The Role of the Family in Determining a Child's Educational Outcomes

Masters Half-Dissertation

Sameeha Kola (KLXSAM001)

This dissertation is presented in partial fulfilment of the requirements of the Masters of Commerce Degree in Applied Economics at the University of Cape Town, May 2011

Abstract

Despite the government's attempts to remedy the inequities of apartheid, there are large racial differences in educational outcomes, with whites performing substantially better than non-whites. In understanding these differences most studies have emphasised the role of school quality, without adequate emphasis on the role of the family. This study has investigated the role of the family in determining educational outcomes, and further how this is different for non-whites and whites. A holistic model of education was used which investigates the role of the family, while controlling for the effect of individual and school level characteristics that impact outcomes. Test scores from a literacy and numeracy evaluation were regressed on family level variables (family structure, parental involvement and socio-economic status), school quality variables and measures of individual ability and effort. The results of this study highlight that parental involvement and socio-economic status play an important role in the educational outcomes of both non-whites and whites, even after controlling for differences in school quality and individual ability and effort. It is not clear that family structure on its own accounts for variations in educational outcomes. Rather it is suggested it operates indirectly through impacting the parental involvement and financial resources available for education. Socio-economic status seems to play a much larger role in explaining the variations in outcomes of white students as compared to non-whites, suggesting a threshold level of socio-economic status above which differences matter more. Parental involvement levels are uniformly high among whites, and therefore does not serve to explain variations in educational outcomes amongst whites. Parental involvement does however have a role in accounting for variations between whites and non-whites and also amongst non-whites. The implications of this study highlight the need to encourage parental involvement amongst non-white families, as a means to improve educational outcomes. The lack of parental involvement often stems from a lack of parental education, impacting the ability to support the educational progress of their children. Programs that enable and equip non-white parents to understand how the educational system works, and how best to assist their children in school, would have lasting benefits by improving the long term prospects of their children.

Plagiarism Declaration

1. I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.

2. This paper entitled "The Role of the Family in Determining a Child's Educational Outcomes" is my own work.

3. I have used the Harvard convention for citation and referencing. Each significant contribution to, and quotation in, this paper from the work(s) of other people has been attributed, and has been cited and referenced.

4. I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

Signature:

Date:

Acknowledgements

I would like to sincerely thank Professor Murray Leibbrandt for his input, support and encouragement.

I acknowledge funding from the National Research Foundation (NRF) and the South African Labour and Development Research Unit (SALDRU).

This paper extensively uses data from the Cape Area Panel Study (CAPS). The Cape Area Panel Study Waves 1-2-3 were collected between 2002 and 2005 by the University of Cape Town and the University of Michigan, with funding provided by the US National Institute for Child Health and Human Development and the Andrew W. Mellon Foundation. Wave 4 was collected in 2006 by the University of Cape Town, University of Michigan and Princeton University. Major funding for Wave 4 was provided by the National Institute on Aging through a grant to Princeton University, in addition to funding provided by NICHD through the University of Michigan.

Table of Acronyms

CAPS	Cape Area Panel Study
DET	Department of Education and Training
HOA	House of Assembly
OLS	Ordinary Least Squares
LNE	Literacy and Numeracy Evaluation

Contents

Section 1: Introduction 8

Section 2: Literature Review 12

2.1. Theoretical Approach 12

2.2. The Determinants of Education 14

2.2.1. Individual Determinants 14

2.2.2. Family Determinants 16

2.2.2.1. Family Structure 16

2.2.2.2. Parental Involvement 20

2.2.2.3. Socio-economic Status 22

2.2.3. School Determinants 24

2.2.4. Neighbourhood Determinants 26

2.3. Concluding Remarks 28

Section 3: Methodology 29

3.1. Research Overview 29

3.2. Model Specification 30

3.2.1. Contemporaneous Specification 32

3.2.2. Value Added Specification 35

3.3. Analysis Method 36

Section 4: Data 38

4.1. Cape Area Panel Study (CAPS) 38

4.2. School Register of Needs Survey (2000) 39

4.3. South African Census (2001) 39

Section 5: Descriptive Statistics 41

5.1. Sample Selection 41

5.2. Descriptive Statistics 41

5.2.1. Outcome Variable 42

5.2.2. Individual Level Explanatory Variables 43

5.2.3. Family Level Explanatory Variables 44

5.2.4. School Level Explanatory Variables 57

5.2.5. Neighbourhood Level Explanatory Variables 59
5.2.6. Value Added Variable 60
5.3. Key Insights 60

Section 6: Regression Analysis 62
6.1. Category Specific Regressions 62
6.1.1. Individual Level Regressions 63
6.1.2. Family Level Regressions 64
6.1.2.1. Family Structure 65
6.1.2.2. Parental Involvement 66
6.1.2.3. Socio-economic Status 67
6.1.3. School Level Regressions 72
6.1.4. Neighbourhood Level Regressions 74
6.2. Education Production Function Regressions 76

Section 7: Conclusions 82

Bibliography 85

Appendix A 90
Appendix B 94
Appendix C 99
Appendix D 100

Section 1: Introduction

Education in South Africa is characterised by stark contrasts along racial lines in terms of both school quality and learners' outcomes, despite the government's efforts to remedy the inequities of apartheid[1]. Education policy since democracy in 1994 has been predominantly aimed at improving access to schooling for previously disadvantaged groups and improving school and teacher quality in previous African (DET) schools through increased funding and various incentive schemes. Although these policies have improved resources in many schools, the quality of previous DET schools is still lagging behind previous white (HOA) schools as the latter are able to improve classroom facilities, extra-curricular activities and financial incentives for teachers though supplementing government funding with school fees (Fiske and Ladd, 2004).

Within the South African context pupil-teacher ratios serve as a good proxy for the resource quality of a school, with larger pupil-teacher ratios being closely associated with lower test scores and reduced educational attainment (Case and Deaton, 1999; Anderson, Case and Lam, 2001). Data from the 2000 School Register of Needs[2] shows that pupil-teacher ratios vary drastically across schools. Previous DET schools have a mean of 33.2 with a maximum of 60 pupils per teacher in some schools, compared to a mean of 24.7 and a maximum of 33 in previous HOA schools. Furthermore, the high variation in pupil-teacher ratios in previous DET schools illustrates the high levels of variation in terms of quality across these schools. Poorly resourced previous DET schools are also characterised by internal governance problems, high levels of teacher absenteeism, violence and sexual abuse (Hoadley, 2007 and Brookes *et al*., 2004 in Lam, Ardington and Leibbrandt, 2008).

Although there has been a steady increase in the mean schooling of Africans since the 1920's (Anderson, Case and Lam, 2001), there are still large racial differences in learner outcomes. The quality of knowledge and skills gained by Africans in schools is still well

[1]Apartheid was a system of government that determined the rights and opportunities of the population in terms of racial classification. Whites were favoured with privilege in all aspects of life including education. Schools were racially segregated, with limited resources allocated to African schools (Fiske and Ladd, 2004).

[2] The data from the School Register of Needs Survey was merged with the Cape Area Panel Study (CAPS) data used in this paper before calculating mean values of the pupil-teach ratio.

below that of their white counterparts even when the level of educational attainment may be the same (Van der Berg *et al*., 2002 in Van der Berg and Burger, 2002; Van der Berg, 2008). African children often start school later and remain enrolled in school well into their teenage years. A study by Lam, Ardington and Leibbrandt (2008) confirms high levels of grade repetition in African populations which accounts for this age disparity: "84% of white students who were in grades 8 and 9 in 2002 successfully advanced three grades by 2005, compared to 44% of coloured students and only 32% of African students (2008: 3). By the age of 19, approximately 90% of white students have completed secondary school compared to 43% of coloured students and 23% of African students (Ardington and Leibbrandt, 2009).

The matriculation exam or 'matric' is a milestone in the educational attainment of a student in South Africa as it represents the completion of secondary school and enables access to tertiary education. As the matriculation exam is standardised at a national level, it is useful in comparing educational attainment across population groups. In 71% of African schools, more than half of all learners failed their matriculation exams in 1999-2000 (Van der Berg and Burger, 2002). The authors also found (using 1997 figures) 60% of the matriculation passes with university exemptions were from predominantly white schools with only 2.5% from predominantly African schools.

