7,163	7,482	7,594	7,761	7,806
Using only people whose top-down and bottom-up values are the same															
All grades	4.8	6.0	7.0	7.4	8.3	7.2	6.8	7.4	7.3	6.5	8.0	7.5	8.5	8.5	7.6
Gr 8-12	2.7	4.8	7.7	11.5	10.1	11.6	7.4	9.3	6.0	8.5	12.4	9.4	10.6	12.1	9.3
Obs.	580	682	851	1,023	1,163	1,306	1,416	1,653	1,845	2,087	2,338	2,591	2,918	3,324	3,077

Note: These statistics, especially those for less recent years, should be read with much caution. See the accompanying discussion.

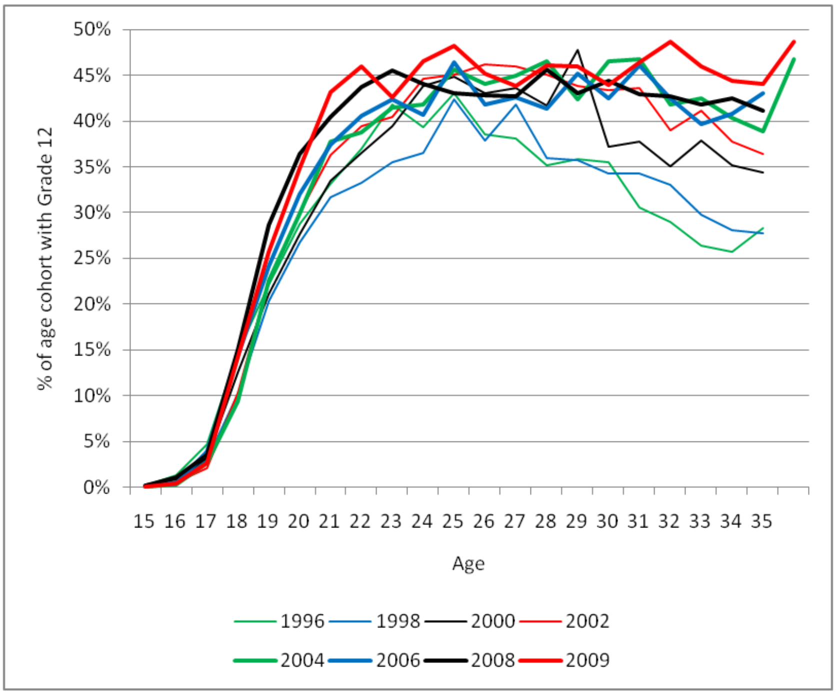
The statistics in Table 15 need to be treated with much caution. Above all, they appear not be consistent with the tendency of learners to attain specific grades at younger ages observed in Figure 4. Could it be that the trend seen in Figure 4 is a result of having used the two specific years 2004 and 2009? The following graphs, which examine the trend with respect to Grade 12 using data from eight different years in the period 1996 to 2009, indicate that this was not the case, and that the trend persists. The first two graphs confirm that in general the attainment by age curves tended to move up (indicating greater attainment) and to the left (indicating younger attainment). In Figure 24, the attainment points reflect the maximum readings from the previous two graphs for Grades 9 and 12. The trend is clearly an upward one for both grades, though this is more pronounced in the case of Grade 9, where attainment rose by more than 10 percentage points. The age points in this graph indicate the maximum age that must be included for a 70% Grade 9 attainment to be observed and for a 35% Grade 12 attainment to be observed. To illustrate, in 1996 at least 70% of 20 year olds had attained Grade 9, but for 19 year olds the percentage was less than 70%. By 2009, the age 20 statistic had become 18, in other words there had been an improvement of two years. The improvement was around just half of this, in other words one year, at the Grade 12 level. Of course the age statistics described here are not the same as the average age of those leaving school at a particular level. Estimates of the latter for Grade 12 were obtained using the patterns from Figure 23. The average across all the eight years was found to be age 21.2, and no significant upward or downward trend over time was observed. How is this reconciled with the age decline referred to earlier? The answer is that whilst the age of completion for a constant percentage of youths (35%) dropped, at the same time attainment was rising increasingly beyond this 35% level through better participation of previously excluded segments of society. Youths from these segments would tend to complete Grade 12 at a higher age, the net effect being no change in the average age of youths completing Grade 12.

