Education: Analysis of the NIDS Wave 1 Dataset

Discussion Paper no. 3

Nicola Branson
Southern African Labour & Development Research Unit
Nicola.Branson@uct.ac.za

David Lam
Department of Economics, University of Michigan
davidl@umich.edu

July 2009

1. Introduction

The National Income Dynamics Study (NIDS) questionnaires devote considerable attention to education. This report analyzes Wave 1 data corresponding to the sections of the questionnaires that are most specifically related to education – Module B of the household questionnaire, Module C of the child questionnaire, and Module H of the adult questionnaire. Many of the questions in these modules are similar to questions on other national surveys such as the Labour Force Survey (LFS), the General Household Survey (GHS), and the Community Survey (CS). In the first part of this report we compare estimates from NIDS to estimates from the 2007 Community Survey and the LFS 2007 for variables that are captured in both surveys. NIDS also includes a number of education variables that are not captured in other national surveys. These include measures of grade repetition, age of starting and ending school, school fees, reasons for not attending school, and plans for post-secondary schooling. While we do not have space to analyze all of the education variables that are unique to NIDS, we will analyze some of the variables that we believe may be of major interest to researchers and policy makers.

2. Comparison of education data in NIDS to CS 2007 and LFS 2007

Many South African surveys collect information on school enrolment and educational attainment. This section compares the NIDS 2008 data for these variables to two other nationally representative surveys – the 2007 Community Survey (CS) and the September 2007 round of the Labour Force Survey (LFS). The sample sizes of the surveys vary considerably, so there is resulting variation in the precision of the estimates. The NIDS sample is the smallest. All estimates presented are weighted using the sample weights provided for each survey.
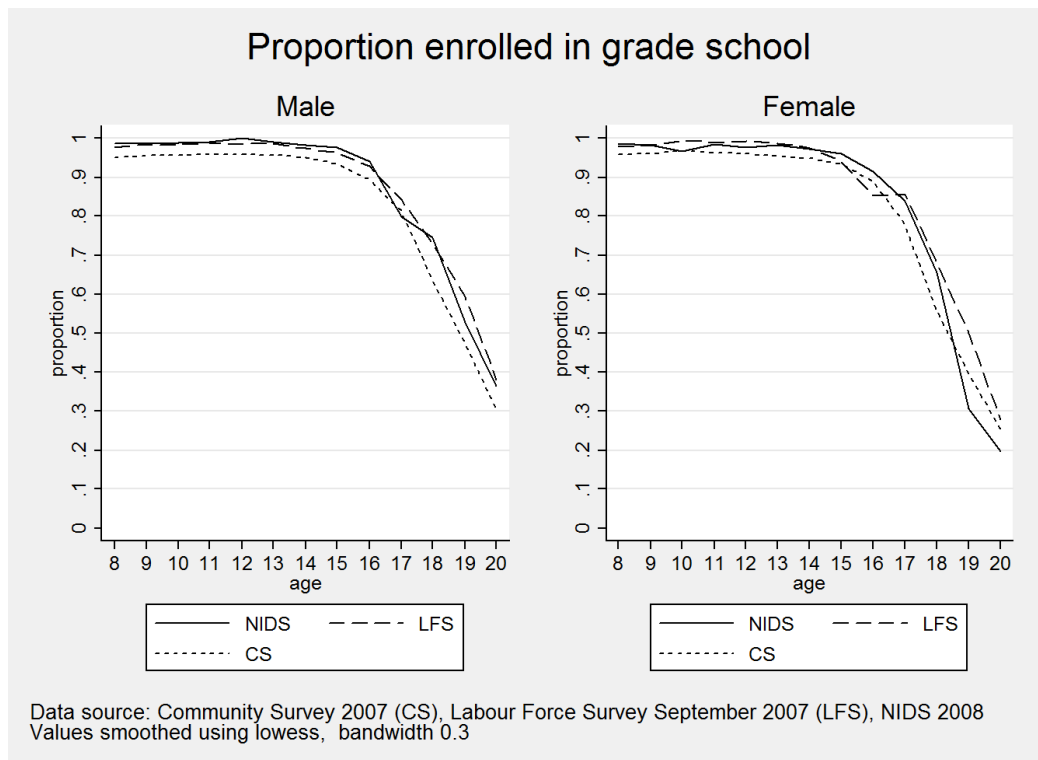
The questions on school enrolment and highest level of education attained are asked in slightly different ways in each of the surveys. In the NIDS data current school enrolment is taken from the questions "*Is this child currently enrolled in school?*" in the child questionnaire and "*Are you currently enrolled?*" in the adult questionnaire. Enrolment in the adult questionnaire included both schooling (that is, primary and secondary school) and post-secondary education. A supplementary question asking "*What level are you currently enrolled in?*" was used to distinguish schooling from other forms of study. The CS question is similar to the NIDS adult question except the wording is "*Does ... presently attend an educational institution?*" The supplementary question was "*Which of the following educational institutions does ... attend?*", with primary and secondary school being two options. The LFS question is "*Which of the following educational institutions, if any, does ... currently attend? Include distance and correspondence education.*" School is an option and is not differentiated into primary or secondary.

Table 1 presents the proportion of the population classified as in school and enrolled by age. Stratification by population group and sex for the in school variable are presented in Figure 1a and b. The NIDS numbers are roughly similar to the two other national surveys, but are most similar to the LFS data. The CS finds lower rates of enrolment at almost all ages than both NIDS and the LFS. Observing the age profile of the proportion of Coloureds in school presented in the right hand panel of Figure 2a, we see that NIDS gives higher estimates of the proportion of 14 to 17 year old Coloureds in school than the other two national surveys.
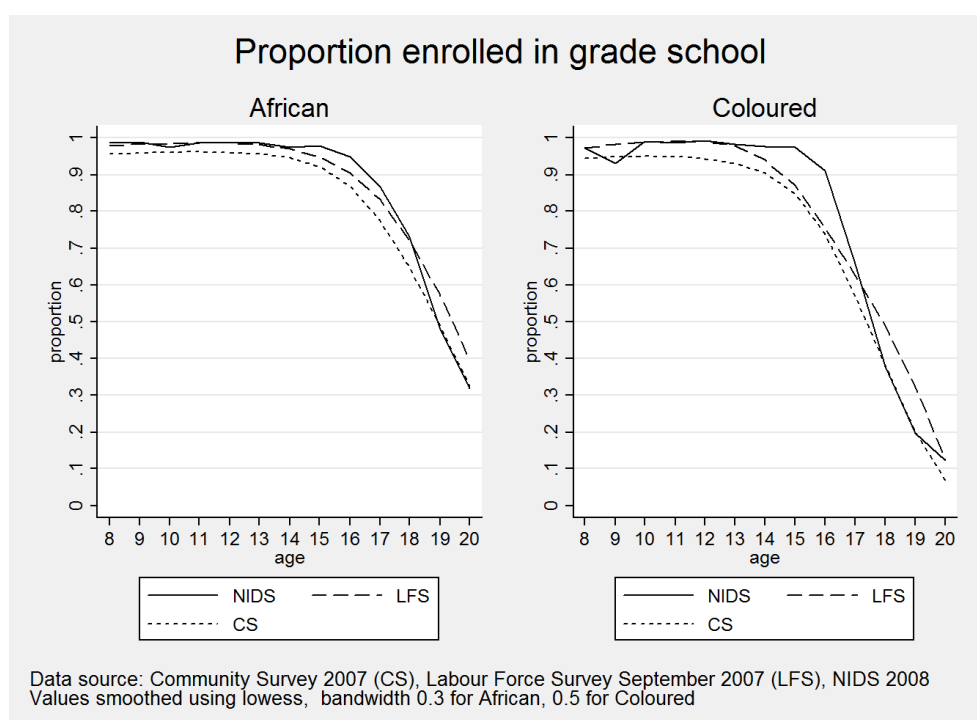
Table 1: Comparison of school enrolment in NIDS, CS 2007 and LFS 2007

