Policy note on pre-primary schooling: An empirical contribution to the 2009 Medium Term Strategic Framework

MARTIN GUSTAFSSON

Stellenbosch Economic Working Papers: 05/10

KEYWORDS: PRE-PRIMARY SCHOOLING, SOUTH AFRICA, AGE-GRADE NORMS
JEL: I21, I28

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

UNIVERSITEIT STELLENBOSCH UNIVERSITY

BER
BUREAU FOR ECONOMIC RESEARCH

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

Policy note on pre-primary schooling: An empirical contribution to the 2009 Medium Term Strategic Framework[1]

MARTIN GUSTAFSSON

ABSTRACT

Various data analysis approaches are used to gauge recent pre-primary enrolment trends in South Africa and the level of compliance with official age-grade norms in Grades R and 1. An analysis of the circumstances of Grade R learners finds that large class sizes are a problem. Two separate logit models are used to examine what factors are associated with better pre-school participation and whether participation in pre-school leads to better learning outcomes in primary school.

Keywords: Pre-primary schooling, South Africa, Age-grade norms
JEL codes: I21, I28

[1] This document was produced by the Social Policy Research Group of 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. Part of this research also serves as an input into a Public Expenditure Tracking Survey of Grade R and pre-school funding for the Departments of Education and Social Development, National Treasury and UNICEF.

1 Introduction

The 2009 Medium Term Strategic Framework (MTSF) of The Presidency includes two key goals relating to pre-school enrolment. Firstly, there is to be 'universal access to Grade R' by 2014. Secondly, the number of learners[2] aged 0 to 4 enrolled in some form of ECD should double by 2014. However, the MTSF also emphasises that these services must be of a high quality. This is obviously important, as a key aim of more formal education for learners before Grade 1 is to improve the quality of learning in Grade 1 and beyond.

This policy note aims to bring new information to bear on the pre-primary enrolment issues being debated in policy circles, largely through the use of a newly available dataset, the National Income Dynamics Study (NIDS) wave 1 dataset, compiled in 2008.

2 A summary of the policy findings

If one only considers what is explicitly labelled 'Grade R' in NIDS, then around 57% of South African children were going through Grade R in 2008. However, a much higher figure of over 80% is obtained if one asks learners currently enrolled in Grade 1 whether they underwent 'pre-primary or Grade R' before starting Grade 1. This 80% figure represents a marked improvement in a short space of time. Only 60% of Grade 4 learners say they went through some form of pre-primary schooling. Separate Annual Survey of Schools data suggest that the 57% level corresponds to formal Grade R only, whilst the 80% level includes non-formal pre-primary schooling. The extent to which primary school learners have undergone some pre-primary schooling is lowest for the poor, though even for the poor the enrolment ratios are fairly good. The ratio is 60% for the poorest income quintile and 90% for the least poor.

NIDS is able to provide new insights into the age of entry of learners into Grade 1, something that influences planning with respect to pre-primary schooling. Despite a flexible admissions policy that allows around half of parents a choice between two years of entry into Grade 1, the practices of parents and schools are even more flexible, with 35% of learners entering Grade 1 either before or after the years stipulated by the policy (the 35% is split about evenly between early and late starters).

In 2008, 1.3 million 0 to 4 year olds, or 26% of the five age cohorts, were enrolled in an institution. Doubling this would obviously mean creating an additional 1.3 million places for children in this age bracket. This becomes easier if new publicly funded Grade R places result in a shift in private spending from Grade R to schooling below Grade R. The 26% enrolment ratio for 0 to 4 year olds assumes that the category 'day-mother/ gogo' is not considered ECD for the purposes of the MTSF target. If the 'day-mother/ gogo' category were included, the enrolment ratio in 2008 would be as high as 51%.

Fees in Grade R are slightly higher than those in primary school, though for the poorest four quintiles annual Grade R fees mostly did not exceed R100 in 2007. Travelling time is slightly shorter for Grade R learners than Grade 1 learners, in general by about 5 to 10 minutes. This gap is likely to be reduced, and travelling times for Grade R learners can be expected to become a bit longer, as more Grade R shifts to public schools. The problem of very large classes seen in primary schools is also visible in Grade R, though to a lesser extent. Around 15% of Grade R learners are in classes exceeding 40 learners. On average around 1.0 day per month is missed by learners in Grade 1, and the figure is virtually the same for Grade R.

A model of the determinants of enrolment for children aged 4 to 6 confirms that enrolment levels are lower in poorer households, but also that children in rural formal areas of the

[2] In South Africa pupils are referred to as learners.

country experience particularly severe barriers when it comes to enrolling in pre-primary institutions. This suggests that special attention needs to be paid to poor children in places such as commercial farms and forestry stations.

The availability of numeracy test scores for some children in the NIDS dataset allowed for a model that would test the impact of pre-primary participation on subsequent learning. The model suggests that in better off communities, what may appear to be learning advantages linked to prior pre-primary participation may in fact be the effects of home background advantages. However, in rural informal settings the impact of pre-primary schooling appears positive and significant, even when one controls for home background effects.

Lastly, important lessons can be learnt from the NIDS survey on how official surveys could be improved to allow for a more effective and accurate monitoring of pre-primary schooling.

3 The data sources

National Income Dynamics Study (NIDS) – Wave 1 of 2008

This study collected data from a nationally representative sample of households in 2008, mainly during February to July, though some follow-up up of refusals occurred in September. Data were collected from 7,305 households and 28,255 individuals. Of the 28,255 individuals, 7,305 had their details collected through a child questionnaire as they were aged 14 or below. For the remainder, a separate and rather different adult questionnaire was used. For 208 children and 1,246 adults (2.1% and 6.7% respectively) no detailed data were collected due to refusals or problems in locating the person or someone who could respond on his behalf. The NIDS dataset attaches two weights to each household, a design weight and a post-stratified weight. The latter, which is constructed to optimise agreement with the 2008 mid-year population estimates of Statistics South Africa (Stats SA), including Stats SA's breakdown by age, race and sex, was used for this analysis. The NIDS dataset and supporting documentation are available at http://www.nids.uct.ac.za.

General Household Survey – 2008

This source was used largely to verify patterns seen in the NIDS data. Stats SA's 2008 General Household Survey (GHS) dataset, which is part of an annual series started in 2003, is based on a collection from around 25,000 households and 95,000 individuals.

Community Survey – 2007

This large sample-based collection by Stats SA covered around 250,000 households and 950,000 individuals. Collection occurred during February of 2007. It was partly aimed at taking the place of the national census that had originally been planned for 2006.

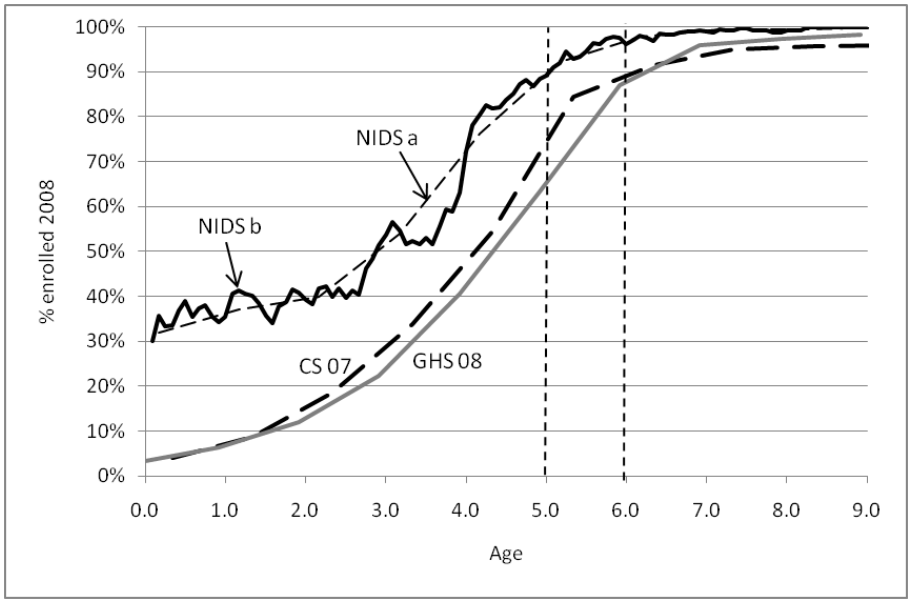
4 Enrolment levels and trends

This policy note relies mostly on data captured through the child questionnaire of the NIDS survey. These data include three key pieces of information which are of great importance for understanding pre-school enrolments, but are seldom if ever captured in nationally representative household surveys in South Africa. Firstly, the month and year of each child's birth is captured, as opposed to just the age on the survey day. This allows for a proper examination of the relationship between age and the level of enrolment. Secondly, the child questionnaire uses 'Grade R' as a specific enrolment category, as opposed to lumping this category together with other pre-primary categories. Not having Grade R as a separate category has been problematic in other datasets, yet as will be seen below even identifying learners as explicitly being in Grade R in a survey still leaves many questions unanswered.