Understanding the determinants of educational attainment is central to understanding these racial inequalities and the effects of different policy initiatives. The level of education achieved by an individual as well as the quality of that education in terms of knowledge acquired is complex and multifaceted. Research in this area spans various disciplines including Economics, Sociology and Psychology. Research in South Africa has been largely focussed on understanding the school level determinants of education, without adequate investigation of the role of the family in perpetuating racial differences in educational outcomes.

This study investigates the role of the family in impacting an individual's educational outcome, within the framework of a larger model that accounts for other determinants of education. Racial groups in South Africa tend to have very different family structures:

African's often live in extended families while whites reside in two parent nuclear families or with single parents (Lu, 2009). Socio-economic differences between population groups also impact how parents relate to their children's educational and personal needs. The interpersonal resources[3] provided by a parent in developing their child is positively related to parental education and income levels (McNeal Jr., 2001; Sayer et al., 2004 in Hango, 2007).

Standard OLS regression analysis is used to explore correlations that exist between family level variables (such as family structure and parental involvement) and educational outcomes. A comprehensive education production model is used, which controls for individual, school and neighbourhood characteristics, with the aim of understanding whether observed correlations can explain racially based educational inequalities in South Africa. Investigation of correlations between other determinants of education and educational outcomes are also highlighted to enable the development of the education production function model.

The strength of this study is the integration of knowledge from various sources to create a comprehensive model of educational attainment based on an education production function. This approach identifies the relationships between individual, family, school and neighbourhood variables on the one hand and educational outcomes on the other. This study creates a view of the education process as a whole and, although emphasis is placed on the family, it is invaluable as a building block for any further research of education in South Africa.

This study reiterates previous research in South Africa and highlights large racial disparities in educational outcomes, with whites on average achieving much higher literacy and numeracy test scores than both Africans and coloureds. Due to the apparent contrast between whites on the one hand and Africans and coloureds on the other hand, this study has grouped African's and coloureds into the category non-white.

[3] Interpersonal resources are also referred to as cultural resources and include the behaviours and attitudes of parents that impact parent-child and parent-school relationships. Examples include parental aspirations and expectations for their child's education, communication between parent and child and involvement in school related activities (Downey, 1995; Fan and Chen, 2001).

The study finds that the family does indeed have a role to play in the educational outcomes of both non-whites and whites, even when individual and school level variables are controlled for. The families' socio-economic status (measured by parental education, family income and household resources) seems to have the largest impact of the family level variables, especially in the case of whites. This is in line with a study by (Van der Berg, 2008), who finds that socio-economic status impacts the educational outcomes of those attending richer (white) schools more. For non-whites, parental involvement measures specifically parental educational expectations also seem to have an important role in explaining the variations in educational outcomes. These results suggest that parental involvement may have a greater role to play in families from a disadvantaged background as it serves to mitigate the effect of socio-economic disadvantage (Hango, 2007). The results of the study emphasise that individual, family and school level variables all have a role to play in explaining educational outcomes and none of these categories should not be neglected in empirical anaylsis.

The next section, 2 provides a literature review which presents the theoretical approach used in this study and identifies the key determinants of education across individual, family, school and neighbourhood categories. Section 3 outlines the model used in this study and the steps by which the analysis was conducted. Section 4 identifies the various data sources used in this study to create a rich data-set that provides information across individual, family, school and neighbourhood levels. Section 5 identifies the variables that have been selected as part of the empirical model and provide descriptive statistics and sample comparisons of means and standard deviations between non-whites (coloureds and Africans combined) and whites. Section 6 presents the regression analysis and results. The final section, 7 highlights the key conclusions of this research, and suggests possible implications for education policy in South Africa.

Section 2: Literature Review

There is no standard theoretical approach used in education studies, nor is there a set of education determinants that are common across studies. In light of this complexity an in depth investigation of the literature was conducted to provide the theoretical background required to research and integrate the various categories of determinants (individual, family, school and neighbourhood) into a single model in a coherent way. The first section of the literature review introduces the theoretical approach used in this study which is based on an integration of the education production function and family specific studies. The second section on the determinants of education provides a detailed account of the explanatory variables that are often used in these studies and the mechanisms through which they operate. This theory has been key to the selection of variables for this study (Appendix A), and the empirical specifications and interpretations of the regressions used in this study.

2.1. Theoretical Approach

In the economic theory of the firm a production function describes the maximum level of output that can be achieved from different combinations of inputs, given available technology. A widely used application of the production function in studies on education is an education production function, which identifies the technology that enables the school as the production unit to identify the optimal mix of inputs to achieve the highest educational outcome for its students (Hanushek, 1986; Pritchett and Filmer, 1999 in Boissiere, 2004). A change in an educational input changes the output by an amount that is determined by the combination of the other inputs into the production process and the technology underlying the input/output relationship (Boissiere, 2004). The strength of this approach lies in its emphasis on the process of education where an individual is subject to the constraints of the production process. This distinguishes the approach from demand

driven approaches to education such as Human Capital Theory[4], which neglect the impact of supply constraints (such as the functioning of schools and teachers), on the educational outcomes of students (Vandenberghe, 1999).

The education production function is modelled on an input/output basis. The output is student achievement; common measures of achievement are years of schooling (Case and Deaton, 1999) and standardised test results (Lamdin, 1996). Less frequently used outputs include enrolment rates (Case and Deaton, 1999), grade repetition (Lam, Ardington and Leibbrandt, 2008), tertiary education continuation, pass rates and student attitudes (Hanushek, 1979). Inputs into education production functions include genetic endowments, individual behavioural inputs such as school attendance, school and teacher quality (Lamdin, 1996; Hanushek, 1979). Although educational output is a discrete measure taken at a point in time the education process is seen as cumulative. Consequently, input measures are typically both contemporaneous[5] and historical, to accommodate for the impact of the past on the current outcome measure (Todd and Wolpin, 2003).

Although theoretically the approach acknowledges a large variety of inputs that impact output, including family and neighbourhood variables, studies have almost exclusively focused on genetic endowments and school and teacher related variables. This is primarily due to data constraints, as few data sets contain rich data on both the family and the school (Boissiere, 2004). This is the biggest weakness of the education production function, as it fails to account for the effect of the family and neighbourhood on educational outcomes. This potentially overstates the impact of school quality on educational outcomes. A recent adaptation of the education production function by Todd and Wolpin (2003)[6] emphasises

[4] Human capital theory pioneered by Becker (1964) is a demand driven theory which holds that educational attainment is determined by an individual's demand for human capital. Human capital "takes many forms including skills and abilities, personality, appearance, reputation and appropriate credentials" (Becker and Tomes, 1986: S6). Education is therefore viewed as an investment good and an individual will choose their level of education by comparing the costs and benefits of each additional year of education.

[5] Contemporaneous measures are current measures taken at a point in time.

[6] Their results showed that controlling for genetic endowments school level characteristics accounted for a very small portion of the racial difference in schooling outcomes in the United States; while differences in family level inputs accounted between 25%-30% of the racial differences. This further supports the need to include both school and family level inputs in analysis (Todd and Wolpin, 2006).

the importance of including both family and school variables as inputs, although the authors don't go far enough and fail to account for the effect of family structure and parental involvement measures such as help with homework and parental educational expectations for their child.

In order to build on the education production function to better accommodate the effect of family and neighbourhood inputs, studies emphasising the role of the family and the neighbourhood on educational outcomes were investigated. Studies of this nature focus on correlations between educational outcomes and specific family or neighbourhood explanatory variables, providing much insight into the functioning of families and neighbourhoods, and the mechanisms through which these variables operate. The weakness of these studies is not adequately controlling for the effect of other categories of determinants of education, (especially school quality), which introduces omitted variable bias and potentially overstates the importance of the family or neighbourhood (Baharudin and Luster, 1998 and Crane, 1996 in Todd and Wolpin, 2003). By integrating the education production function and insights from family and neighbourhood studies, the weaknesses of both approaches are mitigated through the introduction of a holistic model of education.

2.2. The Determinants of Education

Having established the importance of accounting for the effect of individual, school, family and neighbourhood factors on educational outcomes, an in depth investigation of the key determinates of education was conducted. This served to guide the selection of explanatory variables for empirical analysis (Appendix A), and interpret the regression analysis.