Age	In School				Enrolment			
	NIDS		LFS	CS	NIDS		LFS	CS
8	0.99	(619)	0.98	0.96	0.99	(620)	0.98	0.96
9	0.98	(612)	0.98	0.96	0.98	(612)	0.98	0.96
10	0.98	(599)	0.99	0.96	0.98	(600)	0.99	0.96
11	0.99	(644)	0.99	0.96	0.99	(646)	0.99	0.96
12	0.99	(642)	0.99	0.96	0.99	(642)	0.99	0.96
13	0.99	(650)	0.99	0.96	0.99	(651)	0.99	0.96
14	0.98	(645)	0.97	0.95	0.98	(646)	0.98	0.95
15	0.97	(544)	0.95	0.93	0.97	(528)	0.95	0.93
16	0.93	(628)	0.89	0.89	0.95	(581)	0.90	0.90
17	0.82	(581)	0.85	0.80	0.85	(534)	0.86	0.83
18	0.70	(561)	0.71	0.60	0.74	(524)	0.75	0.68
19	0.42	(559)	0.55	0.44	0.55	(522)	0.65	0.56
20	0.28	(483)	0.33	0.28	0.42	(443)	0.44	0.43
Sample size	7767		31606	263876	7549		31606	263876

Figure 1a

Proportion enrolled in grade school

Data source: Community Survey 2007 (CS), Labour Force Survey September 2007 (LFS), NIDS 2008
Values smoothed using lowess, bandwidth 0.3

Figure 1b

Proportion enrolled in grade school

Data source: Community Survey 2007 (CS), Labour Force Survey September 2007 (LFS), NIDS 2008
Values smoothed using lowess, bandwidth 0.3 for African, 0.5 for Coloured

The NIDS highest completed schooling question is worded as follows in both the child and adult questionnaire: *"What is the highest grade in school that you have successfully completed?"* The adult questionnaire has a further instruction: *"Do not count the final year you were in school if you did not successfully complete the year.*" The NIDS team created a combined best education variable which is used as recommended.[1] This variable combines information from this question, a question on the highest level of education completed from the household questionnaire and information about post schooling education from the adult questionnaire. This variable can best be described as answering the question *"What is the highest level of education that ... has successfully completed?"* which is equivalent to the question asked in the LFS. The CS wording excluded the word successfully, asking *"What is the highest level of education that ... has completed?"* and included a category for matric attempted but not successfully completed. Thus there is scope for reporting of highest education level below grade 11 which was not successfully completed.

Table 2 presents the average highest school grade completed, the proportion completing grade 7 and grade 12 for adults who are not currently enrolled in any type of educational institution. As with the enrolment rates, the NIDS data are most similar to the LFS data, with both these surveys presenting a more optimistic picture of educational attainment than the CS data. The biggest discrepancies are in the youngest cohort, with the largest difference between the NIDS

---

[1] This variable is found the individual derived variables data and is called w1_best_edu.
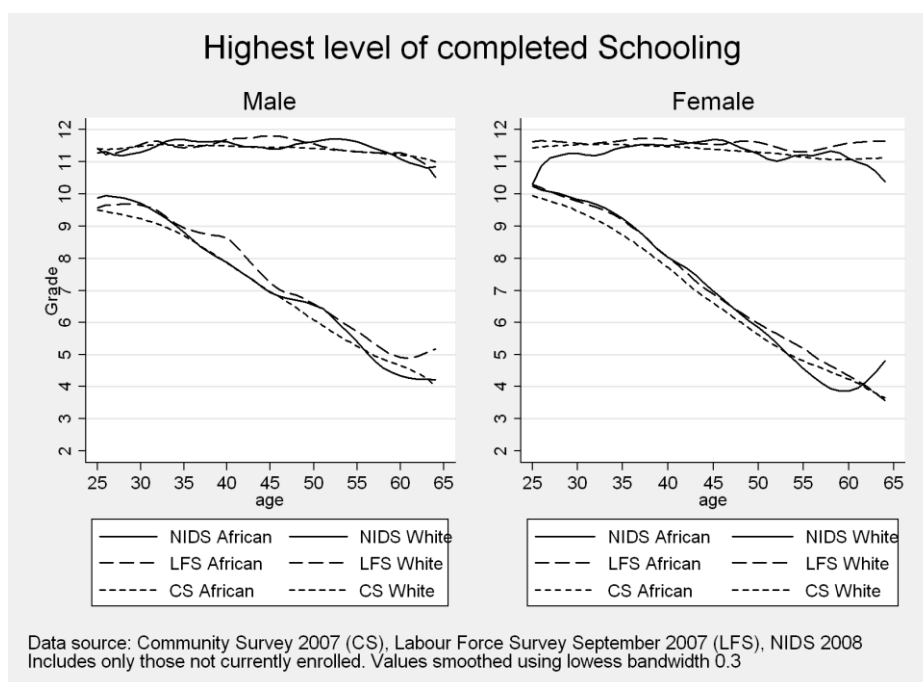
and the CS data. The CS looks particularly different from NIDS and LFS in the proportion completing grade 12. This is clearly represented in the African pattern in Figure 2b, which shows that the discrepancy decreases with age. This might be a result of the fact that the highest educational attainment question in the CS had a separate option for attempted grade 12 but not successfully, potentially signalling an over-estimate in the proportion of people with grade 12 in NIDS and LFS.