Thirdly, whether primary school children did or did not receive pre-primary schooling is captured.

The first graph below compares enrolment in any institution of education or early childhood care across three data sources. These data sources are the 2007 Community Survey (CS 07), NIDS 2008 (NIDS a) and the General Household Survey (GHS 08). The two non-NIDS curves use age in years on the survey day, whilst for NIDS the age respondents would have in April 2008 was used (the curve 'NIDS b' is discussed below). The principal reason why the 'NIDS a' curve is so much higher than the other two at the lower ages is that NIDS counted enrolment with a 'day-mother/ gogo', something the other surveys do not explicitly do. The effect of this in the NIDS data stretch up to age 6. But even above age 6, the NIDS data result in the highest enrolment levels of the three, the difference being around 2 to 4 percentage points. It is thus possible that NIDS may present a slight over-estimate of enrolment. Unfortunately there is no easy way of determining to what extent this may be the case.

Figure 1: Overall enrolment of young children

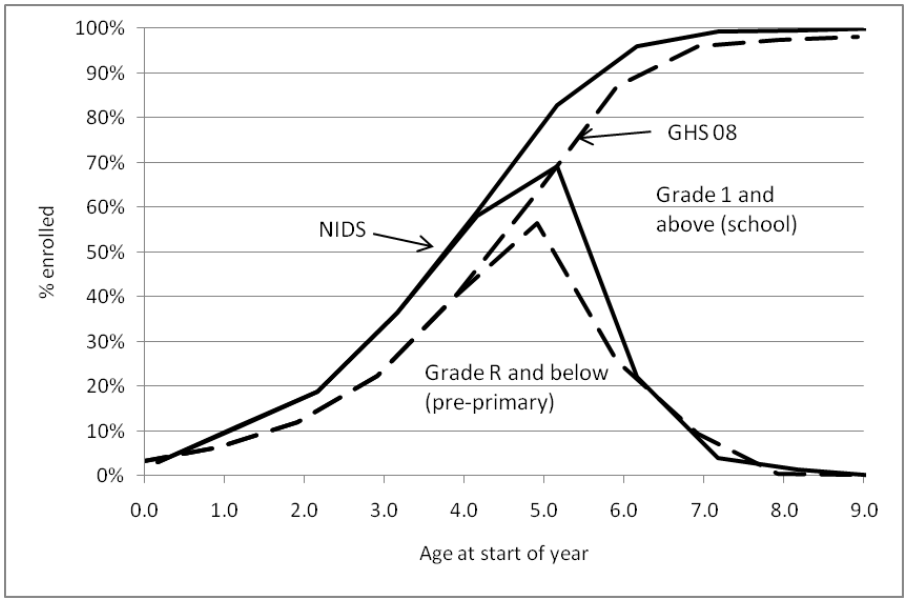


Sources: National Income Dynamics Study, wave 1; 2007 Community Survey; 2008 General Household Survey.
Note: To optimise comparability between the three variables not containing month details (NIDS a, CS 07, GHS 08) the plot against the horizontal axis is adjusted according to the median month in which the survey occurred (or, in the case of NIDS, for which the month of the interview was available, the mean month). The adjustment occurs in such a way that a survey month of June would be plotted against the integer (e.g. 5.0) whilst a survey month before June would appear to the right of the integer (children are younger and so their ages must be adjusted upwards), and a survey month after June would appear to the left of the integer (children are older and so their ages must be adjusted downwards). The assumed survey months were April for NIDS, July for GHS and February for CS.

The following graph takes the comparison a step further by excluding the day-mother category from the NIDS data and separating, in both the NIDS and GHS 2008 datasets, school (Grade 1 and above) from pre-primary (Grade R and below). Both sources display a similarly shaped peak representing pre-primary enrolments, though the NIDS peak is larger than the GHS peak. The left-right alignment would to a large degree be influenced by measurement inconsistencies relating to reported age and the survey date, but the difference in the heights of the peaks are likely to be related to sampling issues, and the way questions about the different levels of education were interpreted. Again, there is no easy way of determining which of the two sources represented in Figure 2 is a better reflection of reality in 2008. Yet

the graph serves as an important reminder that information emerging from any source should be interpreted with care, and may be based on slight under- or over-estimations to the degree illustrated in Figure 2[3].

Figure 2: Split across primary and pre-primary



Sources: National Income Dynamics Study, wave 1; 2008 General Household Survey.

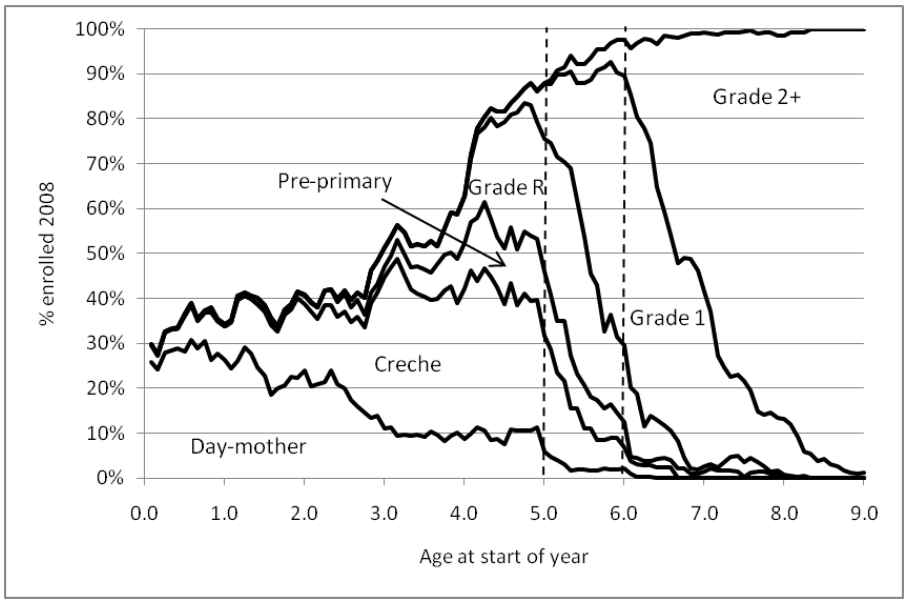
The very unsmooth 'NIDS b' curve in Figure 1 represents valuable policy information that is not available from all or most other nationally representative household datasets in South Africa. Because NIDS collected the month and year of birth of each child, it is possible to determine the age of each child when she started her current grade, i.e. in January of 2008. The 'NIDS b' curve is plotted against the point on the horizontal axis representing one's precise age at the beginning of January 2008, in other words at the beginning of the school year. To illustrate, if one turns 5 in December 2007, then one is considered to be 60 months old at the beginning of January 2008, or 5 years and zero months, or 5.0 years. Given that national admissions policies are based on when a child's birthday is, this information is important. Figure 3 takes the 'NIDS b' curve from the earlier graph and breaks it down by level of education. It is clear that the 'day-mother' category accounts for much of the enrolment up to age 5. It is also clear that enrolment patterns do not neatly follow the admissions policies. If these policies were followed strictly, then no-one born in July 2002 or later would be in Grade 1 in 2008[4]. Put differently, no-one aged 5 years and 5 months (or 5.4 years) or younger in the graph would be in Grade 1 (keeping in mind that the graph represents one's age at the beginning of January). Clearly, a considerable number of learners are in Grade 1 despite being younger than the threshold age. The same can be said of Grade R, where the minimum would be 4.4 years in the graph.