Figure 22: Grade 9 grade attainment by age curves



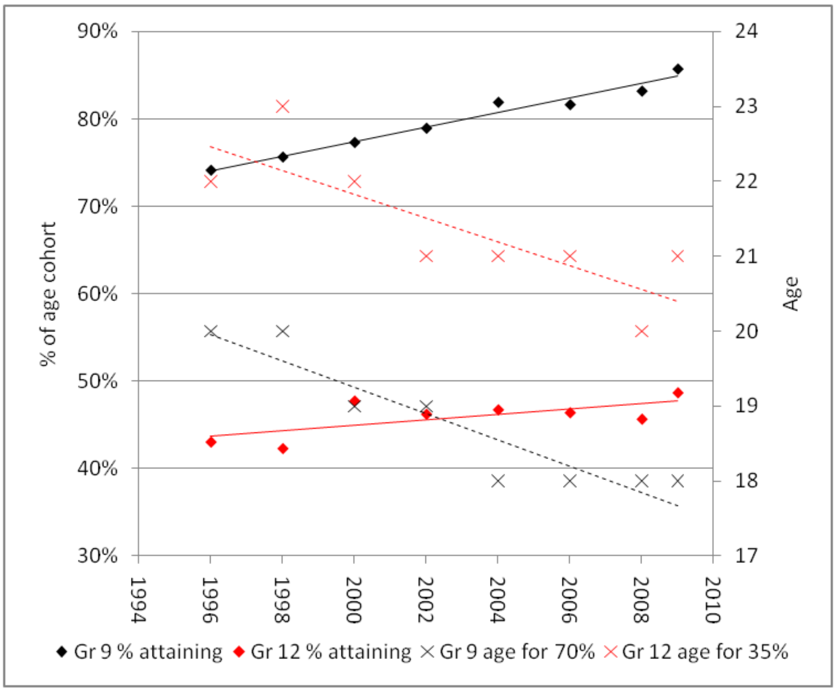
Source: General Household Survey for 2004 to 2009; Labour Force Survey (September) for 2000 to 2002; October Household Survey for 1996 to 1998.

Figure 23: Grade 12 grade attainment by age curves



Source: Figure 22.

Figure 24: Age and attainment trends across years

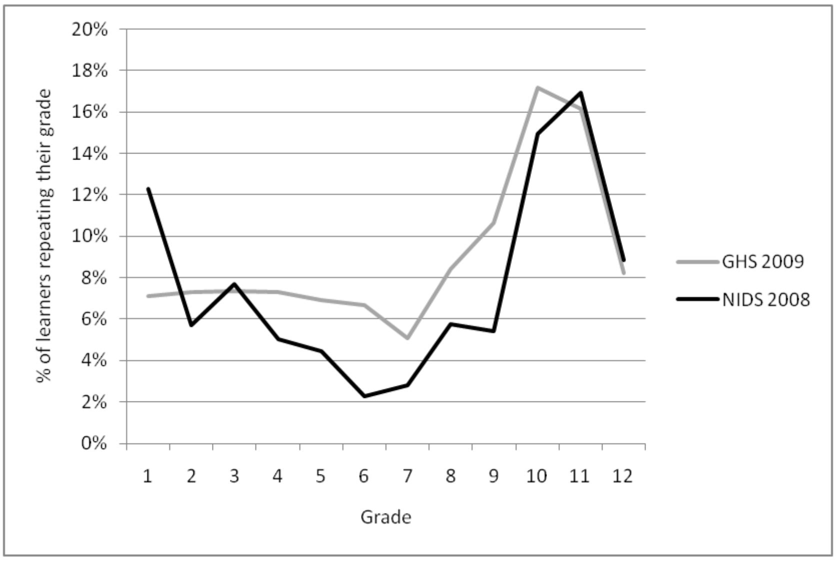


Source: Figure 22.
Note: The dotted lines and their points should be read against the right-hand axis.

The above graphs thus strongly suggest that in the period 1996 to 2009 there was a reduction in grade repetition. There seem to be two other possible explanations. One is that there was a reduction in the phenomenon of staying away from school for a year or so and then returning. This would have an effect similar to grade repetition. The other is that learners started school in Grade 1 at an earlier age. Education planners in South Africa would tend to support the hypothesis that grade repetition declined, though it has been difficult to test this hypothesis due to problems with the grade repetition data collected from schools. School principals are often under pressure to reduce grade repetition, meaning schools are likely to under-report repetition. Moreover, because policy pressures change over the years, the degree of under-reporting is likely to change. Analysts such as Schiefelbein (1975) and Crouch (1991) have proposed alternative and indirect ways of establishing grade repetition levels. Their models have not been tested using South African data. Work along these lines ought to be pursued in South Africa, especially to improve the understanding of historical grade repetition rates.

There are two household data sources other than NIDS which provide grade repetition figures for South Africa. The one is a 2004 collection based on a nationally representative sample of 5,600 households. This collection was commissioned by the Department of Education. The other is the 2009 General Household Survey where, for the first time in the GHS, a question was included on whether the learner is repeating the current grade (it seems as if this question will become a permanent feature of the GHS). How do the NIDS data on grade repetition compare with these two other sources? The following graph compares NIDS and GHS.

Figure 25: Grade repetition in NIDS and GHS



Source: National Income Dynamics Study 2008; General Household Survey 2009.

The values used for the above graph appear in the following table.