Figures 2 a through c present variables stratified by sex for the African and White sub-populations. Higher variability in the NIDS graph is a result of smaller sample sizes. Looking at Figure 2a we see that the NIDS data is most comparable to the other datasets on the highest completed schooling variable for males. The female African pattern shows a divergence in the older ages. In particular, we see that NIDS gives lower estimates between ages 52 and 60 and higher estimates between 60 and 65. Translation of standards into grades could have been a problem among the older cohorts. In addition, the NIDS data estimates lower levels of attainment for White females under 35 than seen in the other datasets. Looking at Figure 1b and c, this underestimation of White educational attainment appears to be driven by an underestimation of the proportion of White 20-35 year olds with grade 12. The NIDS data diverges substantially from the other datasets for White females who are 20-35. In Figure 2.c we see the same uptick in the proportion of Africans with grade 7 that was seen in the highest completed schooling graph.

Table 2: Comparison of educational attainment in NIDS, LFS 2007 and CS 2007

age group	Highest school grade completed			Proportion completing Grade 7			Proportion completing Grade 12		
	NIDS	LFS	CS	NIDS	LFS	CS	NIDS	LFS	CS
25-29	10.13	10.02	9.77	0.92	0.90	0.88	0.42	0.43	0.36
30-34	9.81	9.85	9.43	0.87	0.87	0.84	0.44	0.43	0.35
35-39	9.05	9.22	8.85	0.81	0.81	0.78	0.38	0.41	0.32
40-44	8.31	8.40	8.05	0.73	0.73	0.70	0.30	0.32	0.26
45-49	7.72	7.68	7.41	0.66	0.66	0.63	0.29	0.29	0.23
50-54	6.92	6.95	6.70	0.59	0.59	0.56	0.24	0.21	0.20
55-59	6.06	6.41	6.29	0.51	0.54	0.52	0.18	0.21	0.19
60-64	5.75	6.18	5.93	0.48	0.51	0.49	0.17	0.24	0.18
Sample Size	10728	41532	370370	10728	41532	370370	10728	41532	370370

Figure 2a

Highest level of completed Schooling

Data source: Community Survey 2007 (CS), Labour Force Survey September 2007 (LFS), NIDS 2008
Includes only those not currently enrolled. Values smoothed using lowess bandwidth 0.3

Figure 2b

Proportion completing Grade 12

Data source: Community Survey 2007 (CS), Labour Force Survey September 2007 (LFS), NIDS 2008
Includes only those not currently enrolled. Values smoothed using lowess bandwidth 0.3

Figure 2c

Proportion completing Grade 7

Data source: Community Survey 2007 (CS), Labour Force Survey September 2007 (LFS), NIDS 2008
Includes only those not currently enrolled. Values smoothed using lowess bandwidth 0.3
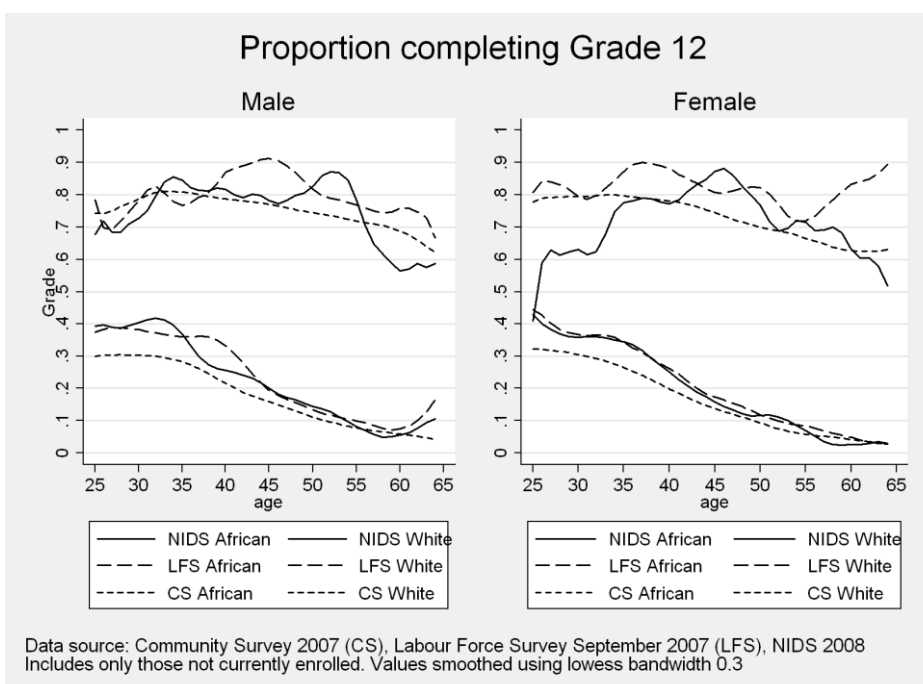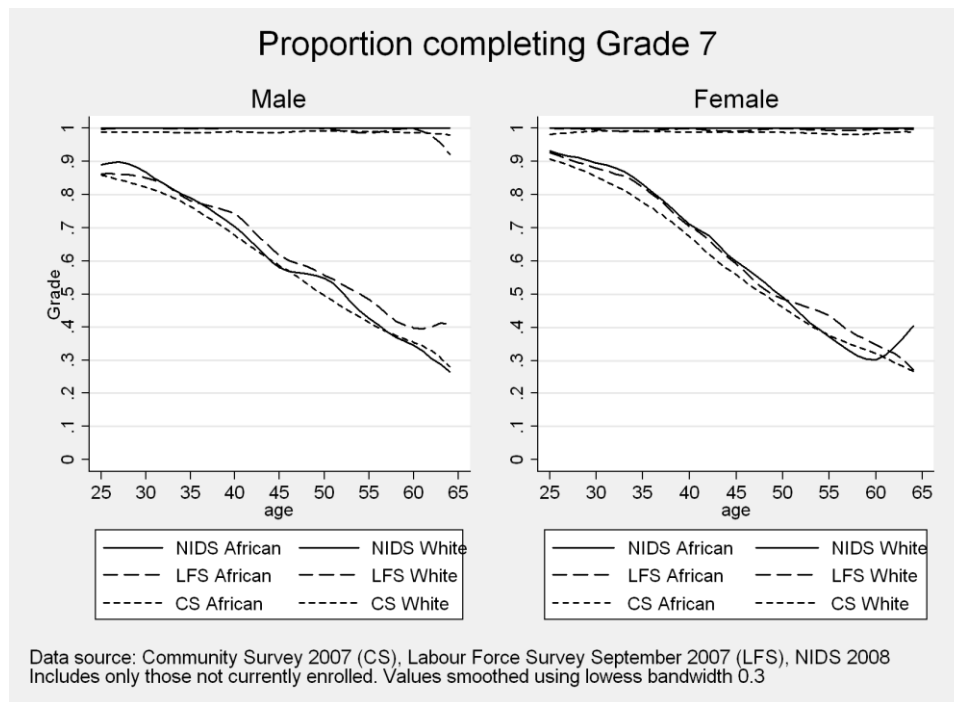
3. Unusual schooling variables in NIDS

This section investigates some of the interesting education variables in NIDS that are not available in other nationally representative South African surveys.