[3] Particular care should be taken in interpreting education level information derived from the 2007 Community Survey. This dataset displays distributions across education levels that are very unlike those seen in other datasets. For example, there are almost no children aged 6 and above in pre-primary institutions, and there are no learners in primary school above age 14. It appears that with respect to education level, age was used to normalise values, resulting in values that are inconsistent with all comparable data, both household data and EMIS data from the Department.

[4] The policy appears in the Education Laws Amendment Act (Act 50 of 2002), which amended the earlier policy of Government Notice 2433 of 1998, which allowed for less flexibility.

At this point it is possible to estimate some magnitudes with respect to the MTSF goal of doubling the enrolment of children aged 0 to 4 years. Using the patterns reflected in Figure 2 by the NIDS curve produces an estimate of 26% of children aged 0 to 4 enrolled, or around 1.3 million. The patterns reflected in Figure 3, which count the 'day-mother' category as a legitimate form of enrolment, produce figures of 51% and 2.6 million. Roughly, then, attaining the MTSF target would imply creating anywhere between 1.3 million and 2.6 million extra places for children in institutions. One's definition of what constitutes legitimate and useful ECD from the perspective of improving the learning opportunities of, in particular, poorer learners would inform which magnitude of improvement was necessary.

Figure 3: Split across all enrolment types

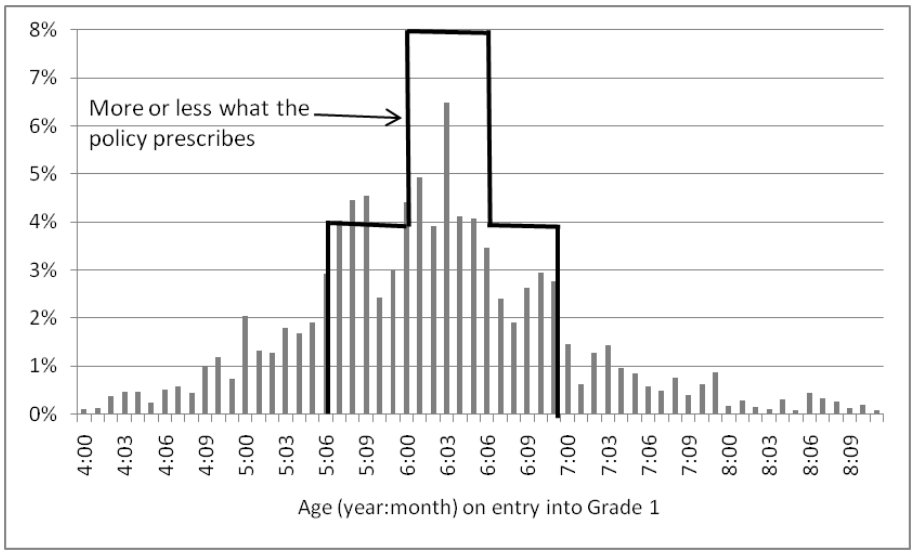


Source: National Income Dynamics Study, wave 1.

Both the quantity and type of pre-primary schooling that is needed is influenced by how early children begin Grade 1. NIDS asked learners already enrolled in Grades 1 and above in which year they started Grade 1. This allowed for the construction of a profile of precise age of entry as in Figure 4. Figure 4 uses data from all children who began Grade 1 in 2006, 2007 or 2008. During these years the admissions policies were stable. The large deviations from the official admissions policy mentioned above are visible. This policy essentially gives parents whose child has a birthday in the first half of the year the choice of two different years for entry into Grade 1, whilst those parents whose child has a birthday in the second half of the year must enter Grade 1 in a specific year, and there is no choice. If one assumed that half of the parents who do have a choice chose the earlier year whilst half chose the later year, then one would expect a profile of precise age entry as illustrated by the stepwise pyramid outlined with thick lines in Figure 4. No-one would enter Grade 1 with an age of less than 5 years and 6 months, and no-one would enter Grade 1 with an age of more than 6 years and 11 months. In fact, only 65% of learners enter Grade 1 at the correct age. Almost exactly half of the remaining 35% enter Grade 1 too early, and half of 35% enter too late. It is widely believed that where Grade R is not offered in schools, schools allow children to enter Grade 1 a year earlier. The data do not support such a hypothesis. When only those learners who have been to pre-school are considered, the pattern hardly changes (65% becomes 68% and the early and late entrants comprise equal parts of the remainder). Clearly, what policy action should be taken with respect to the age of admission into Grades 1 and R (the policy for Grade R is like that for Grade 1 with all criteria pushed back by 12 months) must be informed not just by whether the existing policy is being complied with, but also by a range of factors such as the reasons for early and late entry, the peer effects of having unevenly aged learners, the cognitive capacity

of differently aged learners with respect to the curriculum, and so on. The NIDS data do not allow a proper investigation into these matters.

Figure 4: Age of entry into Grade 1



Source: National Income Dynamics Study, wave 1.

In order to assess how close the country is to universal Grade R, an aim of the MTSF, it is necessary to know what proportion of young South Africans go through Grade R. Can the total Grade R enrolment figures indicated by NIDS be regarded as a reliable indicator of the proportion of learners getting Grade R? Arguably yes, because there is almost no repetition in Grade R, so counting existing Grade R enrolments does not imply double counting. NIDS asks of enrolled learners which grades they have repeated and how many times. These data allow one to obtain a percentage by dividing all grade years which are repeats by all grade years that have occurred. This percentage of grade years which are repeated comes to 18% for Grade 1, 12% to 13% for Grades 2 and 3, and only 1% for Grade R. Ignoring repetition in Grade R would thus not lead to major distortions. The gross enrolment ratio (GER) for Grade R can thus be considered a reliable indicator of the percentage of children who get to take Grade R. This ratio is simply the total number of Grade R learners (*R* below) divided by the size of the grade-appropriate age cohort or the number of five year olds in the population (*P* below).