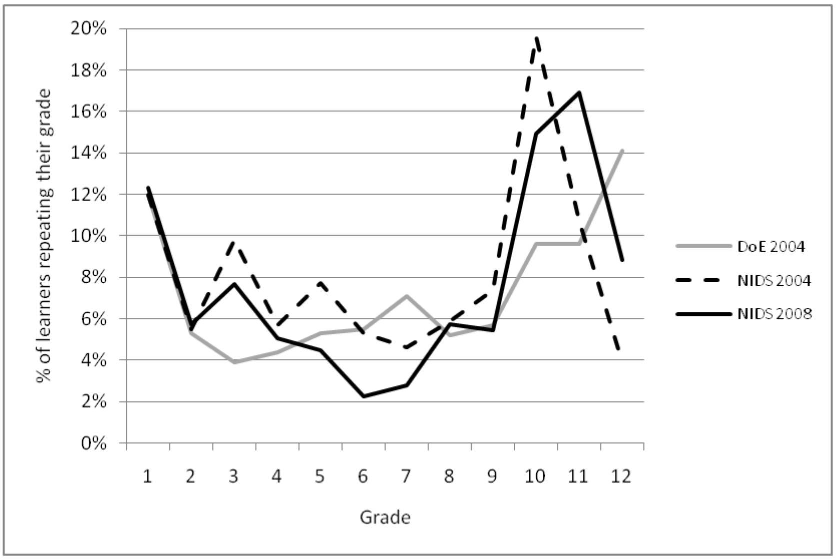
Table 16: Grade repetition in NIDS and GHS

Grade	NIDS 08	GHS 09
1	12.3%	7.1%
2	5.7%	7.3%
3	7.7%	7.4%
4	5.0%	7.3%
5	4.5%	6.9%
6	2.3%	6.7%
7	2.8%	5.1%
8	5.7%	8.4%
9	5.4%	10.6%
10	14.9%	17.2%
11	16.9%	16.2%
12	8.8%	8.2%
8 to 12	10.4%	12.2%
1 to 12	7.6%	8.9%

At the secondary level, the Grades 10, 11 and 12 levels are very similar across the two sources. The Grade 9 level is 5 percentage points higher in the GHS than NIDS and in Grade 8 the difference is around 3 percentage points. The most serious difference relates to Grade 1, which is widely considered to include many repeaters. This is reflected in the NIDS data but, surprisingly, not in the GHS data. Despite these problems, the two sources used for the above graph arguably provide a relatively good picture of what the current grade repetition is, and permit a considerably better picture than any other available sources.

The 2004 DoE collection confirms the existence of a high level of grade repetition in Grade 1 (all three curves in Figure 26 indicate a level of around 12% for this grade). However, it seems to difficult to obtain a clear picture of the historical trend given the very large discrepancies with respect to the shapes of the two 2004 curves, in particular at the secondary level. This underlines the difficulty of using the NIDS imputations of past repeater trends. The smallness of the 2004 DoE sample could be a part of the comparability problem.

Figure 26: Grade repetition in NIDS and 2004 household survey



Source: National Income Dynamics Study 2008 (the 'NIDS 2004' values are from the 2008 collection, as reflected in Table 15); Collection by Department of Education 2004.
Note: For 2004, the overall percentage of learners repeating their grade according to the DoE source is 8.2% for Grades 8 to 12 (against 12.4% using NIDS) and 7.1% for Grades 1 to 12 (against 8.0% using NIDS).

Finally, the following table presents the results of a logit model focussing on the relationship between the NIDS mathematics scores, grade repetition and poverty. Specifically, the model investigates whether lower scores are more likely to be associated with grade repetition amongst better off South Africans than amongst poor South Africans. The model uses six explanatory variables, each one being an interaction term where the mathematics score is multiplied by a dummy variable indicating either the grade in which the learner was enrolled the previous year (2007) or whether the learner is poor or not, where anyone in the bottom three income quintiles was considered poor. The dependent variable is a dummy variable indicating whether one is repeating in 2008 the grade one took in 2007. One would expect all coefficients to be negative. For instance, if one considers two learners who were in Grade 9 in 2007, of whom learner A is repeating Grade 9 in 2008 and learner B is in Grade 10 in 2008, then it is to be expected that learner A's mathematics test score would be lower than learner B's, both because learner B would be a more capable learner generally and because learner B would have the benefit of some Grade 10 tuition. It is not possible to separate the two effects, namely the effect of being a generally better learner and the effect of having received more advanced tuition. However, both effects should work in the same direction and lead to the expectation that higher scores in the mathematics test would be associated with a lower probability of having repeated. Not all the coefficients in Table 17 are negative. However, the two that are of special interest, namely the test scores of poor learners and the test scores of non-poor learners, are negative and are highly statistically significant. What should be of particular interest from a policy perspective is the difference in the magnitudes of the poor and non-poor variables. If the coefficient for non-poor learners is larger it could suggest that scores are better correlated (negatively) with repeating amongst the non-poor and thus that the schools-based assessments that inform grade repetition are more suitable, relative to those in schools serving poor learners. This, in turn, would confirm a hypothesis similar to one found to be true by Lam, Ardington and Leibbrandt (2008). The negative coefficient for non-poor learners in Table 17 is indeed slightly larger than the coefficient for poor learners. However, the difference is small and is not statistically significant, even if the individual coefficients are. Specifically, the model indicates that a one standard deviation improvement in one's test score reduces one's probability of being a repeater by 2.7 percentage points in the case of a non-poor learner and 2.1 percentage points in the case of a poor learner. Whilst the model does not contradict the hypothesis that schools for non-poor learners are better at determining who should repeat a grade, it does not strongly support the hypothesis either, though it does weakly support the hypothesis[17].