3.1 Grade Repetition

Grade repetition has been shown in a number of studies to be a major problem in South African schools. Direct measures of grade repetition are rare in South African household surveys, however, making it difficult to carefully study the issue. Evidence from the census or surveys such as the LFS on age distributions among those currently enrolled in a given grade suggest that there must be high rates of grade repetition, but this evidence is only indirect. NIDS provides the first nationally representative survey with detailed information on grade repetition.

The NIDS questionnaire addresses grade repetition through several questions that are not included in surveys such as LFS, GHS, or the Community Survey: First, there are direct questions on whether the respondent repeated up to four grades and the number of times each of these grades was repeated. There are also questions about the outcome (failed, passed, withdrew) of the grade in which the respondent was enrolled in 2007, with the same questions in the child and adult questionnaire. In addition to these questions on grade repetition and grade outcomes, there are also questions about the age at which the respondent started school and the age at which the respondent completed their highest school grade. The questions on age of starting and stopping school allow for an indirect count of the number of grades repeated by comparing the number of years in school to the highest grade attained. The direct questions on grade repetition are asked only of those under 31, but the questions on age of starting and stopping school are asked for all children and adults.

3.1.1 The quality of the data

a. Direct grade repetition questions

The direct grade repetition questions were asked of all children who had ever attended formal schooling and all adults under 31 who had attained some level of education (i.e. their highest grade completed was not zero). This resulted in a potential sample of close to 13,000 respondents (see table 3 below). The questions have three parts. The first question asked if the respondent had ever repeated a grade. On answering yes, the respondent was asked which grades were repeated and the number of times each grade was repeated. The questionnaire had

space to capture up to four repeated grades. The structure on the grade repetition was the same in the child and adult questionnaire.

The questions appear to have worked well – 91% of the potential sample gave valid answers to the question *"ever repeated a grade"*. The adult response rate was slightly below the child response rate which is in line with recall error. Of those who answered the question, 57% of adult respondents and 25% of child respondents reported having repeated a grade. The lower frequency of repetition in the child sample is a function of these respondents being at different stages of their schooling, while the majority of the adult sample has completed their schooling. The '*which grade'* and '*number of times'* questions were equally well answered.

One concern with the structure of the question is that respondents might have repeated more than four grades (the maximum number that could be recorded on the questionnaire). There are 57 cases where the respondent reported repeating four grades. Thus these people could potentially have repeated more than four. The questionnaire does not report the order in which the grades repeated were captured. Of those that repeated more than one grade, 17% started reporting on higher grades and worked down to lower grades. This percentage is even lower in the child sample. 51 of the 57 cases who reported repeating four grades started by reporting the lower grades they had repeated. Thus if they repeated more than 4 grades they are likely to be higher grades.

Table 3: Completeness of grade repetition questions[2]

	All	Adults	Children
Potential sample	12997	7170	5827
Answered *ever repeated grade* question	11854	6350	5504
% complete	91.2%	88.6%	94.5%
Number who repeated	5,020	3,630	1,383
% repeating	42.3%	57.2%	25.1%
Answered *which grade* question	4,981	3,617	1,364
% complete	99.2%	99.6%	98.6%
Answered *number of times* question	4,965	3,609	1,356
% complete	98.9%	99.4%	98.0%
Number of grades repeated:			
1	3,576	2,417	1,159
2	1,104	923	181
3	244	224	20
4	57	53	4
% that started with highest grade	17.3%	17.8%	14.1%

---

[2] Response rates for all the education questions are presented in the Appendix

b. Indirect grade repetition questions-age start and end school

The child and adult questionnaires included a question asking what year the respondent first attended school. Table 4 shows that this question had a high frequency of missing responses, especially in the adult sample. Only 41% of the adult sample reported valid information. The response of the child questionnaire was much better at 76.1%. Table 3 presents information on the response rate for individuals with complete information on the age they started school. This variable was calculated by subtracting the individual's birth year from the year they started school less one. All individuals born in January were assumed to have had their birthday by the time they started the school year and therefore their age at start of school year was calculated as birth year less age start school. Ages at start of school below four and above 11 were assumed implausible and set to missing (138 children, 284 adults).

Table 4: Response rates to age start/end school

	All	Adults	Children
Potential Sample	20456	14629	5827
	10474	6040	4434
		10538	
Age start school	51.2	41.3	76.1
Age end school		72.0	
% complete for age start school variable:			
by income quintile:			
1	48.7	35.1	71.2
2	52.3	39.2	76.0
3	50.5	39.2	78.6
4	48.8	41.7	77.3
5	57.3	52.3	83.7
by age group:			
5-9	87.5		87.5
10-14	71.1		71.1
15-19	60.4	60.4	
20-24	49.5	49.5	
25-29	44.8	44.8	
30-34	37.6	37.6	
35-39	38.7	38.7	
40-44	33.3	33.3	
45-49	32.5	32.5	
50-54	28.2	28.2	
55-59	26.2	26.2	
60-64	25.2	25.2	
65-69	22.6	22.6	
70-74	23.9	23.9	
75-79	24.1	24.1	
80+	18.1	18.1	

Table 4 also presents the response rate to the age start school variable by household income quintile and age group. It is clear from the table that recall error appears to be a contributing factor. Response rates decrease systematically with age, with a response rate of 60% in the 15-19 age group compared to under 25% in the age groups over 60. Response rates are also positively related to income. Observing the trends, it appears that this variable is probably best suited for analyses on those under age 20 and users should caution again bias towards younger and richer respondents if the entire sample is used.

Table 4 also presents the response rate to the variable *"In what year/what age did you successfully complete this grade?"* This question was asked in the adult and proxy questionnaire. Information for the adult sample only is included in the table. 72% of adult respondents who had completed at least some schooling answered this question with a reasonable value. Information given in years was converted to age by subtracting the birth year of those born in all months except December and subtracting an addition year for those born in December under the assumption that all children born before December had already had their birthday before the school year completed. Values which indicated that the respondent had completed more than two additional grades than the time they spent at school and negative values were set to missing (130 cases).

3.1.2 Evidence on grade repetition

Figure 3 shows the high prevalence of grade repetition in South Africa and the stark differences by population group and gender. By grade 10 over 50% of African males and 40% of African females have repeated at least one grade. There are very low levels of grade repetition among the Indian and White population groups, with Africans having close to double the level seen in the Coloured population group.

Table 5 presents the breakdown of grade repetition by grade. Males are much more likely to repeat at almost all grades than females with the greatest gender difference apparent in the younger grades. Grade repetition is highest in grades 1, 10, 11 and 12 for males, and remains low, below 7%, for females until grade 10. Since the sample includes all those respondents under 31 who had attempted the grade; the sample decreases at the higher grades and is selective of those who have not dropped out.