2.2.1. Individual Determinants

At an individual level, genetic endowments and personality play the prominent role in determining educational attainment. Behman and Taubman (1989 in Miller, Mulvey and Martin, 2001) estimate that 80 percent of the variation in education outcomes are due to genetic endowments (of which intelligence is the largest component). Using a sample of 3000 Australian twins, Miller, Mulvey and Martin (2001) provide more modest estimates,

suggesting that genetic endowments account for between 50 and 65 percent[7] of the variation in educational outcomes. Similarly, Mackintosh (1998, in Deary, Strand, Smith and Fernandes, 2007) in a review article find similar systematic correlations between intelligence and educational outcomes, ranging between 0.4 and 0.7. These findings are not suprising considering the cognitive ability required to acquire knowledge and progress successfully through school. The role of genetic endowments in impacting outcomes is also strongest when there are no other external constraints on acquiring education. For example, when an individual is unable to afford school, the financial constraint plays the primary role in educational outcome while genetic endowments take a secondary role (Burns, 2001).

Conscientiousness and the will-to-achieve are the aspects of personality that have the strongest relationship with educational outcomes[8] (Poropat, 2009). These are linked to academic effort, completing homework tasks, time management and setting goals, where the effort that is put into education by an individual is key in determining educational attainment (Steel, 2007 in Poropat, 2009). South African studies have emphasised the importance of academic effort in determining grade progression and educational attainment (Van der Berg, 2008; Lam, Ardington, and Leibbrandt, 2008). Conscientiousness and will to achieve are closely related to an individual's educational aspirations and expectations. Beal and Crockett (2010) find support for the positive correlation between an individuals future aspirations and expectations and their educational attainment in a sample of approximately 300 American adolescnets. This correlation is seen to operate through positive attitudes and the individuals participation in positive behaviours such as study, and the avoidance of negative behaviours such as substance abuse.

[7] Adjustments to account for assortative mating increase estimates of the contribution of genetic endowments to educational outcomes.

[8] Other aspects of personality that show much weaker relationships to educational attainment include "Agreeableness (reflecting likeability and friendliness), Emotional Stability (adjustment versus anxiety), Extraversion (activity and sociability), and Openness (imaginativeness, broad-mindedness and artistic sensibility) (Poropat, 2009).

2.2.2. Family Determinants

A survey of the literature highlighted a wide variety of family factors used across studies to account for the role of the family in a child's educational attainment. These factors have been categorised into socio-economic status, family structure and parental involvement.

2.2.2.1. Family Structure

Family structure typically includes family size and family type; the former is normally a measure of the number of siblings in a household and the latter a measure on whether the family is a single parent family, nuclear family or blended family in which step parents and/or step children co-reside. Numerous studies have shown that family size is negatively correlated with educational attainment even when controlling for family background characteristics (Downey, 1995; Ginther and Pollak, 2004; Anderson, 2000). The theoretical mechanism through which family size operates is by limiting financial resources and parental time available to provide interpersonal resources[9] to each child in a household as household size increases (van Eijck and de Graaf, 1995). In sociology the sibling resource – dilution theory provides support for the basic notion that having many siblings reduces educational outcomes; stating that "an increase in the number of siblings and a decrease in their spacing dilute the resources that parents can spend on each child (Anastasi, 1956 in van Eijck and de Graaf, 1995: 274)". There is also support for this negative relationship in economics. Becker and Tomes (1976 in de Haan, 2009) propose a theory where a tradeoff exists between child quantity and child quality. In this model it is more expensive for parents to invest in their children if they have many children as the marginal cost of quality increases with family size for any given budget constraint.

A study in Hungary with a sample size of 17146 showed the number of siblings in a family to have a small negative statistically significant relationship to years of education completed. Depending on the age groups of the individuals the coefficient of number of siblings varied from -.07 in the oldest to -.19 in the youngest, suggesting an increase in the importance of family size in explaining the variation in educational attainment (van Eijck and de Graaf,

[9] Interpersonal resources are the behaviours and attitudes of parents that relate to education and include helping with homework and communication regarding education and school performance (Downey, 1995).

1995). Similarly, Ginther and Pollak (2004) and Downey (1995) also find a statistically singnificant negative relationship between educational outcomes and number of siblings when controlling for family background characteristics.

However, recent studies using instrumental variables for family size have indicated that when birth order is taken into consideration this negative relationship becomes statistically insignificant (de Haan, 2009) or almost completely disappears (Black, Devereux and Salvanes, 2005).

In the South African context family size is a more complex measure due to the large presence of extended families, especially in African populations. Competition for parental resources occurs not only from siblings but also from extended family members. However, extended families may also provide the parent/s with a support structure and assist in child rearing activities or with other household tasks freeing up the parent to focus on the children. These two effects operate in opposite directions making empirical observations difficult to interpret (Anderson, 2000; Lu, 2009). Case and Deaton (1999) find a negative effect (-.669) of increased family size on years of completed education in a sample of white 10-18 year olds and a smaller positive effect (.188) in Africans, both were statistically significant. Similarly, Lu (2009) finds a negative effect of sibling size on educational outcomes for whites and no effect for Africans. Whites in South Africa predominantly have family structures observed in the developed world, hence the consistency of results across studies in the USA and South Africa. The results for Africans suggests that the extended families may play a role in reducing the negative relationship between number of children and educational outcomes by assisting in the rearing of children (Anderson, 2000; Lu, 2009).

Regarding family type many USA studies show that children living in traditional nuclear families have higher educational achievement than children from all other family structures (Ermisch and Francesconi, 2001; Biblarz and Raftery, 1999; Sandefur and Wells, 1999; Ginther and Pollak, 2004; Anderson, 2000). These studies use different measures of family type that is typically guided by data constraints. Studies suggesting no effect normally do not control adequately for other family characteristics specifically family size. For example, if family size is not controlled for a child from a single mother with no siblings may show an

advantage in educational attainment when compared to children from two parent familes with many children (Biblarz and Raftery, 1999).

There are various theories[10] that provide an explanation of these empirical results. In economics theories focus on the increased availability of parental resources; two parent families have the advantage of providing complementary financial and interpersonal resources to the child. Financial resources typically provided by a father are complemented with interpersonal resources that a mother may focus on providing (de Haan, 2009; Biblarz and Raftery, 1999). Children living in alternate family structures generally receive less encouragement, support and hand's on help with homework than children in nuclear families (Astone and McLanahan 1991, in Kan and Tsai, 2005).

Some argue that the correlations between family type and educational outcomes may represent the importance of unobserved family characteristics. Children from divorced parents may perform poorly academically not because they are a part of a single parent family but perhaps because their parents are "less competent at family life (Biblarz and Raftery, 1999: 326)" or due to the high level of conflict in households who divorce (Sandefur and Wells, 1999). Where the unobserved characteristic jointly determines family structure and outcomes, problems of endogeneity arise. Most fixed effect studies which controlled for the potential endogeneity of family structure have found statistically significant negative effects of alternate family types as compared to traditional nuclear families suggesting that endogeneity is not a significant concern with regards to family structure (Ermisch and Francesconi, 2001; Evenhouse and Reilly, 2004 in Ginther and Pollak, 2004).

Sandefur and Wells (1999) using a sample of 4312 find that a child's years of completed schooling is negatively related to living in alternate family structures compared to nuclear families. The strength of the relatioship is reduced when controls for parental education and income, family size and race are added to the model. The authors use various different measures of family structure such as number of disruptions to family structure, the number

[10] These include socialisation, learning and loss of parental control theories. All these theories emphasise the way parent – child interactions differ across family structures. For example single parent families are coupled with inconsistent or reduced parenting time and often have to deal with trauma from the death of a parent or divorce (Ginther and Pollak, 2004; Biblarz and Raftery, 1999).

of years lived in a traditional nuclear family and the family type in which the child resides at age 14. The author concludes that the number of years lived in a traditional nuclear family is the most effective measure of family structure as such a measure accommodates historical impacts on the family. Similarly, Ginther and Pollak (2004) find that "both types of children in blended families stepchildren and their half-siblings who are the joint children of both parents are similar to each other and substantially worse than outcomes for children reared in traditional nuclear families (Ginther and Pollak, 2004: 671)"; controlling for birth order, number of siblings, parental income and parental education.