Table 17: Grade repetition logit model

Variables	*Coef-ficients*	*Change in value*	*Change in probability*
Was in Gr 7 * Score	-0.03	-0.50 to 0.50	-0.002
Was in Gr 8 * Score	0.07	-0.50 to 0.50	0.004
Was in Gr 9 * Score	0.12**	-0.50 to 0.50	0.007
Was in Gr 10 * Score	0.23***	-0.50 to 0.50	0.013
Is poor * Score	-0.36***	-0.50 to 0.50	-0.021
Is non-poor * Score	-0.45***	-0.50 to 0.50	-0.027
Constant	0.10		
N	1370		
Pseudo R^2	0.104		

*Note: The dependent variable is whether one is repeating one's current grade. *** indicates that the estimate is significant at the 1% level of significance, ** at the 5% level. The change in probability is obtained using the* mfx compute *command in Stata.*

[17] Various transformations, including the use of natural logarithms of the mathematics score, led to similar conclusions to what is presented here.

Appendix E: Comparative statistics on grade-on-grade progress

The following table includes the figures used for Figure 16. PISA 2003 mathematics data were used for the countries other than South Africa (ZAF). The South Africa figures are based on NIDS. PISA countries were chosen purely on the basis that the principal focus had to be on developing countries and that countries should be selected against which South Africa is commonly compared. For each PISA country, learners, who are all age fifteen, were broken down by repetition status, where this was determined by responses to a series of questions on grade repetition. For instance, '111' means the learner had not repeated at the primary, lower secondary or upper secondary level, whilst '11-' means that there was no repetition at the primary and lower secondary levels and that there was no response for upper secondary. The value 2 refers to the fact that there was grade repetition at one of the three schooling levels. By performing the analysis by repetition status the risk that factors associated with grade repetition would bias the results was minimised. One would expect, for instance, fifteen year olds in Grade 9 to have repeated more than fifteen year olds in Grade 10, on average, and the education research evidence points to the fact that learners with more repetition in their enrolment history perform worse, on average, compared to their grade peers. The intraclass correlation coefficients were then calculated, for pairs of grades at a time, using grade as the class divider. For example, in Brazil, when Grades 9 and 10 learners with no history of grade repetition were analysed, one's grade was found to explain 5.3% of the overall variation in mathematics scores amongst these learners. Only where at least 100 observations for each of the two grades being analysed were available, did the calculation of the intraclass correlation coefficient occur. Where more than one repetition status existed for a country, the average across the available statistics was used for Figure 16. An asterisk indicates that the statistic should be ignored as the higher grade displayed a lower average score than the lower grade. This only occurs in the case of non-repeaters in Grades 9 and 10 in Mexico (and, as discussed previously, in Grades 8 and 9 in South Africa). The right-hand pane of Table 18 indicates that grade-on-grade gains in many countries are substantial. For example, the gain in Brazil between Grades 8 and 9 is equal to about a third of the difference between never-repeating Grade 9 learners in Brazil and never-repeating Grade 9 learners in Australia.

Table 18: Details on grade-on-grade improvement

	Repetition	*Intraclass correlation coefficients*				*Average PISA score*			
Country	*status*	*8*	*9*	*10*	*11*	*8*	*9*	*10*	*11*
ZAF			0.000*	0.000	0.041				
BRA	111			0.053			400	430	
	11-		0.136			310	353		
	2--		0.102			286	327		
FIN	11-		0.093			517	552		
TUN	111				0.031			425	444
	2--		0.151			309	342		
THA	111				0.330			441	522
	11-			0.064			397	423	
AUS	11-			0.052			495	528	
	2--			0.049			431	476	
MEX	11-			0.006*			398	398	
	1--			0.104			331	365	
	21-		0.030			320	336		

Appendix F: Regression models on test results and the labour market

The following table provides the detailed outputs of the models used for Figure 18 above, which focuses on the conditional and unconditional relationship between years of schooling and mathematics test results, using the NIDS 2008 dataset. Model A analyses the unconditional relationship between the two variables. Model B examines the relationship conditional on age and Model C uses, in addition to age, dummy race variables and the maximum level of education of either parent as socio-economic status conditioning variables. Race is used here largely due to the strong linkages between race and the quality of education received in the apartheid system.