Figure 3

Proportion who have ever repeated by Grade
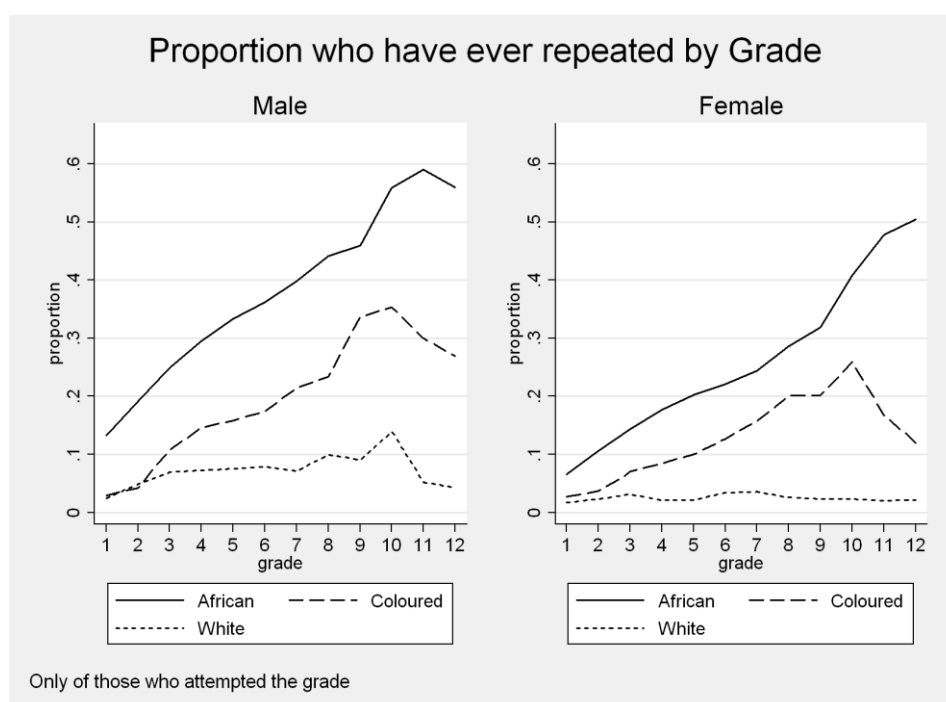
Only of those who attempted the grade

Figure 4 shows the proportion who has ever repeated a grade at each grade level, using only the sample that repeated at least one grade. A large proportion of African male grade repetition happens in grade 1. The proportion with at least one failed grade then rises gradually until grade 9, with a sharp increase starting at grade 9. About 30% of grade repetition occurs between grade 9 and 12 for African males. African females have an even higher concentration of repetition at the higher grades, with over 40% of repetition happening between grade 9 and 12. 45% of Coloured male grade repetition happens before grade 4 and 50% after grade 8, with very little grade repetition between grade 4 and grade 8. This is very different to the pattern for Coloured females which is more evenly spread out across the grades.

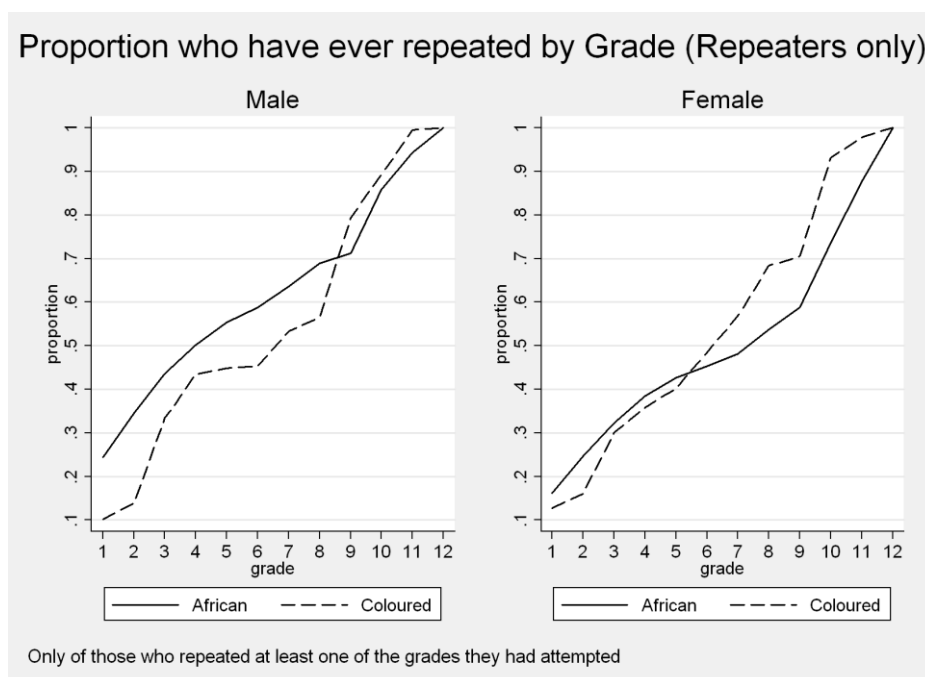**Table 5: Percentage repeating each grade, grades 1-12**

Males

Grade	All Males %	N	African %	N	Coloured %	N	Indian %	N	White %	N
1	11.77	(5226)	13.23	(4404)	2.95	(634)	0.00	(58)	2.49	(129)
2	8.35	(4814)	9.36	(4063)	1.97	(573)	0.53	(56)	2.34	(121)
3	8.80	(4492)	9.56	(3800)	7.06	(528)	0.00	(52)	1.73	(111)
4	8.03	(4063)	8.87	(3433)	5.84	(478)	0.00	(47)	0.00	(104)
5	6.91	(3627)	7.81	(3068)	3.02	(417)	0.00	(45)	0.00	(96)
6	5.45	(3232)	6.24	(2727)	1.55	(371)	0.00	(41)	0.00	(92)
7	6.19	(2850)	6.87	(2401)	4.01	(330)	0.00	(35)	0.45	(83)
8	9.99	(2435)	11.00	(2052)	6.25	(280)	0.00	(29)	2.18	(73)
9	8.89	(2036)	9.16	(1716)	15.32	(226)	1.85	(25)	1.04	(68)
10	20.67	(1582)	22.76	(1341)	10.60	(161)	0.00	(17)	8.67	(62)
11	16.26	(1092)	18.11	(936)	9.58	(96)	5.39	(15)	0.00	(44)
12	10.62	(652)	12.60	(541)	0.46	(65)	0.00	(12)	0.00	(34)