There are very few studies in South Africa that have investigated the effect of family type on educational outcomes. Studies by Van der Berg (2008) and Lu (2009), find no evidence for an educational advantage to living in a nuclear family when controlling for family background characteristics. The former uses a measure of the proportion lived with both parents, the latter uses measures of co-residence at age 14. Similarly, Case and Deaton (1999) find no disadvantage for children from single female headed households as compared to other family types.

Contrary to these studies Anderson (2000) found that family type to be highly correlated with various educational outcomes for Africans, even while controlling for parental education and income. The strongest relationship was found for children living without either biological parent, who "were less likely to be enrolled in school, had completed fewer grades, were older for their grade if enrolled, and has less money spent on their school fees and school-related transportation costs (Anderson, 2000: 5)," than children from traditional nuclear families. Children from female headed households (predominantly single parent households), also displayed lower educational attainment. Nimubona and Vencatachellum (2007) also find that children from female headed households (irrespective of race) complete fewer years of education than alternate family types.

Studies on family structure emphasise that the effect of family structure operates directly as well as indirectly through its effect on limiting the financial and interpersonal resources available for education. Empirical results therefore show that higher levels of income

compensate for the negative effects of family size and type (Sandefur and Wells, 1999; Anderson, 2000).

2.2.2.2. Parental Involvement

Parental involvement essentially comprises the interpersonal resources provided to children and can be loosely defined as the behaviours and attitudes parents display that impact their child's educational outcomes. Parents typically choose the level and type of parental involvement they provide based on the specific skills they have, time and financial constraints and requests from children and schools. Psychologists and educators have long established that parents involved in their child's development and educational needs tend to have children who do better academically. These children develop positive attitudes and behaviours towards education and develop intellectually in ways that support success at school (Downey, 1995; Hoover-Dempsey and Sandler, 1995).

There is no universally accepted set of measures that comprise parental involvement; with data availability primarily driving the decisions of measures to include in empirical studies. Different measures of parental involvement impact a child's educational outcomes differently; composite measures of parental involvement are therefore discouraged as results become difficult to interpret accurately (Fan and Chen, 2001; Jeynes, 2005; Tan and Goldberg, 2009). Studies often show different effects for fathers' and mothers' suggesting where possible the parental involvement activities of each parent should be investigated separately (McBride *et al*., 2005 in Tan and Goldberg, 2009; Hango, 2007). Studies also suggest that parental involvement reduces the negative impacts of family income constraints on the child through supplementing financial resources with interpersonal resources that build social capital (Hango, 2007; Tan and Goldberg, 2009).

There are various approaches used to categorise the separate aspects of parental involvement. The most widely used approach distinguishes between school-based involvement strategies and home based involvement strategies. The former includes parent-teacher conferences, volunteering at school, communicating with teachers and participating in school governance (Hill and Tyson, 2009). Parental training by teachers and involvement

between schools and communities are sometimes identified as home based strategies (Hill and Tyson, 2009), however these should rather be grouped as school-based as they pertain to the relationship between parents and schools. Home based interventions are best understood in terms of Comer's analysis, (1995 in Hill and Tyson, 2009: 741) as "parental reinforcement of learning at home", and includes activites such as help with homework, reading and playing together, educationally based extra curricular activities (e.g. visits to museums), dicipline, encouragement and motivation. Grolnick and Slowiaczek (1994 in Hill and Tyson, 2009) build on this model to include educational aspirations and expectations for their children as well as the communication of these expectations.

Empirical studies show large inconsistencies in terms of the measures of parental involvement used and the existence, strength and direction of the relationship between parental involvement and a child's educational outcomes. Meta analysis studies conducted in the USA provide insight into these inconsistencies by highlighting the most commonly used measures of parental involvement and the average results across studies. Fan and Chen (2001) and Jeynes (2005) identify the most commonly used measures of parental involvement as parental educational aspirations and expectations for their child, parent-child communication regarding education, home structure (which includes activities such as time management, reading and play activities, help with homework and providing an atmosphere that encourages study), parenting style and participation in school based interventions. Fan and Chen (2001), Jeynes (2005) and Hill and Tyson (2009) all show on average small to moderate positive realtionships between almost all parental involvement measures and a child's educational outcomes.

In a meta analysis of 25 studies, Fan and Chen (2001) find parental aspirations and expectations for their children's education has the strongest relationship (.40) to a child's educational attainment and home structure has the weakest relationship (.09). In their meta analysis of 41 studies Jeynes (2005) find similar results with parental aspirations and expectations having a strong relationship (.58) to a child's educational attainment. Reading with a child (.42) and parenting styles (.31) also show significant positive correlations to a child's educational outcomes. Although the measure was statistically insignificant checking a child's homework was negatively correlated to educational outcomes (- .08). Hill and

Tyson (2009) in their meta analysis of 50 studies find both home based and school based parental involvement to be positively correlated to achievement in middle school with the exception of parental help with homework.

Parental involvement may possibly be endogenous with common unobservable factors impacting both parental involvement and educational outcomes. For example a child's ability and consequent performance in school may also encourage greater (or lesser) parental involvement (Maccoby and Martin, 1983 in Kan and Tsai, 2005). Most studies however, do not account for this potential endogeneity because the mechanisms of operation are not established. There is no clear theory that determines if and how parents adjust their interpersonal resources dependent on their child's characteritics and performance.

2.2.2.3. Socio-economic Status

The commonly used measures of socio-economic status are parental education and family income. Although these two measures are correlated (higher education is normally associated with higher levels of income) each measures a different aspect of socio-economic status and should be considered separately. Income measures, proxy for the available financial resources and is subject to greater variations over time, while parental education is related to the stability and long term status of the family. Other measures used are parental occupation and household resources (such as books and computers) (Sirin, 2005).

Numerous international studies confirm positive correlations between parental education and family income and educational outcomes (Aakvik, Vaage and Salvanes, 2005; Davis-Kean, 2005; Ermisch and Francesconi, 2001; Dubow, Boxer and Huesmann, 2009; Blanden and Gregg, 2004). These variables operate both directly and indirectly to impact childrens educational attainment. Direct links include access to financial resources, quality schools and neighbourhoods which are largely improved by higher levels of parental income and education. Indirect links are seen through choices of family structure, parental expectations of children, and behaviours such as reading and playing with children, parental involvement

in school activities and parental warmth (Kodde and Ritzen, 1988; Davis-Kean, 2005; Davis-Kean and Sexton, 2009; Anderson, Case and Lam, 2001).

Regarding the relative impact of family income and parental education Blanden and Gregg (2004) using careful econometric specification found a causal connection between family income and educational attainment, and suggest that parental education is only important through its impact in largely determining family income. Duncan and Brooks-Gunn (1997 in Davis-Kean and Sexton, 2009) disagree and find that parental education is far more important than income in the long run. They find that although income has a significant positive effect at age 4, this effect declines and almost disappears by adolecence.

In South Africa, Van der berg (2008) emphasises the importance of home resources particiularly access to books and computers as measures of socio-economic status, and find a positive relationship between home resources and educational outcomes. South African studies also find parental education and family income to be positively correlated to educational attainment. A study by Lam, Ardington and Leibbrandt, (2008) using data from the Cape Area Panel Study[11] found parental schooling to have weak positive effects on a child's grade advancement. However, these effects were only statistically significant for father's education in the coloured sample. The relationship between family income and educational outcomes was strongly positive and statistically significant for whites and coloureds but not significant for Africans. Lam (1999 in Anderson, Case and Lam, 2001: 46) estimates that "the schooling advantage of African children whose mothers' have 12 years of schooling compared to those whose mothers' have less than 4 years of schooling is roughly two full grades for both 13 and 17 year olds", with a similar size effect for fathers' education. No differences are seen dependent on the gender of the child. Nimubona and Vencatachellum (2007), however find a mother's level of education is approximately 10% more important than the father's level of education.

[11] The current study also uses Cape Area Panel Study (CAPS) data.