Table 19: Regression of test results on years of schooling

Variables	*Model A*	*Model B*	*Model C*	*Mean*	*Std. dev.*
Age		-0.01	0.00	27.6	17.0
Age squared		0.00	0.00	1048	1219
Is coloured			0.26***	0.08	0.28
Is Indian			-0.13	0.02	0.15
Is white			0.30***	0.10	0.29
Attained Gr 6	0.05	0.06	-0.16	0.06	0.24
Attained Gr 7	0.04	-0.24	-0.51	0.07	0.26
Attained Gr 8	0.43***	0.34	0.53	0.08	0.27
Attained Gr 9	0.14	0.65	0.55	0.07	0.26
Attained Gr 10	0.43***	0.01	-0.03	0.10	0.30
Attained Gr 11	0.55***	0.18	0.18	0.09	0.28
Attained Gr 12	0.55***	0.09	0.07	0.16	0.36
Attained 13 years	0.77***	-0.43	-0.47	0.06	0.24
Attained 15 years	1.52***	0.06	0.15	0.03	0.16
Age*Gr 6		0.00	0.00	1.66	8.04
Age*Gr 7		0.01	0.01	2.23	9.13
Age*Gr 8		0.00	0.00	2.58	10.23
Age*Gr 9		-0.02*	-0.02	1.91	7.67
Age*Gr 10		0.01	0.01	3.23	10.74
Age*Gr 11		0.01	0.01	2.38	8.18
Age*Gr 12		0.01	0.01	5.10	12.75
Age*Gr 13		0.03***	0.03***	2.23	9.32
Age*Gr 15		0.04***	0.03*	1.10	7.19
Max. parent education			0.01**	4.99	5.06
Constant	-1.05***	-0.80*	-0.84*		
N	2188	2187	1838		
R^2	0.095	0.112	0.142		
Adjusted R^2	0.092	0.104	0.131		

*Note: The dependent variable is the normalised test score across the four tests, which has a mean of -0.51 and a standard deviation of 1.09 if all observations in the dataset are considered. *** indicates that the estimate is significant at the 1% level of significance, ** at the 5% level and * at the 10% level. Means and standard deviations are those across the 2188 observations used for Model A. In Model B, the two variables Attained Gr 8 and Age*Gr8 are jointly significant (using a Wald test) at the 5% level. The same can be said for the variables relating to 10, 11, 12, 13 and 15 years of schooling. In Model C, joint significance is also found at 10, 11, 12, 13 and 15 years of schooling.*

Table 20 presents the results of an earnings function, where monthly non-zero income according to the NIDS dataset is regressed on a number of explanatory variables, many of which are commonly used in South African analyses of this kind. In Model A, the only education variables included are those relating to the years of schooling of the income earner. The R^2 value for Model A is similar to those of comparable earnings functions produced by Keswell and Poswell (2004: 841), in other words the NIDS data explain income relatively well. Model B uses the same variables as Model A, but restricts observations to those where the mathematics test score is present. This reduces the sample by about 80%. Model C includes all variables. Model D excludes only the mathematics test score variable. The fact that the coefficients in Model D are similar to those in Model C (with the exception of the lower years of schooling variables) suggests that the model based on a smaller number of observations, namely Model C, can be regarded as relatively reliable and that the mathematics test score coefficient in Model C would probably not have changed greatly had values been available for Model D. The NIDS values on the highest level of education of the respondent's parents (whether the parents were alive or not) were tested in the earnings function and were found, rather surprisingly given the theoretical importance of this variable, to be insignificant. Parent education was therefore excluded from the final models in Table 20.

Table 20: Earnings function

Variables	*Model A*	*Model B*	*Model C*	*Model D*	*Mean*	*Std. dev.*
Attained Gr 8	0.40***	0.07	0.01	0.36***	0.10	0.30
Attained Gr 9	0.31***	-0.11	-0.18	0.26***	0.14	0.34
Attained Gr 10	0.45***	-0.16	-0.24**	0.34***	0.16	0.37
Attained Gr 11	0.53***	0.22*	0.00	0.35***	0.16	0.37
Attained Gr 12	1.13***	0.83***	0.45***	0.74***	0.19	0.39
Attained 13 years	1.16***	0.90***	0.53***	0.77***	0.05	0.22
Attained 14 years	1.82***	1.86***	1.27***	1.31***	0.03	0.18
Attained 15 years	2.25***	2.14***	1.51***	1.71***	0.02	0.14
Age	0.11***	0.09***	0.10***	0.12***	26.8	10.6
Age squared	0.00***	0.00***	0.00***	0.00***	830	707
Is female	-0.55***	-0.61***	-0.63***	-0.56***	0.57	0.50
Is coloured	0.24***	0.16*	0.03	0.14***	0.15	0.36
Is Indian	0.60***	0.21	0.05	0.40***	0.00	0.07
Is white	0.76***	0.70***	0.42***	0.55***	0.03	0.18
Writes English			0.15**	0.16***	0.53	0.50
Did Gr 12 math			0.20***	0.13***	0.17	0.37
Computer literate			0.35***	0.46***	0.31	0.46
Maths test score			0.12***		-0.50	1.07
Constant	4.45***	4.88***	4.77***	4.26***		
N	5165	1008	1008	5165		
R^2	0.485	0.555	0.578	0.508		
Adjusted R^2	0.484	0.549	0.570	0.506		