Females

Grade	All Females %	All Females N	African %	African N	Coloured %	Coloured N	Indian %	Indian N	White %	White N
1	5.96	(5927)	6.65	(4951)	2.76	(752)	0.00	(72)	1.71	(152)
2	4.18	(5591)	4.71	(4669)	1.51	(709)	0.00	(68)	0.86	(145)
3	5.23	(5270)	5.73	(4396)	3.93	(674)	0.00	(66)	1.25	(134)
4	4.69	(4910)	5.19	(4100)	2.09	(626)	4.52	(62)	0.00	(122)
5	3.71	(4530)	4.18	(3775)	1.64	(580)	0.62	(60)	0.38	(115)
6	3.70	(4148)	4.02	(3465)	2.81	(522)	0.00	(52)	1.45	(109)
7	3.83	(3713)	4.13	(3100)	4.38	(467)	0.00	(47)	0.00	(99)
8	6.27	(3285)	6.97	(2761)	4.89	(395)	0.00	(43)	0.00	(86)
9	6.82	(2850)	7.67	(2429)	4.01	(305)	0.00	(40)	0.05	(76)
10	14.17	(2349)	15.72	(2017)	11.83	(231)	2.92	(35)	0.00	(66)
11	16.50	(1704)	19.03	(1470)	4.71	(151)	0.00	(27)	0.00	(56)
12	15.66	(1041)	19.40	(858)	0.76	(110)	0.00	(21)	0.00	(52)

Notes to table 5: Sample restricted to under 31 year old who have attempted the grade in question. In other words, each row is calculated off a different sample, with sample size shown in the columns called N. Weighted proportions shown.

Figure 4

Proportion who have ever repeated by Grade (Repeaters only)

Only of those who repeated at least one of the grades they had attempted

3.1.3 Outcome of 2007 grade

In addition to collecting data on grades repeated, NIDS asks about the outcome of the grade attended in 2007. Identical questions appear in the child and adult questionnaires. Table 6 shows the distribution of responses by population group for children aged 7-14 in the child questionnaire and aged 15-19 in the adult questionnaire. The response "withdrew from school before completing the grade" is rarely selected for those under age 15 – under 1% for all groups. We generate an indicator that students failed the grade, using only those who report either having passed or failed. Table 7 shows the proportion who failed their 2007 grade by gender

and population group for three-year age groups. As we would expect from the above results on grade repetition, males are more likely to fail than females at most ages. One exception is the 19-21 age group, where females have higher failure rates for both Africans and Coloureds. This is a relatively small and somewhat unusual group, since it includes only those who are still in secondary school at these older ages. Failure rates are quite high in the 16-18 age group – 14% to 22% for African and Coloured students. Failure rates are very low for Indian and White students in all age groups.

It is clear from analyzing these variables that rates of failing grades and repeating grades are high in the South African schooling system, especially for Africans and Coloureds. We now turn to look at what effect this will have on grade attainment given the amount of time spent in school and the age distribution in the classroom.

Table 6: Distribution of outcomes of 2007 grade

	African	Coloured	Indian	White	Total
Age 7-14					
Withdrew	0.3	0.0	0.0	0.0	0.3
Failed	7.2	3.4	0.0	1.0	6.4
Passed	92.5	96.6	100.0	99.0	93.3
Age 15-19					
Withdrew	1.2	1.0	5.3	2.5	1.3
Failed	14.3	17.6	0.0	2.6	13.8
Passed	84.2	79.9	94.7	84.5	84.2

Table 7: Proportion who failed 2007 grade by age

	African		Coloured		Indian		White		
Age group	Female	Male	Female	Male	Female	Male	Female	Male	Total
Failure rate									
7-9	0.05	0.14	0.02	0.10	0.00	0.00	0.00	0.02	0.08
10-12	0.06	0.10	0.02	0.01	0.00	0.00	0.03	0.00	0.07
13-15	0.02	0.04	0.01	0.11	0.00	0.00	0.00	0.04	0.03
16-18	0.13	0.17	0.23	0.10	0.00	0.00	0.04	0.00	0.15
19-21	0.30	0.26	0.15	0.43	0.00	0.00	0.00	0.00	0.26
Sample size									
7-9	699	693	99	133	12	9	25	27	1,697
10-12	773	760	135	125	9	10	26	21	1,859
13-15	725	758	119	97	10	10	30	16	1,765
16-18	615	640	54	64	13	7	10	19	1,422
19-21	319	360	17	13	3	1	16	8	737

3.1.4. Grade attainment given years spent in school

In this section we use the NIDS variables on the age at which respondents started school and ended school. NIDS is the first large nationally representative survey to include these variables, which have the potential to answer a number of important questions about schooling. Table 8

presents the mean age of starting school, the mean age of leaving school (that is, primary or secondary school), the mean number of years spent in school (the difference between the leaving age and the starting age) and the mean of the highest grade attained by age group. Only those who are no longer enrolled are included in the sample. All individuals with missing data for the age of starting school are assumed to have started school at the average age others in the same age group and population group started school.

As seen in the top panel of Table 8, males and females start school at similar ages, although there are some important differences in the age at starting school across population groups.

The mean age of starting school for the 25-29 age group was just over 6.4 years for Africans, compared to around 5.5 for Coloureds, Indians, and Whites. There is a systematic decrease in the age Africans start school across cohorts – the difference between the oldest and the youngest cohort is over a year[3]. The mean age at leaving primary or secondary school is 19.4 years for Africans aged 25-29. This compares to 17.7 for Coloureds and Indians and 17.4 for Whites. The age of leaving school appears to have been relatively stable over time. The largest trend appears to be for Coloureds, where there is some evidence that young people are staying in school longer.

The bottom panel of Table 8 uses the starting age and ending age to calculate the number of years spent in primary and secondary school. Calculated values less than the highest grade minus two or more than double the highest grade are assumed to be implausible and set to missing (645). Africans aged 25-29 spent 12.9 years in school, compared to a mean below 12 years for Whites. It is striking that Africans aged 25-29 spent about 1.5 years longer in school than Whites. In spite of these extra years attending school, the last section of the table shows that Africans aged 25-29 ended up with more than a full grade less schooling than Whites, 10.1 for Africans compared 11.2 for Whites. Coloured 25-29 year-olds completed an average of 10.14 grades, 0.08 grades more than Africans, although they spent close to a year less in the schooling system. This is in line with the higher rates of dropout among Coloured shown, for example, in the Cape Area Panel Study.

While results for the older cohorts in Table 8 must be interpreted with caution due to the high percentage of missing values on the questions about age of starting and ending school, the evidence for the younger cohorts is further confirmation that grade repetition is an important factor affecting progress through school in South Africa, especially for Africans. The data make it

---

[3] Caution however needs to be taken in this comparison, due to the increase in the number of imputed ages that respondents start school by age cohort

clear that the lower grade attainment for Africans is not the result of Africans dropping out of school early. On the contrary, Africans are actually spending over a year longer enrolled in school than other groups, but end up with an average of over a full grade less in completed grade attainment.