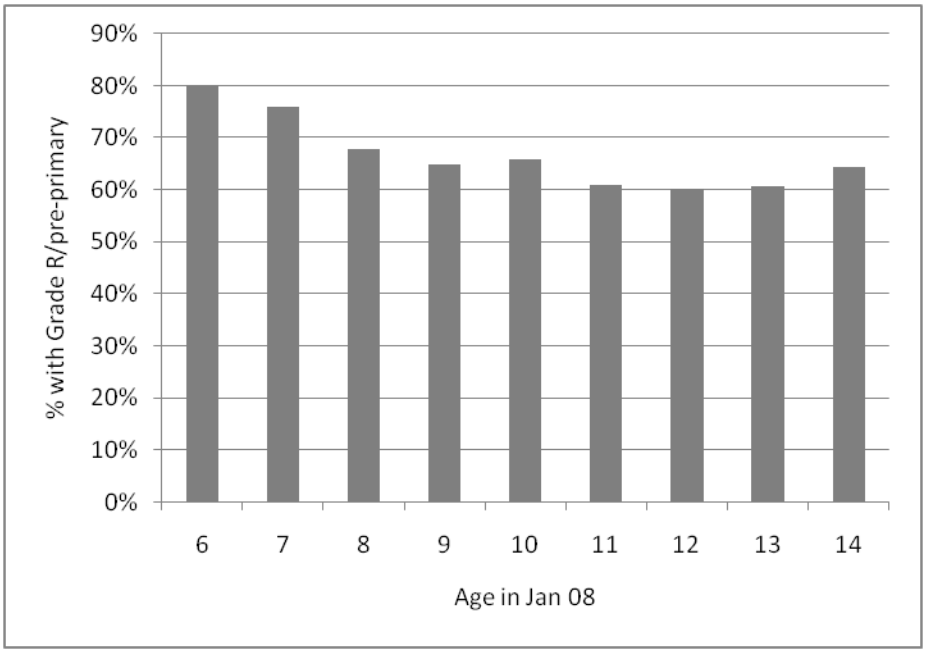
$$g = \frac{R}{P}$$

If only Grade R enrolments in the age range of 4 to 8 years are considered, then the result is a percentage of 57%. If all Grade R enrolments in the range 0 to 9 years are considered, the result is 62%. Restricting the calculation to 4 to 8 year olds seems appropriate. It is unlikely that learners outside this range are truly enrolled in Grade R. The 57% obtained from NIDS can be compared to a 52% figure for 2008 where official Grade R enrolments, in public and independent schools, are divided by one age cohort. Importantly, the 52% figure does not take into consideration Grade R offered outside proper schools (schools with at least Grade 1), for instance in pre-primary schools (Department of Education, 2008; Stats SA, 2008). The Annual Survey of Schools of the Department of Education asks how many Grade 1 learners in public and independent schools received 'formal pre-primary or Grade R' and how many received 'non-formal pre-primary or Grade R'. The data from that source for 2008 indicate that 59% had formal pre-primary, 19% had non-formal pre-primary and 22% had no pre-primary (own analysis of the dataset). This seems to confirm that what is considered Grade R

in NIDS is mostly just formal Grade R, and that more informal services fall under some other NIDS category.

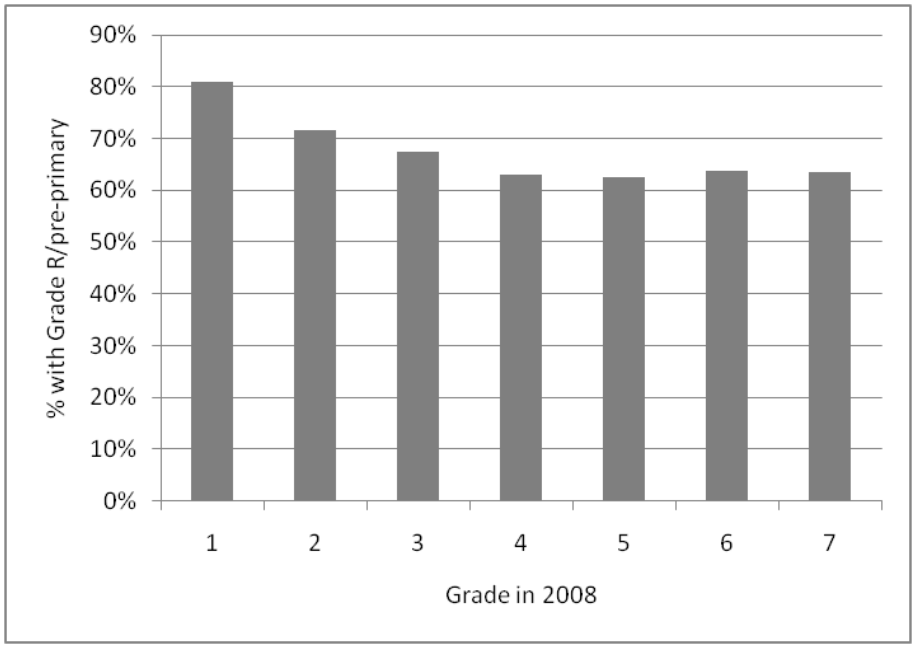
The following two graphs illustrate what percentage of learners who are enrolled in at least Grade 1 received 'pre-primary or Grade R before going to Grade 1/Sub A'. Figure 5 provides an analysis by age and Figure 6 by grade. Some kind of pre-primary schooling was received by 80% of the youngest learners in the schooling system and the trend seems to have improved markedly in recent years. The 80% level is very close to the sum of the formal and non-formal pre-primary values found in the Annual Survey of Schools dataset (59% plus 19%). Figure 6 suggests there was an improvement in coverage in the year immediately preceding Grade 1 from around 60% to 80% in the period 2004 to 2007. In fact, the period 2003/04 to 2006/07 saw very large increases in provincial spending on ECD of 59% in real terms, suggesting that the recent improvements are to a large extent the result of public interventions (National Treasury, 2007: 126).

Figure 5: Learners who have pre-school (by age)



Source: National Income Dynamics Study, wave 1.

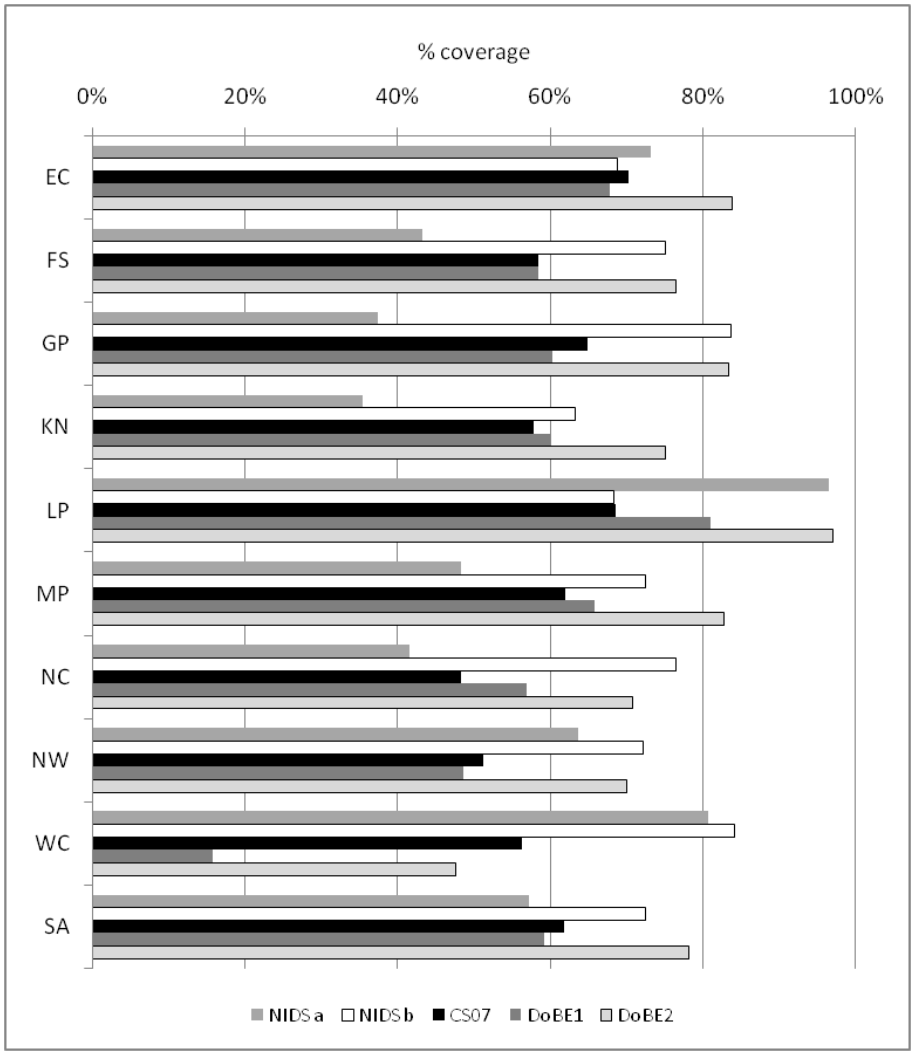
Figure 6: Learners who have pre-school (by grade)



Source: National Income Dynamics Study, wave 1.