*Note: The dependent variable is the natural log of monthly labour market income, using the relevant NIDS variables listed in Argent (2009: 21). *** indicates that the estimate is significant at the 1% level of significance, ** at the 5% level and * at the 10% level. Means and standard deviations are those across the 1008 observations used for Model B and Model C.*

Table 21 provides the results of five logit models where the dependent variable is whether the person is employed or not. Model A includes all variables, whilst Models B to E each enter just one of four key educational quality variables that are of interest: whether the respondent believes he or she can write well in English, whether the respondent took mathematics in Grade 12, whether the respondent considers himself or herself computer literate and the mathematics test score of the respondent. The last four models suggest that computer literacy and the mathematics test score are more statistically significant predictors of being employed than English writing skills and having done mathematics in Grade 12. The last column in Table 21 provides the estimated expected increase in the probability of being employed, given certain shifts in the values of the explanatory variables. For example, having attained 15 years of education (as opposed to less than Grade 8) increases one's probability of being employed by 33 percentage points.

Table 21: Employment logit model

Variables	*Model A*	*Model B*	*Model C*	*Model D*	*Model E*	*Change in value*	*Change in probability (Model A)*
Attained Gr 8	-0.13	-0.14	-0.28*	-0.27*	-0.26		
Attained Gr 9	0.02	0.04	-0.32**	-0.30*	-0.28*		
Attained Gr 10	-0.02	-0.01	-0.08	-0.04	0.00		
Attained Gr 11	0.20	0.25	-0.14	-0.07	-0.02		
Attained Gr 12	0.21	0.37	0.14	0.29**	0.36***		
Attained 13 years	0.29	0.47	0.09	0.30*	0.37**		
Attained 14 years	0.67	0.97**	0.91***	1.19***	1.28***		
Attained 15 years	3.74***	4.02***	1.87***	2.16***	2.26***	0 to 1	0.33
Age	0.09*	0.08	0.11***	0.10***	0.10***	+1	0.02
Age squared	0.00	0.00	0.00***	0.00***	0.00***		
Is female	-1.28***	-1.28***	-0.90***	-0.90***	-0.90***	0 to 1	-0.26
Is coloured	0.31	0.36	0.29**	0.38***	0.38***		
Is Indian	3.53***	3.61***	0.82**	0.93**	0.97**	0 to 1	0.31
Is white	0.34	0.54	0.32	0.50**	0.51**		
Writes English	-0.08				-0.09		
Did Gr 12 math	0.20			0.04			
Computer literate	0.32*		0.42***			0 to 1	0.07
Maths test score	0.15**	0.15**				+1.09	0.03
Constant	-1.58*	-1.38*	-1.71***	-1.58***	-1.56***		
N	1943	1943	8769	8769	8769		
Pseudo R^2	0.183	0.180	0.120	0.117	0.117		

*Note: The dependent variable is the being employed (as opposed to being strictly unemployed or 'discouraged unemployed'). The mean for this dependent variable is 0.64 in the smaller sample (1621 observations) and 0.70 in the larger sample (7845 observations). *** indicates that the estimate is significant at the 1% level of significance, ** at the 5% level and * at the 10% level. The change in probability is obtained using the* mfx compute *command in Stata. The estimate for age in the final column takes into account both of the age variables. The 1.09 increase in the maths test score represents an increase of one standard deviation (measured across all test-takers).*

Appendix G: Grade 12 subject choices

The 2009 examinations dataset that was analysed included 623,229 examination candidates. This number was reduced to 528,986 candidates when only those with seven or eight subjects with non-missing scores who took life orientation as one subject and were in a school with a quintile value (indicating the school was public) were included. These 528,986 learners were distributed across 5,877 schools and formed the basis for the analysis that follows.

The next table provides the raw numbers behind Table 7. The reason why the overall total is less than 528,986 is that learners who had fewer than or more than four non-language subjects (other than life orientation) were excluded. The reason why learners would have fewer than four non-language subjects is largely that some learners took a non-official language, such as French, instead of a non-language subject.