2.2.3. School Determinants

Insight into the role of the school in educational attainment is largely provided by education production function studies. Harbison and Hanushek (1992: in Boissiere, 2004) provide a classification of key school input into hardware, software, teacher and management/institutional inputs. Hardware refers to tangible school resources impacting the quality of the school facilities, including school buildings, classroom, sports, computer and sanitation facilities. Software refers to smaller scale tangible resources such as textbooks and stationary, and intangible school provisions such as the curriculum and teaching styles endorsed by the school. Teacher inputs include measures of teacher availability such as pupil-teacher ratios, as well as teacher quality such as education, teaching times and experience levels. Management and institutional inputs are measures of governance for example examination methods and access to funding. The selection of inputs is however ultimately decided by data constraints (Boissiere, 2004).

Studies report very different findings of the importance of school inputs due to the use of different control variables, data quality, regression methods and sample selection criteria (Boissiere, 2004). In a review of 147 American articles using the education production function methodology, Hanushek (1986: 1162) concluded that "there appears to be no strong or systematic relationship between school expenditures and student performance". Although he acknowledged some studies (13 of 65)[12] showed a statistically significant positive correlation between expenditures (which serves as a general input measure of school resource and quality) and student test scores, this relationship became marginal or ceased to exist when family background variables were controlled for.

Investigating the relationship between teacher /pupil ratios and student test scores Hanushek (1986) also found great disparity; 89 of the 112 studies showed statistically insignificant findings; 9 showing a statistically significant positive relationship and 14 a statistically significant negative relationship. Similar disparities were found when

[12] 3 of the 65 studies found a statistically significant negative relationship between school expenditures and student test scores.

investigating the relationships between teacher education[13], teacher experience[14] or teacher salary[15] on the one hand and test scores on the other. Later studies by Hanushek (2003) have served to confirm these trends across school inputs, even when extended to include school facilities and school administration inputs. Hanushek (2003: F39) acknowledges that there are instances where "small classes or added resources have an impact. It is just that no good description of when and where these situations occur is available".

Using the same studies as Hanushek (1986), Hedges, Laine and Greenwald (1994) disagree with Hanushek's conclusions that school inputs do not always matter. Using more sophisticated synthesis methods[16] than the vote counting method used by Hanushek, the authors conclude that expenditures per pupil and student outcomes are positively related sytematically across studies and the magnitude of the relationship is "large enough to be of practical importance"(Hedges, Laine, and Greenwald, 1994: 5). Krueger (2003: F35) also disagrees with Hanushek's review methods, stating that "he inadvertently places a disproportionate share of weight on a small number of studies that frequently used small samples and estimated mis-specified models", hence resulting in flawed conclusions. Krueger's (2003) re-analysis of Hanushek's conclusions regarding pupil/teacher ratio's showed statistically significant positive relationships to increase from 14.8% to 33.5% through using adjusted weights to account for the quality differentials across studies.

In South African studies pupil-teacher ratios are widely used as a proxy for school facilities and general school quality (Case and Deaton, 1999). Using data from a 1993 survey, Case and Deaton (1999: 1050) "find strong and significant effects of pupil-teacher ratios on

[13] 85 of the 106 studies showed statistically insignificant findings; 6 showed a statistically significant positive relationship and 5 a statistically significant negative relationship.

[14] 69 of the 109 studies showed statistically insignificant findings; 33 showed a statistically significant positive relationship and 7 a statistically significant negative relationship.

[15] 50 of the 60 studies showed statistically insignificant findings; 9 showed a statistically significant positive relationship and 1 a statistically significant negative relationship.

[16] The method used accounts for outliers, estimate size across studies and correlation of error terms across studies using the same sample.

enrolment, on educational achievement, and on test scores for numeracy[17]" for African and coloured students. For example, in African schools a reduction of the pupil-teacher ratio by a quarter would increase average educational attainment (measured in years) by 0.52 (between ages 10-18) and enrolment by 0.02 a year (Case and Deaton, 1999). Case and Yogo (1999) echo the importance of pupil-teacher ratios estimating that an average reduction of the pupil-teacher ratio by 10 would increase average educational attainment by 0.6 years. Contrary to these studies Van der Berg (2008), find no strong evidence to support these views and conclude that pupil-teacher ratios have only a limited role to play in educational outcomes in South Africa.

Glewwe and Jacoby (1994) bring to attention the potential issue of endogeneity where both the educational outcome and the school attended could be determined by an external factor such as an individual's motivation, innate ability or their parents preferences. In South Africa this tends to be less of a problem, although choice in schools exists, there are large racial, locational and income barriers which limit the range of schools a child can attend. This makes school selection largly exogenous, especially for African families in poverty (Case and Deaton, 1999).

2.2.4. Neighbourhood Determinants

Neighbourhood variables used in studies are typically proxies to account for the socio-economic status of a neighbourhood and include poverty rates, racial compositions, average incomes and the proportion of adults in managerial positions. Variable selection is primarily driven by data availability and there is no clear justification why some measures should be preferred (Ellen and Turner, 1997; Wilson, 2001). The role of the neighbourhood in influencing educational attainment lacks consensus. Wilson (2001) found that students dropping out of high school lived in American neighbourhoods with lower socio-economic statuses. These results were statistically significant but accounted for a very small variation in the outcome variable (i.e. drop out from high school). A different study found the racial composition of neighbourhoods was not significant in accounting for variation in

[17] The effect of pupil-teacher ratios on the educational achievement of white students showed a small statistically insignificant negative correlation.

educational attainment in similar American neighbourhoods (Datcher, 1982 in Wilson, 2001). Jencks and Mayer (1990) provide a review of neighbourhood effects and conclude that there is no universal explanation for their role; and this is emphasised by a diversity of empirical results. Although there is some evidence that the socio-economic characteristics of neighbourhoods and peers are negatively correlated to educational outcomes, this should not be taken as a rule.

Ellen and Turner (1997) identify three main methodological issues that give rise to the large variations in empirical results on the importance of neighbourhoods. Firstly, there are a diverse set of measures used accross studies which influence conclusions. Secondly, neighbourhood effects are assumed to be linear, which is simplistic. A poverty rate of 2% may have no effect on educational outcomes whereas a poverty rate of 30% may largely impact educational outcomes. This suggests the existence of a minimum threshold above which negative effects may be felt (Ellen and Turner, 1997). Thirdly, and most importantly, neighbourhood effects are overstated in simple single equation models due to the endogeneity of family and neighbourhood effects (Wilson, 2001; Jencks and Mayer, 1990; Evans, Oates and Schwab, 1992). For example a parent concerned about their child's education may move to a neighbourhood with the characteristics that enable positive development; this parent may also assist their child with homework and have a positive communication style. If the role of the parent is not taken into account one may attribute the child's performance at school solely to the enabling environment provided by the neighbourhood and overstate the neighbourhood effect. Consequently, empirical studies controlling for family characteristics have smaller or no evident neighbourhood and effects (Ellen and Turner, 1997).

Surprisingly a South African study by Burns (2001) using data from KwaZulu-Natal found neighbourhood effects to operate independently from family background characteristics. Results showed that living in a community where the proportion of households earning a monthly income of over R3000 is high and the unemployment rate is low significantly

reduces schooling gaps[18]. This finding should be taken cautiously as the study did not test for robustness using alternate model specifications.

2.3. Concluding Remarks

The literature review clearly sets out the key determinants of education, the variables commonly used to measure these determinants, and the expected mechanisms through which these determinants operate. This information was used to investigate which variables available in the dataset would be best suited for use in the empirical analysis. This process of variable selection is highlighted in Appendix A. Furthermore, understanding the complex nature of the education process serves to assist interpretations of empirical analysis. The next section on the methodology used in this study, builds on aspects of the literature review and sets out a model which integrates the various categories of determinants of education in a coherent manner.

[18] The schooling gap is "the potential years of education (i.e. number of years of education an individual would have if they started school at age 6 and advanced one grade every year) minus the number of years of education actually attained (Burns, 2001: 10)".