The following graph offers a comparison of various measures, by province. 'NIDS a' is based on current enrolment in what is explicitly referred to as Grade R – here the national average, already discussed above, is 57%. 'NIDS b' is based on data on 7 to 9 year olds who were enrolled in at least Grade 1 in 2008 and indicated whether or not they had received Grade R or pre-primary schooling before Grade 1 – the national figure here is 73%. 'CS07' refers to the percentage of 4 to 6 year olds *not* in schooling (Grade 1 or above) who are enrolled in a pre-primary education institution according to the 2007 Community Survey – here the national figure is 62%. One should keep in mind that the problem with this measure is that it would double count learners who experience more than one year of pre-primary schooling. 'DoBE1' is the proportion of Grade 1 learners in public and independent schools who received formal Grade R according to the 2008 Annual Survey of Schools – the national figure is 59%. Finally, 'DoBE2' is the proportion of Grade 1 learners in public and independent schools who received any pre-primary schooling, whether formal or non-formal, according to the 2008 Annual Survey of Schools – the national figure is 78%.

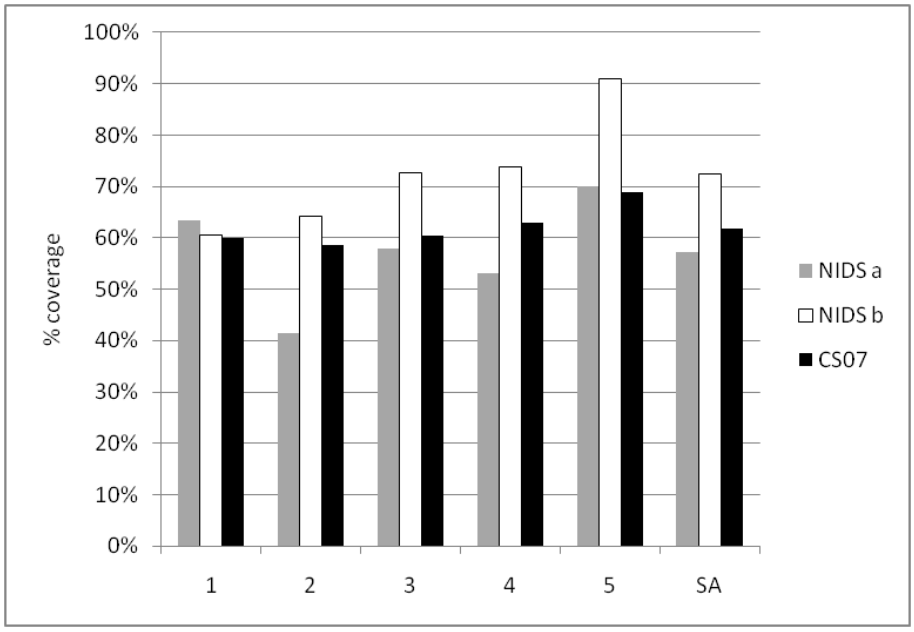
Figure 7: Pre-school coverage by province



Sources: National Income Dynamics Study, wave 1; 2007 Community Survey; Annual Survey of Schools dataset 2008.

The following graph represents three of the measures from the previous graph, by income quintile.

Figure 8: Pre-school coverage by income quintile



Sources: National Income Dynamics Study, wave 1; 2007 Community Survey.
Note: Income quintile 1 has the lowest incomes. The quintiles are quintiles of all individuals in the population where ranking is according to total household income.

It appears as if there are two possibilities for a Grade R enrolment ratio, depending on whether one wants to count only formal Grade R, or any pre-school attendance before Grade 1, whether this be in a formal or non-formal institution. Importantly, the NIDS dataset and the Annual Survey of Schools appear to agree fairly closely with respect to both possible enrolment ratios. Counting only formal Grade R gives a percentage of around 57%, and including non-formal pre-primary increases the percentage to around 80%. The higher value would in fact be above 80% as the 80% level reflected the earlier Figure 5 is in fact largely a reflection of the situation in 2007, and we can expect year on year improvements. Real growth in provincial budgets for ECD beyond 2006/07 exists in every year, and the total amount of provincial spending is expected to quadruple in real terms between 2006/07 and 2011/12 (National Treasury, 2009: 160).

The NIDS sample is too small to provide highly reliable statistics at the provincial level for many variables. It is thus important to compare provincial NIDS figures to comparable figures from larger samples. Figure 7 suggests that coverage in the year preceding Grade 1 is indeed exceptionally high in Eastern Cape and Limpopo, even if one considers just formal Grade R ('DoBE1'). Given the levels of poverty in these two provinces, and the widely held belief that poverty inhibits pre-primary participation, this finding is somewhat surprising. Figure 7 moreover seems to suggest that participation in pre-primary schooling is low in Western Cape, despite that province's socio-economic advantage. The Community Survey ('CS07') source indicates a level of enrolment below the national average, and the Annual Survey of Schools source yields statistics that are the lowest for all provinces.

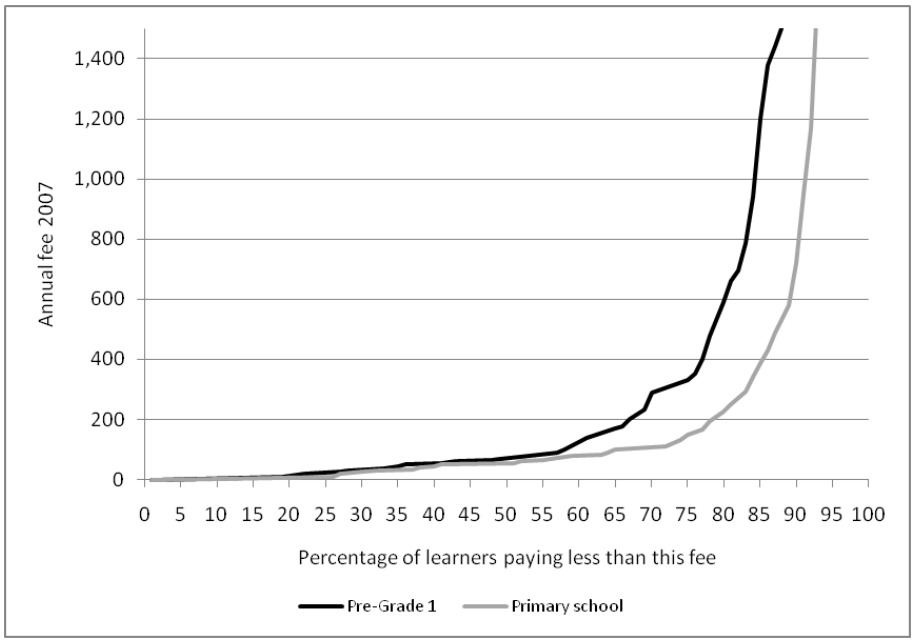
Figure 8 reflects a substantial positive association between income and 'NIDS b', or the percentage of enrolled primary learners who have been through some pre-primary schooling. A figure of 60% in quintile 1 compares with a figure of 90% in quintile 5.

5 The nature of pre-primary participation

A NIDS question asks what the annual school fee was in the previous year. The distribution of values for primary school learners (learners in Grades 1 to 7) and for pre-primary learners

is illustrated below. As one may expect, given the more private nature of pre-primary schooling when compared to primary schooling, fees in pre-primary are slightly higher. However, as illustrated in Figure 10, for 80% of the country the fee was below R100. This is well below the actual cost of even the cheapest possible service, suggesting a high prevalence of public subsidies (to compare, it costs around R6,000 per year for the state to educate every Grade 1 learner).