Table 8 Age of starting and ending school

Age group	Age start school						Age end school					
	Male	Female	African	Coloured	Indian	White	Male	Female	African	Coloured	Indian	White
25-29	6.33	6.32	6.47	5.63	5.26	5.75	19.16	19.06	19.39	17.65	17.74	17.41
30-34	6.53	6.44	6.65	5.86	5.21	5.66	19.60	19.87	20.28	17.84	17.39	17.26
35-39	6.76	6.63	6.96	5.93	5.10	5.54	19.77	19.41	19.96	18.32	16.26	18.72
40-44	6.60	6.61	7.07	5.95	5.02	5.14	19.00	18.91	19.56	16.69	17.56	18.06
45-49	6.58	6.74	7.08	5.64	5.56	5.69	18.56	19.34	19.96	16.42	18.26	17.70
50-54	6.81	6.84	7.33	5.83	7.69	5.07	18.81	19.46	20.46	17.16	17.99	17.45
55-59	7.09	7.07	7.74	6.27	6.00	5.18	17.96	17.05	17.62	16.21	15.20	17.88
Total	6.61	6.60	6.90	5.85	5.60	5.42	19.19	19.24	19.77	17.45	17.41	17.79

Age group	Number of years in school						Highest school grade completed					
	Male	Female	African	Coloured	Indian	White	Male	Female	African	Coloured	Indian	White
25-29	12.81	12.72	12.91	12.04	12.48	11.68	10.03	10.30	10.06	10.14	11.07	11.23
30-34	12.83	13.27	13.37	12.08	12.22	11.60	9.79	9.88	9.62	10.20	11.51	11.48
35-39	12.48	12.80	12.78	11.82	11.28	13.19	8.91	9.11	8.56	9.81	10.38	11.60
40-44	11.84	11.67	11.74	10.30	12.58	12.20	8.36	8.23	7.60	7.97	10.63	11.39
45-49	11.74	11.88	11.81	10.70	12.64	12.24	7.79	7.57	6.61	8.33	9.95	11.68
50-54	11.07	11.17	10.74	10.42	10.20	12.23	7.13	6.68	5.85	6.45	10.02	11.23
55-59	11.10	10.47	9.33	10.65	10.04	12.52	6.44	5.59	4.34	6.42	7.18	11.46
Total	12.35	12.45	12.59	11.45	11.93	12.24	8.81	8.66	8.26	8.88	10.31	11.44

Notes to Table 8: Sample intended to be 25-59 year olds who were not studying at the time of the survey. The age start school and number of years in school numbers should be considered with caution since this question was poorly answered. Missing values are imputed to the average value within age group and population group cells. The probability of non-response on these variables increases with age, at an increasing rate which is in line with recall difficulties. In addition, young, White, better educated men are less likely to have missing data. The number of years in school is calculated using the age start and end school variables. Weighted means presented.

3.1.5 Age distribution in the classroom

Given the high levels of grade repetition it is of interest to see what affect this has on the age distribution within grade. To isolate the effect of being old for the current grade due to differences in the age that individuals start school from the effect of grade repetition, we look at the age distribution by grade in the following way. We define *appropriate grade* to be the grade an individual would be in if they progressed one grade each year until their current age. Then those who are in a grade which is below this appropriate grade have failed and those that are

above this appropriate grade have skipped grades. We graph the distribution of appropriate grade[4] for those currently in Grade 2 and grade 9 in Figures 5a and b respectively.

Figure 5a shows that the majority of children currently in grade 2 are in the appropriate grade for their age given the age they started school[5]. There is however, a significant amount of density to the right of grade 2, indicating that many of the children in grade two should really be in a higher grade given their current age and the age they started school.

Figure 5a

'Appropriate Grade' for those currently in Grade 2

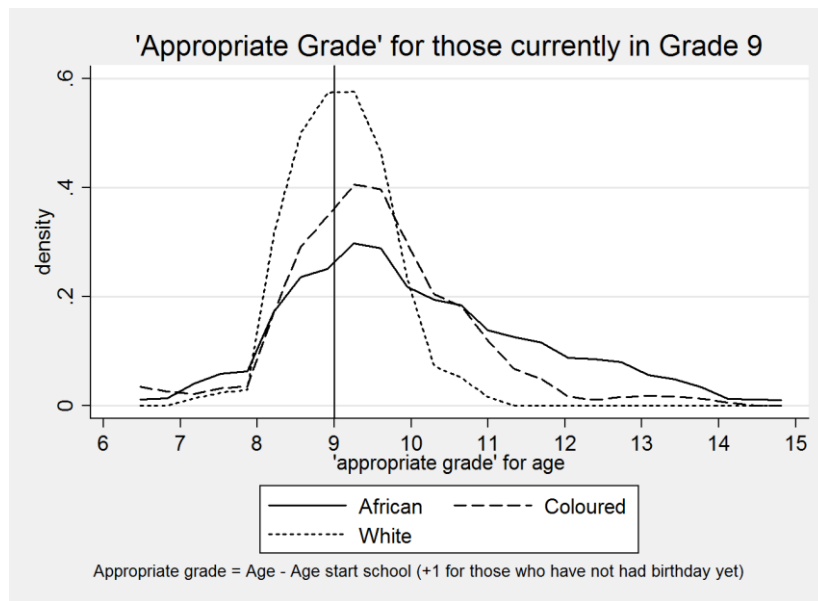Appropriate grade = Age - Age start school (+1 for those who have not had birthday yet)
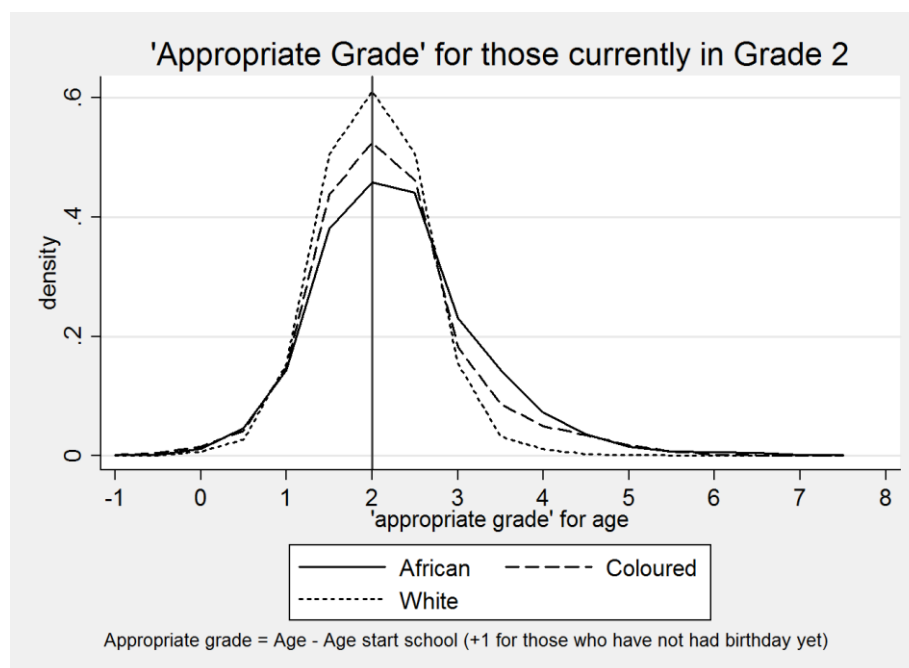
By grade 10, Figure 5b shows that this tail has increased substantially, especially for Africans and Coloureds, and that the mean has shifted to the right, indicating that the average student is below the appropriate grade for their age. Students are up to 5 years below the grade they could be given their age. This means those students who are 15 years old, the age you should be in grade 9 if you start school at age 6 and progress without failing, have fellow classmates that are up to 5 years their senior.