Section 3: Methodology

3.1. Research Overview

This study has two key research questions. Firstly, 'What is the role of the family in educational attainment in a South African context?' In answering this question I aim to describe the correlation between educational outcomes (using a standardised literacy and numeracy score) and family level variables. The second question is 'To what extent is the role of the family in educational attainment different for whites and non-whites (Africans and coloureds) in a South African context?' This question builds on the first question and seeks to understand the relationship between family and educational attainment on the one hand and race on the other hand. This relational question supplements the descriptive elements of the first question in a way that is vital in the South African context where racial differences in educational outcomes are extreme.

Research into the effects of family specific factors in a child's educational outcomes is often conducted without adequately controlling for the other determinants, especially those at the school level. This makes it difficult to understand the true importance of family specific factors in the process of educational attainment as a whole. To provide accurate estimates of the role of the family, other determinants of education that impact an individual's educational outcomes need to be controlled for. This has added a further dimension to this study, i.e. to identify the key individual, school and neighbourhood factors that are related to educational outcomes in the South African context and to adequately control for these variables in the empirical analysis, where appropriate[19]. Consequently, this study investigates the role of the family within the context of a holistic model of educational attainment that includes individual, family and school and factors.

[19] Neighbourhood factors have been investigated separately and have not been included in the holistic model due to the very high levels of correlation between race and neighbourhood factors. This is discussed in more detail in later sections of this paper.

3.2. Model Specification

This study uses an augmented education production function to investigate the role of the family while controlling for the effects of individual and school factors. The traditional education production function is used as the foundation into which additional determinants of education commonly only used in family specific studies, are incorporated. The strength of this approach is its wide ambit in understanding the entire process of educational attainment and accounting for variables that are normally neglected in traditional education production functions despite theoretical and empirical evidence of their importance in other studies. The variables included in the empirical model were selected from a very rich dataset (outlined in section 4), based on the findings from a large variety of literature. The variable selection process is presented in Appendix A. The strength of the education production function is retained through using its methodology and ease of interpretation while its weaknesses are mitigated through including detailed variables on family structure, parental involvement and individual variables (other than innate ability), to supplement school level variables.

The production function on which this study is based is specified below. For any specified individual educational attainment at time t (A_t) is a function of individual level characteristics (I_t) comprising of genetic endowments and cumulative individual inputs up to time t; family level characteristics cumulative to time t (F_t); school level inputs cumulative to time t (S_t); neighbourhood inputs cumulative to time t (N_t) and Ɛ the residual which includes all the unmeasured characteristics that impact educational attainment (Hanushek, 1979).

$$A_t = f(I_t, F_t, S_t, N_t) + \varepsilon$$

Test results are the preferred output measure of educational attainment at time t (A_t) in education production functions as they illustrate differences in cognitive achievements across individuals and hence the quality of achievement. It is also true that higher test scores are normally associated with the long term goals of schooling, namely success after school. Tests of basic arithmetic and reading skills are a measure of the foundational educational concepts required both for success at higher levels of education and integration and success in the labour market and therefore more accurate at estimating the true value

of education than tests at higher levels of education (Hanushek, 1986). Externally validated tests are also more accurate than school specific tests as they provide a uniform measure of ability that is not influenced by teacher quality (Hanushek, 1979). Other measures such as years of schooling, attendance rates or attitudes towards education do not take into account the actual skills and knowledge gained through education and are therefore an inefficient proxy measure for the real outcome of education, which is skills development to improve productivity and success after schooling (Lamdin, 1996; Hanushek, 1979, 1986).

This model builds on a traditional education production function in two ways. Firstly, individual level characteristics are seen as a combination of both genetic endowments (innate ability) and activities related to aspects of personality such as study time and individual educational expectations, whereas traditional education production functions typically only address innate ability. Secondly, the model builds on the traditional education production function view of the family which typically only includes family income and parental education (if any family level characteristics at all) to further include aspects of family structure and parental involvement. Ideally, the model should also include neighbourhood level variables. However, in this study the decision was made not to include neighbourhood variables in the empirical education production function (section 6.2) due to the high level of correlation between race and neighbourhood factors[20] in the South African context due to the legacy of apartheid (Burns, 2001).

It is important to note that there are various relationships that exist between inputs. For example, higher levels of parental education are often associated with higher levels of income, greater parental involvement with children and access to better schools and neighbourhoods and sometimes smaller families. An individual's educational aspiration or expectation may be related to the level of parental involvement or the genetic endowments of the individual. Although these correlations between input variables may exist it is vital to include an input where there is theoretically and empirically sound evidence to suggest its importance in determining educational outcomes (Todd and Wolpin, 2003). The importance

[20] Various neighbourhood variables were regressed against race to ascertain the degree of explanatory power that race has in each case, with resulting correlations of between 63% and 77%. These regressions are presented in Appendix E.

of establishing a clear theoretical reason why inputs should be included in the analysis has been emphasised as the most important aspect of making decisions on input variables (Becker, 1964; Hanushek, 1979, 1986; Krueger, 2003). Hanushek (1979) insists that "independent" explanatory variables will seldomly be independent, multicollinearity must be accepted in most estimations and accounted for in the interpretation of results.

The education production function specified above "portrays the educational production relationship as cumulative; past inputs have some lasting effect, although the value in explaining output may diminish with more distant inputs (Hanushek, 1979: 364)". Cumulative specifications of the educational production function require large amounts of data, and in practise are incredibly difficult if not impossible to specify empirically. Hence, it is ultimately data availability that influences the selection of input measures (Hanushek, 1979; Todd and Wolpin, 2003). "As in other areas of empirical research, compromises are frequently necessary between what is conceptually desirable and availability of data (Hanushek, 1986: 1158)".

As a result of data constraints empirical analysis of education production functions typically need to make strong assumptions in order to establish validity, which has led to a variety of different model specifications. The two main specifications used are the contemporaneous and value added specifications. These specifications are used despite their weaknesses because they are often the only specifications available to the researcher with a limited data set (Lamdin, 1996; Dee, 2004; Heckman and Li, 2004; Hanushek, 1986). This study uses both the contemporaneous and value added specifications to compare results under the different assumptions of each specification.[21]

3.2.1. Contemporaneous Specification

The contemporaneous specification relates educational attainment (often as a test score) solely to current or cotemporaneous measures of educational determinants (e.g. quality of the school currently attended). Genetic endowments and historical measures (e.g. the quality of previous schools attended) are not required to build an empirical model based on

[21] The cumulative specification was not feasible in this study due to the lack of historical data collected prior to the standardised test score in 2002, which is used as the outcome variable in this study.

this specification (Todd and Wolpin, 2003; Hanushek, 1979). This specification enables analysis when there are severe data limitations and therefore despite its shortcomings is widely used in studies (Lamdin, 1996; Dee, 2004). When using this specification the residual includes (1) omitted contemporaneous variables, (2) historical inputs, (3) genetic endowments and (4) measurement error (Todd and Wolpin, 2003, 2006).

The contemporaneous specification of the education production function is specified below. For a given individual, educational attainment at time t (A_t) is a function of current individual level characteristics (I_{t*}) comprising current individual inputs (genetic endowments are either omitted or accounted for with the use of proxy measures); current family level characteristics (F_{t*}); current school level inputs (S_{t*}); current neighbourhood inputs (N_{t*}) plus Ɛ the residual which captures the unobserved and omitted characteristics and historical inputs that impact educational attainment (Todd and Wolpin, 2003; Hanushek, 1979).

$$A_t = f(I_{t*}, F_{t*}, S_{t*}, N_{t*}) + Ɛ$$

The contemporaneous specification is based on the following assumptions. "*(i)* Only contemporaneous inputs matter to the production of current achievement **or** *(ii)* Inputs are unchanging over time, so that current input measures capture the entire history of inputs. **And in addition to *(i)* or *(ii)*,** *(iii)* Contemporaneous inputs are unrelated to (unobserved) endowed mental capacity (Todd and Wolpin, 2003: F16)" and other genetic endowments.