Table 22: Participation in most common non-language combinations

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Overall
MATH + GEOG + PHSC + LFSC	18,028	13,589	18,557	12,535	7,057	69,766
MLIT + ACCN + BSTD + ECON	14,937	11,350	16,833	10,250	4,187	57,557
MATH + ACCN + BSTD + ECON	13,347	8,447	12,947	8,600	3,593	46,934
MATH + PHSC + LFSC + AGRS	14,743	8,225	10,239	2,668	1,225	37,100
MLIT + HIST + GEOG + LFSC	8,128	5,223	6,814	4,014	1,764	25,943
MLIT + GEOG + LFSC + AGRS	4,632	3,614	3,994	1,130	525	13,895
MATH + PHSC + LFSC + ACCN	1,352	990	2,938	3,070	3,459	11,809
MLIT + HIST + GEOG + TRSM	2,898	1,650	2,944	2,374	667	10,533
MATH + PHSC + LFSC + CATN	1,628	1,698	2,271	2,367	1,960	9,924
MLIT + HIST + GEOG + AGRS	4,721	1,653	2,256	617	79	9,326
Other combinations	36,983	25,632	45,739	43,829	64,043	216,226
Total	121,397	82,071	125,532	91,454	88,559	509,013

The next two tables provide a similar analysis, except here the breakdown is by province instead of quintile. Table 25 below provides a view similar to that of Table 8, by province.

Table 23: Cumulative participation by province in non-language combinations

	EC	FS	GP	KN	LP	MP	NC	NW	WC	Overall
MATH + GEOG + PHSC + LFSC	15	14	13	12	18	14	10	18	8	14
MLIT + ACCN + BSTD + ECON	20	34	27	22	30	29	16	35	14	25
MATH + ACCN + BSTD + ECON	33	40	33	35	40	35	20	39	20	34
MATH + PHSC + LFSC + AGRS	45	43	34	43	54	46	22	42	20	42
MLIT + HIST + GEOG + LFSC	51	48	38	48	60	49	33	48	27	47
MLIT + GEOG + LFSC + AGRS	52	49	39	49	67	54	39	55	27	49
MATH + PHSC + LFSC + ACCN	55	52	41	52	68	55	42	56	30	52
MLIT + HIST + GEOG + TRSM	57	53	43	55	70	57	44	57	31	54
MATH + PHSC + LFSC + CATN	59	58	45	57	71	59	47	59	33	56
MLIT + HIST + GEOG + AGRS	61	59	45	57	76	63	47	62	34	58
Other combinations	100	100	100	100	100	100	100	100	100	100

Table 24: Participation my province in non-language combinations

	EC	FS	GP	KN	LP	MP	NC	NW	WC	Overall
MATH + GEOG + PHSC + LFSC	9,695	3,342	11,624	14,666	14,856	6,636	958	4,696	3,293	69,766
MLIT + ACCN + BSTD + ECON	3,166	4,966	11,321	13,532	9,696	7,508	634	4,135	2,599	57,557
MATH + ACCN + BSTD + ECON	8,634	1,411	5,947	15,549	8,462	3,237	361	1,132	2,201	46,934
MATH + PHSC + LFSC + AGRS	7,721	742	511	10,118	11,704	5,229	177	668	230	37,100
MLIT + HIST + GEOG + LFSC	3,363	1,143	3,815	6,432	4,470	1,369	1,044	1,706	2,601	25,943
MLIT + GEOG + LFSC + AGRS	960	195	80	1,493	6,193	2,665	594	1,651	64	13,895
MATH + PHSC + LFSC + ACCN	1,997	659	2,345	3,485	817	519	355	459	1,173	11,809
MLIT + HIST + GEOG + TRSM	1,303	310	1,874	4,374	1,076	822	156	192	426	10,533
MATH + PHSC + LFSC + CATN	1,056	1,243	1,488	2,252	1,125	739	297	537	1,187	9,924
MLIT + HIST + GEOG + AGRS	1,475	146	46	581	4,012	2,163	10	807	86	9,326
Other combinations	24,886	9,990	47,383	53,975	19,772	18,141	5,135	9,594	27,350	216,226
Total	64,256	24,147	86,434	126,457	82,183	49,028	9,721	25,577	41,210	509,013