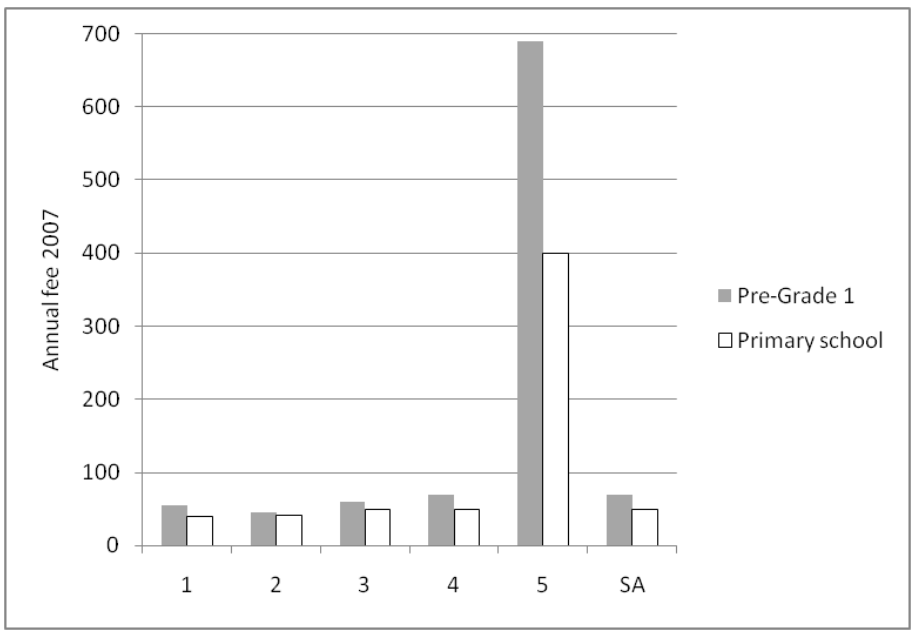
Figure 9: Distribution of fees



Sources: National Income Dynamics Study, wave 1.

.

Figure 10: Fees by income quintile

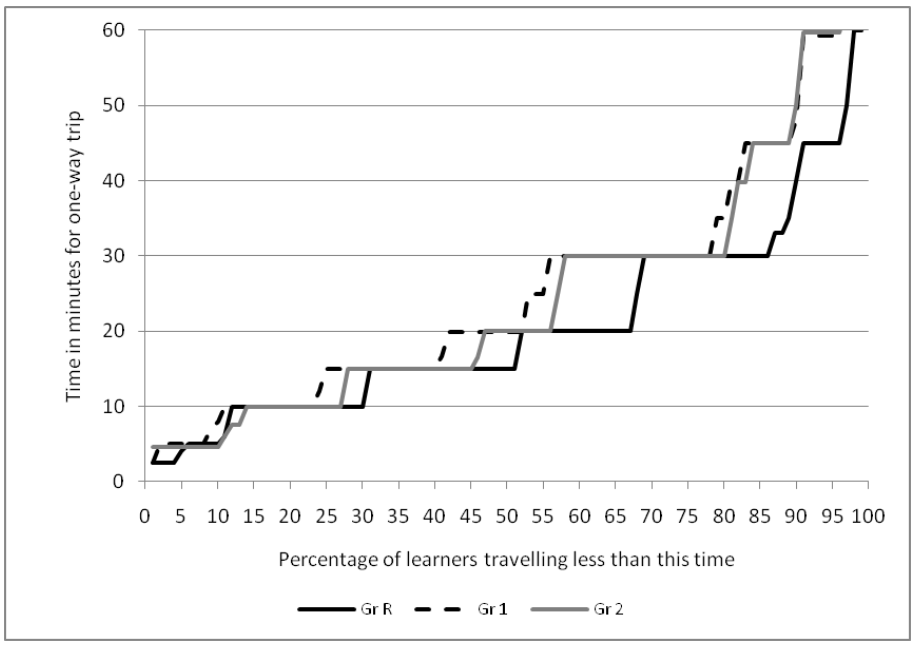


Sources: National Income Dynamics Study, wave 1.

.

Figure 11 indicates that it takes learners a bit less time to reach pre-primary services than primary services. The difference is not very large, often just 5 to 10 minutes, and it would be the result of the existence of pre-primary sites that are more numerous and geographically closer to people's homes than primary schools. A shift towards more Grade R in public schools is likely to increase the travelling time for the affected learners by a small amount.

Figure 11: Travelling time to school

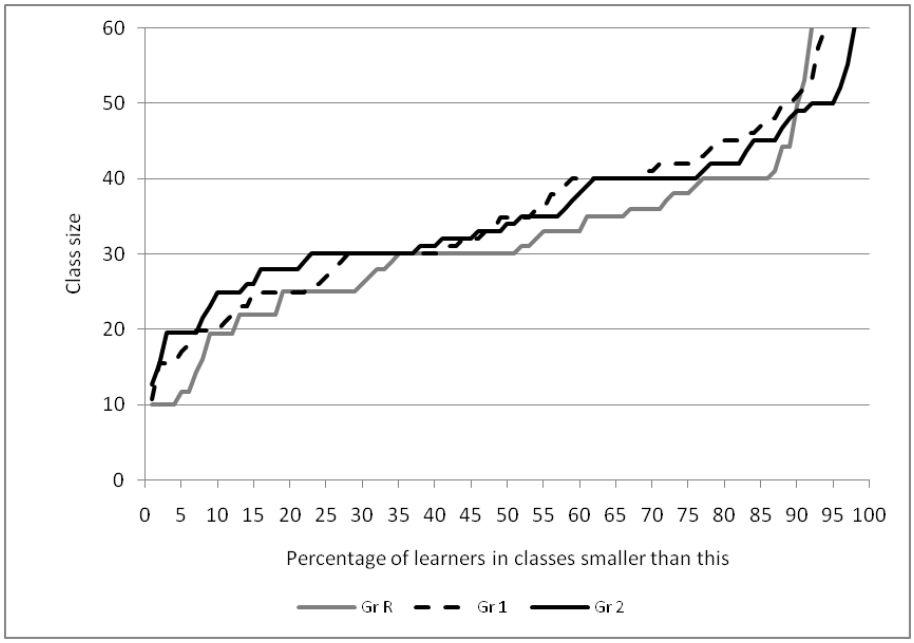


Sources: National Income Dynamics Study, wave 1.

.

Class sizes in many schools are unacceptably high in South Africa (though learner/educator ratios for the system are relatively equitably distributed across institutions and much lower than the highest class sizes). The next graph illustrates that around 30% of Grades 1 and 2 learners are in classes with more than 40 learners. Over 10% are in classes with more than 50 learners. The situation in Grade R is similar to that in the higher grades, but slightly better. For instance, only around 15% of Grade R learners are in classes exceeding 40 learners.

Figure 12: Class size

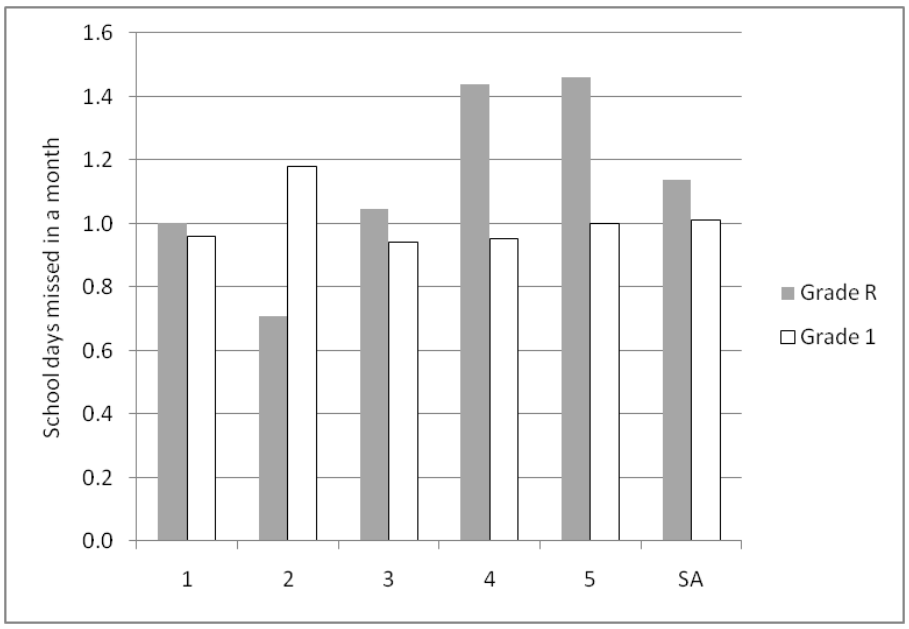


Sources: National Income Dynamics Study, wave 1.