---

[4] Graphs restricted to include those who are a maximum of two years below and 6 years above the appropriate grade

[5] Age start school uses the data from the question which asked the age the respondent started school if available and imputes with mean start age within population group and age group for all others.

Figure 5b

'Appropriate Grade' for those currently in Grade 9

Appropriate grade = Age - Age start school (+1 for those who have not had birthday yet)

4. Conclusion

This report is organized in two parts. The first part assesses how comparable key questions in the NIDS education data are to similar questions in the LFS and CS national surveys. The second part highlights how certain questions in NIDS which are not available in any other national surveys, add value to an analysis. We find that the levels of enrolment and completed education in NIDS are comparable to the LFS data, both of which present a slightly more optimistic picture than the CS data. We focus on grade repetition and show that this topic can be approached from three separate directions in the NIDS data - using the direct question asking about grade repetition, the result (pass, fail or withdrew) of the 2007 school year or indirectly calculated using the age start and end school questions. Each dimension adds value and depth not previously possible using national survey to topic important to both researchers and policy makers.

Appendix

NIDS Section H: Education response rates - Adult sample

Variable #	Variable Description	Total # who responded	Total # who should have responded	Response Rate
h1	Highest school grade completed	15547	16932	91.8
h2	Year in which respondent completed highest grade	10709	14737	72.7
h5	Year in which respondent first attended Grade 1/Sub A	6358	14737	43.1
h6	Highest school grade in mathematics completed	12679	14737	86.0
h7	Respondent has successfully completed some from of tertiary studies	13349	14737	90.6
h8	Highest level of tertiary education completed	1377	1556	88.5
h11	The respondent repeated school grades	6281	7127	88.1
h12_1	Respondent repeated grade 'x' (1st Answer)	3524	3561	99.0
h12_2	Grade was repeated 'x'times (1st Answer)	3518	3561	98.8
h12_3	Respondent repeated grade 'x' (2nd Answer)	1189	1192	99.7
h12_4	Grade was repeated 'x'times (2nd Answer)	1177	1192	98.7
h12_5	Respondent repeated grade 'x' (3rd Answer)	335	337	99.4
h12_6	Grade was repeated 'x'times (3rd Answer)	316	337	93.8
h12_7	Respondent repeated grade 'x' (4th Answer)	122	124	98.4
h12_8	Grade was repeated 'x'times (4th Answer)	104	124	83.9
h13	Respondent pursued some form of education in 2007	6387	7127	89.6
h14	Main reason respondent did not enrol in 2007	3199	4064	78.7
h15	Education level respondent was enrolled in 2007	3045	3100	98.2
h17	Subject or Programme respondent was studying in 2007	230	240	95.8
h20_1	Amount spent on School fees	2753	3100	88.8
h20_2	Amount spent on Uniform	2588	3100	83.5
h20_3	Amount spent on Books and Stationery	2331	3100	75.2
h20_4	Amount spent on Transport to school	2266	3100	73.1
h20_5	Amount spent on allowances and other school related expenses	2205	3100	71.1
h22	Result of schooling in 2007	3058	3100	98.6
h23	Main reason respondent withdrew before completing educational year	56	57	98.2
h24	Respondent is currently enrolled	6367	7127	89.3
h25	Main reason respondent did not enrol in 2008	3573	3653	97.8
h28	Level respondent is currently enrolled in	2654	2714	97.8
h30	Subject or Programme respondent is currently studying	246	248	99.2
h31	Matric is a prerequisite for current Educational institution	247	248	99.6
h32	Respondent intends to continue at school until completion of matric	2246	2466	91.1
h33	Respondent intends to continue studying after completion of matric	2159	2466	87.6
h34	Respondent is computer literate	15602	16932	92.1
h35	Respondent has driver's license	15543	16932	91.8
h36	Respondent's reading level in home language	15583	16932	92.0
h37	Respondent's writing level in home language	15583	16932	92.0
h38	Respondent's reading level in English	15584	16932	92.0
h39	Respondent's writing level in English	15590	16932	92.1

NIDS Section C: Education response rates - Child sample

Variable #	Variable Description	Total # who responded	Total # who should have responded	Response Rate
c2	Which of the following institutions the child (under 8) currently attends	4842	4993	97.0
c3	Has this child ever attended school	4405	4661	94.5
c6	Highest school grade completed	5874	5871	100
c7	Year child first attended Sub A./Gr. One	4633	5871	78.9
c8	Attendence of pre-primary before Sub A./Gr. One	5449	5600	97.3
c9	Has the child ever repeated a grade	5860	5871	99.8
c10_g1	The child failed grade	1367	1388	98.5
c10_1	Number of times child failed grade	1367	1388	98.5
c10_g2	The child failed grade	217	217	100
c10_2	Number of times child failed grade	203	217	93.5
c10_g3	The child failed grade	43	43	100
c10_3	Number of times child failed grade	37	43	86.0
c10_g4	The child failed grade	19	19	100
c10_4	Number of times child failed grade	18	19	94.7
c11	Child attended school in 2007	5876	5871	100
c12	Main reason child was not enrolled in 2007	521	527	98.9
c13_1	Amount spent on School fees	4618	5349	86.3
c13_2	Amount spent Uniform	4542	5349	84.9
c13_3	Amount spent on Books and Stationery	3668	5349	68.6
c13_4	Amount spent on Transport to school	3613	5349	67.5
c13_5	Amount spent on other school related expenses	3615	5349	67.6
c15	Result of child's schooling in 2007	5329	5349	99.6
c16	Reason for withdrawal before end of educational year	14	14	100
c17	The child is currently enrolled in school	5766	5871	98.2
c18	Reason for not enrolling in 2008	72	80	90.0
c19	Level child is currently enrolled in	5645	5686	99.3
c20_1	What is the usual mode of transport to school has been: 1st Answer	5506	5686	96.8
c21	Time taken to reach school	5585	5686	98.2
c22	Number of learners in child's classroom	4324	5686	76.0
c23	Number of days child was absent from school,month	5491	5686	96.6