Assumptions *(i)* and *(ii)* are required to eliminate the measurement error that results when cumulative inputs are not included in education production functions. (Hanushek, 1979). Assumption *(i)* does not hold under any approach to education. It is widely accepted across disciplines that the education process is cumulative and past inputs matter. Assumption *(ii)* is used in the contemporaneous specification of this study. It is more aligned to reality to assume unchanging inputs when inputs are measured in the earlier years of education before significant time has lapsed for changes to occur (such as in the present study); or in circumstances where change is not facilitated by the external environment. This is the case in South Africa, those from disadvantaged backgrounds do not readily move into economic advantage or to significantly better schools in short spaces of time because of the high

income and education constraints faced by their parents. Therefore not catering for historical school and neighbourhood inputs is unlikely to create significant bias because the quality of the neighbourhood or school do not largely vary over time. Furthermore, as parental education and family income does not readily change in the South African context, (specifically for Africans), one can assume that family level variables (such as parental financial and interpersonal resources[22] provided for education) would also not change significantly over time. This means that observed measures of these variables are largely representative of historical measures hence, family level variables would typically be subject to smaller measurement errors.

Assumption *(iii)* is related to the effect on estimates when genetic endowments that impact the capacity to learn are not accounted for. Under this assumption, although the residual variance would increase through omitting measures of genetic endowments, there would be no bias on the other estimated coefficients. There is evidence however that child intelligence is positively correlated to family background characteristics (Miller, Mulvey, and Martin, 2001), which consequently results in an upward bias on the estimates of family level inputs (Todd and Wolpin, 1996, 2003). Proxy variables (such as parental education and family income) that theory suggests are highly correlated to unobserved genetic endowments are normally used to reduce this bias.

The data set used in this study does not have data on genetic endowments; hence in the contemporaneous specification parental education levels are used as proxy measures for the child's genetic endowments in an attempt to reduce the omitted variable bias. This is based on the observation that parents with higher levels of education and better test scores generally have more favourable genetic endowments (i.e. higher scholastic abilities) and therefore have children with higher scholastic abilities (Bowles and Nelson, 1974 in Kodde and Ritzen, 1988). This is in lines with the approach used in various international studies where maternal test scores, parental education levels or a combination of both are included

[22] Interpersonal resources are the non-financial resources parents provide their children with to promote educational achievement. For example help with homework, participating in school and communication on school related matters (Downey, 2005; Fan and Chen, 2001).

as proxy measures for a child's genetic endowment (Miller, Mulvey and Martin, 2001; Kodde and Ritzen, 1988).

It is important to note parental education not only accounts for genetic endowments but various other unobservable family level characteristics. For example it is suggested that parental education is a measure of family socio-economic status and accounts for variations in family structure and parental involvement (Davis-Kean and Sexton, 2009; Hango, 2007; Davis-Kean, 2005). Therfore the inclusion of parental education in the contemporaneous specification suggests that these regressions may control for a large component of family level characteristics even when specific family level variables are not included in the regression.

3.2.2. Value Added Specification

The value added specification is an adaptation of the contemporaneous specification that includes a lagged baseline achievement measure in order to reduce the bias from omitted historical inputs and genetic endowments (Todd and Wolpin, 2003)". Achievement measures are normally standardised test scores from earlier in the child's life; that is before the education outcome being studied. However, any achievement score pertaining to the individual can be used, such as pass rates or grade repetition (Hanushek, 1979). The value added specification used in this study uses a measure of the proportion of grades passed prior to taking the LNE test[23] in 2002.

The value added specification is presented below. Educational attainment at time t (A_t) is a function of the elements of the contemporaneous specification (I_{t^*}, F_{t^*}, S_{t^*}, N_{t^*}) and baseline achievement (B_A) measured in this study by the "proportion of schooling passed", plus the residual Ɛ which measures unobserved and omitted inputs that impact achievement.

$$A_t = f(I_{t^*}, F_{t^*}, S_{t^*}, N_{t^*} B_A) + \varepsilon$$

This specification rests on the assumption that the inclusion of a baseline achievement measure reduces measurement error (as compared to the contemporaneous specification)

[23] The LNE test score is the outcome variable in this study.

through accounting for historical inputs and genetic endowments by proxy. The level of bias in the value added specification is therefore dependent on the extent to which the baseline achievement measure accounts for the true effect of historical inputs and genetic endowments. When using this specification the residual is reduced to include only (1) omitted variables and (4) measurement error (Todd and Wolpin, 2003). Consequently, many authors have given preference to the value added specification over the pure contemporaneous specification (Hanushek, 1979, 2003; Todd and Wolpin, 2006; Kruger, 2000 in Todd and Wolpin, 2003).

3.3. Analysis Method

Contemporaneous and value added empirical specifications were defined using variables across individual, family, school, and neighbourhood categories. The process of variable selection involved an in depth analysis of the literature on the determinants of education, the modelling methods used in education production function studies and the available data. This process is highlighted in Appendix A. Multiple regression analysis using OLS methods was used to analyse the empirical specifications and identify correlations between the outcome and explanatory variables. As this study is concerned in identifying and explaining racial differences, all regressions were run separately for whites and non-whites (Africans and coloureds). This decision is supported by the kernal density graphs of the outcome variable (presented in section 5) which reveal a much greater area of common support between Africans and coloureds, with minimal overlaps with whites. Van der Berg and Burger (2002), also model non-whites (combining Africans and coloureds) and whites separately; based on the similarity in the school and socio-economic characteristics between Africans and coloureds on the one hand, and the differences with whites.

Firstly, category specific (i.e. individual, family, school and neighbourhood) regressions are presented to identify the relationships between category specific variables and the outcome variable. The family level category further includes separate regressions for each sub-category of family variables; family structure, parental involvement and socio-economic status. The explanatory variables for the individual, family and school level regressions are then combined into an education production function regression that addresses the impact

on educational attainment holistically. This regression allows investigation of the role of family level variables while controlling for variables across other categories. As previously highlighted neighbourhood variables are not included into the education production function due to the high level of correlation with race.

Both category specific regressions and the education production function are run in terms of both the contemporaneous and value added specifications of the education production function. In light of the general view on the preference of the value added specification (Hanushek, 1979; Todd and Wolpin, 2003), these regressions are included in the main text of this study while contemporaneous specifications are included in the Appendix. Key differences between the results of the two specifications are highlighted.

Section 4: Data

This study uses data primarily from the Cape Area Panel Study (CAPS) (2002-2005). Other data sources are used as supplements include the School Register of Needs Survey (2000) and the South African Census (2001). The data set used provides the richest array of individual, household, community and school data that is currently available in South Africa. This is especially the case with regards to CAPS data on the family and home environment.

4.1. Cape Area Panel Study (CAPS)

CAPS[24] is a longitudinal study of the lives of young adults and their families in the greater metropolitan area of Cape Town in South Africa. Wave 1 was conducted between August and December 2002 on a randomly selected sample of 4752 young adults aged between 14 and 22 and their households. A maximum of 3 young adults per household were included in the young adult sample. The sample consists of 42% African, 44% coloured and 14% white. Africans and whites were oversampled in order to achieve a sample that comprises of equal numbers of coloured and Africans and half as many whites, and is not representative of the population breakdown in Cape Town. Sample weights are used in this study to adjust the racial proportions closer to the actual population in Cape Town. A third of these youth were re-interviewed in 2003 (Wave 2a) with the remaining two-thirds being re-interviewed in 2004 (Wave 2b). The full sample was then re-visited in both Wave 3 and Wave 4 in 2005 and 2006 respectively (Lam *et al*, 2008a).

Household response rates in Wave 1 in households with identified young adults were high in African and coloured areas (89% and 82% respectively) and low in white areas (48%). Given household participation, the participation of young adults was high across all racial groups. Overall attrition between wave 1 and wave 3 was 19% (Lam *et al*, 2008b), this study only uses wave 1 data and hence attrition rates do not influence results.

CAPS comprises a rich panel data set which includes young adult and household data on various aspects of education and schooling outcomes, employment, living standards,

[24] Technical documentation and additional detail on CAPS data is available online at www.caps.uct.ac.za. CAPS data sets are available for download.

residential movements, income, health, sexual relationships, family structure and parental involvement with youth's. Wave 1 data also include a 20 minute literacy and numeracy test that was administered to all young adults (Lam *et al*, 2008b). The test was available in both English and Afrikaans. Almost all Xhosa young adults (99%) took the test in English. This indicates a level of disadvantage to Xhosa young adults who took the test in their second language (Ardington and Leibbrandt, 2009).