.

Attendance problems as reported by households in terms of school days missed (where schooling was occurring for other learners) in the month prior to the survey date reveals no great differences between Grade R and Grade 1. If anything, more days are missed by Grade R learners from better off households, perhaps reflecting the greater ability of these households to keep the learner at home, and a smaller dependence on possible meals served by the institution.

Figure 13: School days missed



Sources: National Income Dynamics Study, wave 1.

.

6 Reasons why children participate in pre-primary schooling

The NIDS dataset allows for a relatively robust model explaining enrolment in an institution by young children. Such a model can assist in the design of strategies aimed at improving this enrolment. The model presented below is one of several that were tested. Below, the dependent variable is a 0-1 binary variable where 1 means the child attends any education institution other than a 'day-mother'. Only children aged 4 to 6 in January 2008 were considered. As seen in Table 1, 75% of these children were enrolled somewhere. Explanatory variables were as follows: the monthly income of the household; whether the child lived in an informal rural area (ex-homeland); whether the child lived in a formal rural area (commercial farming area, mainly); whether the child was a girl; whether the child was aged 5 or 6 (meaning the reference age was 4); whether the child was the eldest child in the household (where anyone aged 14 and below was considered a child); the child's weight divided by the child's height (the two variables were physically measured as part of the survey); whether the child suffered from a serious illness; the years of education of the most educated member of the household; and whether the biological mother of the child lived with the child.

Table 1: Statistics for 4 to 6 year olds

	Mean	10th percentile	Median	90th percentile
Is enrolled	0.75			
Monthly household income	5244.89	742.38	2400.00	11280.93
Is in informal rural area	0.42			
Is in formal rural area	0.06			
Is a girl	0.48			
Is aged 5	0.06			
Is aged 6	0.06			
Is eldest child	0.49			
Weight for height	0.23	0.14	0.20	0.33
Has serious illness	0.06			
Highest education level in household	10.22	7.00	11.00	12.00
Mother lives at home	0.79			

Source: National Income Dynamics Study, wave 1.
Note: Variables without distribution statistics are 0-1 dummy variables.

Key results of the model appear in the next table (full results appear in the appendix). The explanatory variables were able to explain 23% of the variance in the dependent variable. Many variables do not significantly explain enrolment even at the 10% level of significance, yet they were retained within the model as their non-significance is telling for policymakers. As one might expect, better household income is significantly associated with better enrolment. Specifically, a R1,000 increase in a household's monthly income is associated with a 0.5 percentage point increase in a child's probability of being enrolled. This underlines the importance of a pro-poor targeting strategy in the expansion of publicly funded pre-school services. Being in an ex-homeland appears to play no role, but being in a formal rural area does. Specifically, the probability of being enrolled for children in such an area is 7 percentage points lower than for children in urban areas (urban areas can be considered the reference area). Importantly, the child's gender is not significantly linked to his or her probability of being enrolled. Age, as one would expect, is highly significantly linked to enrolment. The older the child, the greater the chances of her being enrolled. Being the eldest child is significantly and negatively associated with enrolment. Parents may not always be aware of the importance or availability of pre-school if they have not recently had an older child in pre-school (or school). Neither of the two health measures in the model appear significantly linked to enrolment. More educated adults in the household are significantly associated with better enrolment, which underlines the importance of educating less educated parents about their rights and responsibilities with respect to pre-school. Finally, if the child's mother lives with the child this does not appear significant in the model. To sum up, a

strategy that targets spending and advocacy towards poorer and less educated households is important. This is in fact what the current policies are geared towards[5]. It seems as if special attention needs to go towards providing access to pre-primary services for the poor working in formal sector sites such as commercial farms and forestry stations.

Table 2: Determinants of enrolment amongst 4 to 6 years olds

	Change in value	Change in probability	Significance
Monthly household income	+1000	0.0050	**
Is in informal rural area	0 to 1	-0.0076	
Is in formal rural area	0 to 1	-0.0720	**
Is a girl	0 to 1	-0.0099	
Is aged 5	0 to 1	0.0857	***
Is aged 6	0 to 1	0.1759	***
Is eldest child	0 to 1	-0.0306	*
Weight for height	+0.01	0.0003	
Has serious illness	0 to 1	-0.0421	
Highest education level in household	+1	0.0090	***
Mother lives at home	0 to 1	0.0035	

*Note: ***, ** and * indicate that the coefficient is significant at the 1%, 5% and 10% levels respectively.*

7 The value of pre-primary schooling for subsequent learning

The 2008 wave of NIDS included the testing of numeracy competencies amongst respondents, using four different tests with differing levels of difficulty. The use of the test with child respondents is indicated in the next table. Most children did not write the test. For instance, amongst 12 to 14 year olds only 39% did. One can regard the sample of test-writers as non-random as households were asked whether they would like to participate in the test. Analysis using the numeracy scores must thus be interpreted cautiously. Reportedly the choice of test in the NIDS survey depended on the grade of the child. Children in a higher grade were meant to write a more difficult test. At least for the children aged 14 and below this did in fact not seem to be the case. A comparison of test taken against the current grade of the learner revealed no clear pattern. The fact that the criteria governing the choice of test are not clear underlines the need to view the analysis with caution.

Table 3: Numeracy test participation amongst children

Age	No test	Test 1 (least difficult)	Test 2	Test 3	Test 4 (most difficult)
6	571	1	1		
7	628	2			
8	618	7			
9	622	13	2		
10	609	7	17		
11	553	35	82	1	
12	387	58	180	16	1
13	430	36	146	77	
14	248	10	69	72	5
Total	4,666	169	497	166	6

Despite the drawbacks with the data, NIDS offers an exceptional opportunity to explore the conditional correlation between participation in pre-school (children in school are asked whether they went through pre-school) and achievement in the numeracy test. This could

[5] See for instance Government Notice 26 of 2008.

throw some light on the impact of pre-primary schooling on subsequent learning in school. Table 5 reflects the results from two models that seemed informative. Descriptive statistics on the variables used (or derived) are provided in the next table. Only scores from test 2 were considered, and these were transformed so their mean was 500 and standard deviation 100. (Using scores from more than one test and dummy explanatory variables for which test was written was attempted, but this yielded results that were difficult to interpret.) Four dummy variables for grade participation were used, the reference value being participation in Grades 3 or 4. Four dummy variables indicating the type of pre-school were developed by multiplying whether the child had been to pre-school by dummies on the four types of areas considered in the survey. Of course this approach assumes that the child went to a pre-school in the type of area in which he currently lives, which may not be true. Yet this classification seemed preferable to having all pre-primary schooling covered by just one variable. Three variables relating to the home background were selected: number of adults in the home, where an adult is anyone aged 15 and above; household income; and school fees paid in the previous year. The latter two are commonly used as proxies for socio-economic status. As explained below, the number of adults in the home appeared to be positively correlated with poverty.

Table 4: Statistics for Grades 3 to 10 learners

	Mean	10th percentile	Median	90th percentile
Score	490.30	396.39	496.26	622.13
Is in Grade 5	0.10			
Is in Grade 6	0.27			
Is in Grade 7	0.44			
Is in Grade 8 and above	0.18			
Class size	39.09	23.00	41.00	54.00
Got pre-school in rural formal area	0.02			
Got pre-school in rural informal area	0.22			
Got pre-school in urban formal area	0.28			
Got pre-school in urban informal area	0.06			
Adults in household	3.24	1.00	3.00	6.00
Household income (log of monthly)	7.84	6.70	7.78	8.95
School fees (log of annual)	3.54	0.00	4.09	6.21

Source: National Income Dynamics Study, wave 1.
Note: Variables without distribution statistics are 0-1 dummy variables.