Table 25: Individual subject participation across provinces

		EC	FS	GP	KN	LP	MP	NC	NW	WC	Overall	Learners
DNCE	Dance Studies	5	0	28	0	0	2	0	0	63	100	330
MUSC	Music	14	4	16	29	0	3	1	4	30	100	1,291
VSLA	Visual Arts	5	2	34	27	1	4	1	5	21	100	6,116
DRMA	Dramatic Arts	3	3	16	65	0	1	0	0	12	100	5,460
DSGN	Design	3	2	40	7	0	6	0	4	39	100	2,250
HIST	History	14	4	15	25	15	9	3	5	11	100	84,030
GEOG	Geography	11	4	14	23	20	10	2	7	7	100	199,725
MATH	Mathematics	15	5	16	26	17	9	1	5	6	100	269,697
MLIT	Mathematical Literacy	9	6	18	22	15	11	3	7	9	100	259,260
PHSC	Physical Science	14	6	17	23	18	9	1	6	6	100	205,368
LFSC	Life Sciences	14	5	14	24	19	10	2	6	7	100	277,569
INFT	Information Technology	4	4	32	34	5	4	1	5	10	100	5,528
CATN	Computer Applications Technology	10	9	24	17	6	7	3	5	18	100	46,847
AGRS	Agricultural Sciences	17	2	1	20	34	19	1	5	1	100	84,818
ACCN	Accounting	11	6	18	28	13	9	2	5	7	100	162,426
BSTD	Business Studies	11	6	19	26	12	9	2	5	9	100	190,601
ECON	Economics	12	6	18	25	17	10	1	5	6	100	142,446
MCHT	Mechanical Technology	9	7	27	28	5	6	3	8	8	100	6,842
ELTT	Electrical Technology	8	8	32	22	8	6	3	8	6	100	6,197
GRDS	Engineering Graphics and Design	8	7	29	25	6	6	2	8	9	100	24,715
TRSM	Tourism	12	4	19	32	9	13	2	3	6	100	67,865
HOSP	Hospitality Studies	10	3	26	34	7	10	2	4	5	100	10,065
CNST	Consumer Studies	11	8	21	15	7	7	2	6	22	100	31,083
CVLT	Civil Technology	7	7	27	18	4	6	1	8	21	100	9,334
RLGS	Religion Studies	26	0	4	15	16	15	0	2	22	100	1,621
	Overall	12	5	17	24	16	10	2	6	8	100	528,986

What follows is an explanation of the method used to arrive at the figures in Table 9 and Table 10. The explanation proceeds with reference to the treatment of a fairly typical (but fictitious) learner. The learner whose results are given in Table 26 did fulfil the requirement that a score of at least 40 (out of 100) had to be achieved in one's home language (in this case isiZulu). The learner also did fulfil the requirement that at least three subjects had to carry a score of at least 40. However, he failed to fulfil the requirement that at least six subjects had to carry a score of at least 30. The learner only had five subjects with at least 30. Therefore the learner did not obtain a National Senior Certificate in 2009.

Table 26: Results of one learner

Subject	Score
IsiZulu HL	70
English FA	52
Life Orientation	71
Mathematics	21
Geography	35
Physical Science	23
Life Sciences	30

What the computer program did for every non-passing learner was to first check what alternative non-language subjects were available in the learner's school through an examination of the subjects taken by all other learners in the school (in other words records outside of the sample of 5,000 learners were considered). In the case of the learner in question, a matrix of possible switches could look as follows:

Table 27: Subject switching matrix

Alternative subjects offered in the school ▶ Subjects of the learner ▼	Mathematical literacy	Accounting	Business Studies	Economics	Engineering Graphics and Design	Hospitality Studies	Tourism
Mathematics	✓✓						
Physical Science		✓✓	✓✓	✓	✓	✓	✓

There were seven possible subject switches available to the learner, indicated by the cells with ticks in the matrix. Only mathematics and physical science are included in the matrix from amongst the learner's subjects as these were the subjects where a score less than 30 was achieved. For each possible switch, the availability of data amongst all candidates to estimate an alternative score was checked. For example, to test the physical science to tourism switch, a search was run through the database (not just the sample of 5,000) to see which learners took mathematics, geography, tourism and life sciences (this is the combination of the learner, with tourism replacing physical science). If at least 500 learners took this combination, then the estimation of an alternative score was deemed possible. The 500 threshold ensured that enough observations in the dataset were available for a statistically reliable estimate, but it also ensured that very unusual combinations, which are likely to clash with general timetabling practices in schools, would not be considered feasible alternatives. Of the seven possible subject switches, three could be simulated because the 500 observation threshold was passed. These three are marked with double ticks in the matrix.

For the three switches where estimates for the alternative subject were possible, the estimate was found using an ordinary least squares regression approach. For instance, to estimate the expected accounting score, the coefficients *b* for the following regression equation were found, where Y was accounting scores of learners, X_1 was the mathematics scores of learners, X_2 the geography scores, and X_3 the life sciences scores.

$$Y = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3$$

The values for the four b coefficients were then used to obtain an estimated accounting score for the learner in question. The regression equations predicted actual scores rather well. The R^2 value for the equation just described to estimate accounting scores was 0.80. Of the 239 regression equations, each with a different combination of Y and X values, that were possible and necessary in the analysis of the sample of 5,000 learners, half had R^2 values exceeding 0.70 and 90% had R^2 values exceeding 0.58.

In the case of the learner in question, all three switches for which estimates were possible resulted in a score of at least 30 for mathematical literacy, accounting or business studies. In other words, either of the three switches would give the learner the required six subjects with a score of at least 30, and hence the National Senior Certificate.