Model 1 below explores just the relationship between the numeracy score and structural education factors, without considering the quality of educational inputs or home background (further details on this model appear in Table 7 in the appendix). Although the grade variables do not individually display a statistically significant association with the score, jointly they are significant at the 1% level, in other words highly significant (this was tested using a Wald test). One would of course expect a higher grade to be associated with better performance in the same test. Class size is a highly significant predictor in the model. Partly, this variable would reflect socio-economic factors as more disadvantaged and remote schools have difficulties attracting teachers and raising fee revenue in order to pay for additional teachers beyond the publicly funded staff establishment. With respect to pre-primary participation, a model like Model 1 with a single pre-primary dummy variable would display a highly significant association between pre-primary participation and the test score. When four dummies for the four types of areas are used (as in Model 1), pre-schooling in a formal urban area emerges as a highly significant predictor of test performance and pre-schooling in an informal rural area as significant at the 10% level. Both of these significant associations are positive.

Table 5: Pre-primary schooling as a determinant of school learning

	Model 1		Model 2	
	Coef-ficient	Signif-icance	Coef-ficient	Signif-icance
Is in Grade 5	-1.07		-1.16	
Is in Grade 6	-19.00		-21.99	
Is in Grade 7	29.64		23.72	
Is in Grade 8 and above	39.59		32.27	
Class size	-0.72	***	-0.77	***
Got pre-school in rural formal area	24.73		-6.76	
Got pre-school in rural informal area	21.97	*	23.01	**
Got pre-school in urban formal area	31.45	***	4.46	
Got pre-school in urban informal area	-12.19		-21.90	
Adults in household			-5.86	**
Household income (log of monthly)			19.45	***
School fees (log of annual)			7.62	***

*Note: ***, ** and * indicate that the coefficient is significant at the 1%, 5% and 10% levels respectively.*

All three home background factors, when brought into the model (see Model 2), are statistically significant (see Table 8 in the appendix for details). The level of education of adults was used in various ways in testing Model 2, but household income seemed consistently a better predictor of the capabilities of adults in the household. The negative association between the number of adults in the household and the child's performance in the test appeared significant in various permutations of Model 2, and more significant than the presence of the biological mother or father in the household. The adult count variable can largely be considered a proxy for overcrowding, unemployment and a general sense of poverty within the household. The inclusion of home background factors renders the pre-school variables less significant than they were in Model 1. If the four dummy variables were replaced by a single pre-primary variable, this single variable would not appear significant in Model 2. Of the four, only the rural informal variable remains statistically significant (and positive). The urban formal pre-primary variable is no longer significant, and the magnitude of its association has shrunk, suggesting that in Model 1 this variable was largely masking the effects of socio-economic advantages in the home.

What conclusions can be drawn from the above? Within urban formal contexts, it will appear within schools as if learners in primary school who have received pre-primary schooling perform better than those who have not. When such observations are made, it is important to take into account the fact that more educated and advantaged households are more likely to send their children to pre-school, and that some of what appears to be the benefits of pre-school are in fact the benefits of home background factors. In rural informal contexts, on the other hand, the positive impact on pre-schooling seems to be more robust and independent of home background. This underlines the importance, expressed in existing policies, of making pre-schooling available to the rural poor. Of course the above analysis was performed using a small and non-random sample. The overall explanatory power of the two models is moreover low – 7% and 14% for Models 1 and 2. The analysis presented here should thus be considered a contribution to the debate on the value of pre-primary schooling and not a definitive analysis which on its own serves as a basis for hard policy conclusions.

References

Department of Education (2008). *School realities 2008*. Pretoria.
National Treasury (2007). *Provincial budgets and expenditure review 2003/04 to 2009/10*. Pretoria.
National Treasury (2009). *Provincial budgets and expenditure review 2005/06 to 2011/12*. Pretoria.
Statistics South Africa (2008). *Mid-year population estimates 2008*. Pretoria.

Appendix

The details of the logit model referred to in section 6 above appear below.

Table 6: Logit model details for children aged 4 to 6

n							1348
Pseudo R squared							0. 2286
	Coef.	Std. Err.	Z	P>z	LB	UB	dy/dx
Monthly household income	0.000	0.000	2.480	0.013	0.000	0.000	0.0000
Is in informal rural area	-0.125	0.245	-0.510	0.611	-0.605	0.356	-0.0076
Is in formal rural area	-0.852	0.396	-2.150	0.031	-1.629	-0.076	-0.0720
Is a girl	-0.163	0.218	-0.750	0.453	-0.590	0.263	-0.0099
Is aged 5	1.718	0.246	6.980	0.000	1.236	2.201	0.0857
Is aged 6	3.503	0.396	8.850	0.000	2.727	4.279	0.1759
Is eldest child	-0.460	0.268	-1.710	0.087	-0.986	0.066	-0.0306
Weight for height	0.505	0.950	0.530	0.595	-1.357	2.366	0.0306
Has serious illness	-0.557	0.486	-1.140	0.252	-1.510	0.396	-0.0421
Highest education level in household	0.148	0.038	3.900	0.000	0.073	0.222	0.0090
Mother lives at home	0.056	0.268	0.210	0.833	-0.468	0.581	0.0035
Constant	-0.700	0.545	-1.280	0.199	-1.768	0.368	0.0000

Note: Dependent variable is a 0-1 binary variable with 1 meaning the child (aged 4 to 6) is enrolled in an institution. Lower and upper bounds use the 5% level. dy/dx is the marginal effect on the overall probability as estimated by Stata's mfx compute command.

The details of the two models summarised in Table 5 appear in the following two tables. The first is a model without the home background variables whilst the latter includes these variables.

Table 7: Linear regression model details Grades 3 to 10 learners (1)

n						478
R squared						0.0904
Adjusted R squared						0.0729
	Coef.	Std. Err.	t	P>|t|	LB	UB
Is in Grade 5	-1.073	41.670	-0.030	0.979	-82.956	80.811
Is in Grade 6	-19.003	40.087	-0.470	0.636	-97.776	59.770
Is in Grade 7	29.644	39.772	0.750	0.456	-48.510	107.797
Is in Grade 8 and above	39.594	40.526	0.980	0.329	-40.042	119.230
Class size	-0.723	0.253	-2.860	0.004	-1.220	-0.225
Got pre-school in rural formal area	24.730	28.121	0.880	0.380	-30.530	79.989
Got pre-school in rural informal area	21.974	11.247	1.950	0.051	-0.127	44.074
Got pre-school in urban formal area	31.450	10.416	3.020	0.003	10.982	51.919
Got pre-school in urban informal area	-12.190	19.121	-0.640	0.524	-49.763	25.383
Constant	489.878	40.859	11.990	0.000	409.588	570.167

Note: Dependent variable is a score in a numeracy test with mean of 500 and standard deviation of 100. Lower and upper bounds use the 5% significance level.

Table 8: Linear regression model details Grades 3 to 10 learners (2)

n						441
R squared						0.1628
Adjusted R squared						0.1393
	Coef.	Std. Err.	t	P>|t|	LB	UB
Is in Grade 5	-1.160	42.865	-0.030	0.978	-85.411	83.091
Is in Grade 6	-21.993	40.756	-0.540	0.590	-102.099	58.114
Is in Grade 7	23.720	40.474	0.590	0.558	-55.834	103.273
Is in Grade 8 and above	32.265	41.247	0.780	0.434	-48.806	113.337
Class size	-0.769	0.258	-2.980	0.003	-1.276	-0.262
Got pre-school in rural formal area	-6.760	30.867	-0.220	0.827	-67.430	53.911
Got pre-school in rural informal area	23.013	11.513	2.000	0.046	0.383	45.643
Got pre-school in urban formal area	4.461	12.171	0.370	0.714	-19.462	28.384
Got pre-school in urban informal area	-21.902	20.271	-1.080	0.281	-61.744	17.940
Adults in household	-5.857	2.304	-2.540	0.011	-10.386	-1.328
Household income (log of monthly)	19.454	5.746	3.390	0.001	8.161	30.748
School fees (log of annual)	7.620	2.152	3.540	0.000	3.391	11.850
Constant	344.823	58.527	5.890	0.000	229.786	459.860

Note: Dependent variable is a score in a numeracy test with mean of 500 and standard deviation of 100. Lower and upper bounds use the 5% significance level.