4.2. School Register of Needs Survey (2000)

The School Register of Needs Survey mandated by the Ministry of Education was conducted in both 1996 and 2000 to collect school level data to investigate and monitor school quality and informed education policy making. The 2000 Survey was extensive and covered 27 148 public and private schools and 390 schools for learners with special needs. Data collected included the school information such as type of school, physical infrastructure and facilities, classroom-pupil ratios, technology, criminal incidents, textbooks and stationary; and teacher information such as percentage of teachers privately paid and pupil-teacher ratios (Department of Education, 2000).

A subset[25] of the full School Register of Needs data set was merged with CAPS data using school codes to create a data set that provides information across individual, family and school categories (Lam, et al., 2008a).

4.3. South African Census (2001)

The 2001 census provides comprehensive population data including community-level demographic, housing, education, health, income, employment and access to public services data at the individual and household level.

Census data categorised by mainplace and subplace level was used to develop a CAPS Wave 1 community level data set[26]. This data set also included generated variables from the census data such as mean years of education for adults, mean household income and

[25]The subset was created in a format to enable merging with CAPS data. It is named the CAPS Wave 1 school level data file and available by request from www.caps.uct.ac.za.
[26] Data file available by request from www.caps.uct.ac.za.

poverty indicators (Lam, et al., 2008b). The CAPS Wave 1 community level data set was merged with CAPS data and the School Register of Needs data using CAPS cluster codes to provide a comprehensive data set with individual, family, school and community variables required for educational attainment analysis.

Section 5: Descriptive Statistics

5.1. Sample Selection

The first step in sample selection was to exclude all those young adults not enrolled in high school (i.e. grade 8 to grade 12) in 2002. There is extensive research on the importance of school level variables on educational outcomes (Boissiere, 2004; Krueger, 2003), and this study could only include school variables if the specified sample attended school. Furthermore, inclusion of school level variables ensures individual, family, school and neighbourhood category regressions are all based on a similar sample. The second step was to exclude all young adults with a missing LNE score, the outcome variable used in this study. Very few observations were dropped as a result of a missing LNE score, 4 Africans, 4 coloureds and 1 white. The sample[27] used consists of 831 Africans, 825 coloureds and 234 whites.

5.2. Descriptive Statistics

This section provides a description of the sample used in this study. In addition each variable's mean and standard deviation is compared across non-whites (Africans and coloureds) and whites. Lastly, the relationships between key variables (particularly family level variables) and standardised LNE scores (the outcome variable) are investigated separately for non-whites and whites. All means and standard deviations presented in this study have been weighted to account for sample design and household and young adult non-response. The weights adjust this sample to provide a representative sample in terms of the demographics of Cape Town. These results are descriptive and therefore fail to account for the effect of other related variables.

[27] The sample size has small differences across regressions included in this study, as some variables have missing values assigned to some observations and regression analysis automatically excludes these observations from the analysis. Comparisons across contemporaneous and value added specifications of category specific regressions are done despite small differences in sample size as the differences are not significantly large.

5.2.1. Outcome Variable

The outcome variable used in regression analysis is the standardised test score from the CAPS Literacy and Numeracy Evaluations (LNE). As highlighted in section 3, test scores are the preferred measure of educational attainment in education production function studies. As the LNE test included both basic literacy and numeracy questions, the LNE score is a combination of the results from both modules. The table below shows the summary statistics for the actual LNE score (out of a total of 45) by race. It is immediately clear that the mean score for Africans is much lower than the mean scores for whites, with coloured's falling somewhere in the middle. In the African sample, some young adults failed to get a single question correct.

TABLE 1: LNE SCORE BY RACE

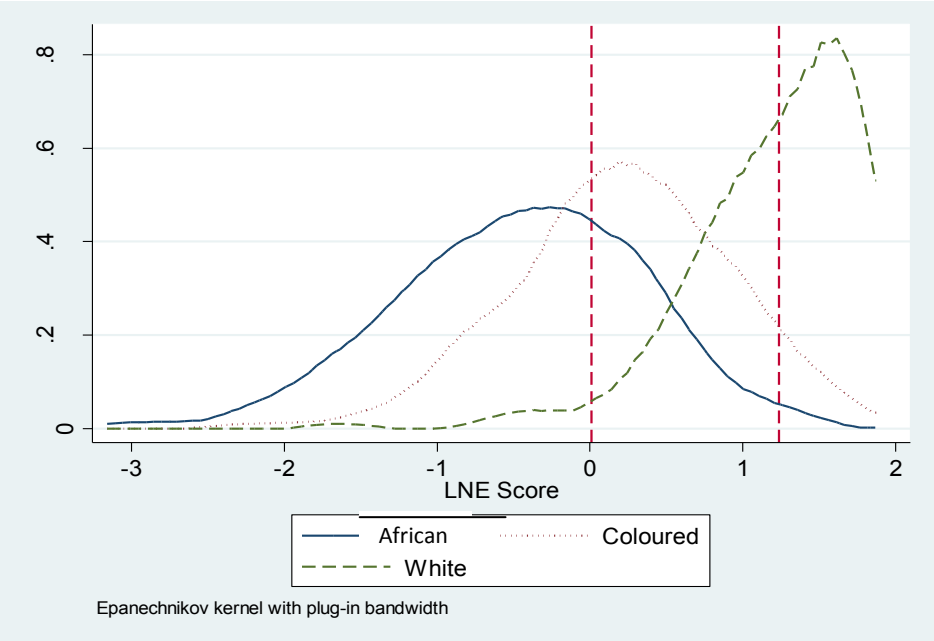
	Obs	*Mean*	*Std. Deviation*	*Min*	*Max*
African	831	22.88	7.37	0	42
Coloured	825	29.01	6.49	6	45
White	234	38.00	4.86	12	45
Total	1890	28.93	8.19	0	45

(Source: Cape Area Panel Study, 2002)

The kernel density graphs of the standardised LNE score below show the distribution of the outcome variable by race, and clearly illustrates that the African sample has both the lowest mean and widest distribution. In contrast, the white sample has the highest mean, with a distribution which is skewed left. This indicates that most whites achieved higher scores than other race groups. The coloured sample also has a wide distribution although it has a higher mean value than that for Africans, illustrated by the graph being to the right of the graph for the African sample. The kernel densities demonstrate that the area of overlap between Africans and whites is minimal, with the African and coloured distributions revealing a much greater area of common support. These distributions therefore support the study design used in the following chapter which runs regressions separately across non-whites (Africans and coloureds) and whites. The large gap between non-whites and whites is expected given the legacy of apartheid; non-whites continue to obtain lower educational attainment than whites (Case and Deaton, 1999; Fiske and Ladd, 2004). This is reflected in the kernel densities by placing vertical lines at each group's respective mean

LNE score. The line at the value 0.01 represents the mean LNE score for non-whites and the line at 1.24 represents the mean LNE score for whites.

Figure 1: Kernel Density Plot - Standardised LNE score by race



(Source: Cape Area Panel Study, 2002)

5.2.2. Individual Level Explanatory Variables

Studies demonstrate that the most important individual level variables in the analysis of educational attainment are genetic endowments, conscientiousness and individual educational expectations (Miller, Mulvey, and Martin, 2001; Poropat, 2009; Beal and Crockett, 2010). In this study, genetic endowments is proxied by "proportion of schooling passed"[28] in the value added regressions; and by parental education[29] in the contemporaneous regressions. Individual effort on study "study time" is included as a measure of conscientiousness. Dummy variables are used to estimate individual educational expectations, these are referred to as, "educational expectation no matric", "educational expectation matric" and "educational expectation tertiary". Demographic individual level variables "age" and gender "male" are also presented. The table below presents the

[28] This variable controls for past achievement at school and is used in value added regressions to account for the effect of genetic endowments and historical inputs into the production process. The descriptive statistics of this variable is included in section 5.2.6.

[29] Descriptive statistics for parental education are provided in section 4.2.3., which contains the descriptive statistics for family level variables.

weighted descriptive statistics for the individual level variables used in this study. The table shows that "study time" has few missing observations, while the educational expectations variables have a large number (105) of missing observations.

Table 2. Descriptive Statistics: Individual Level Variables

Variables	*Obs*	*Mean*	*Std. Deviation*	*Min*	*Max*
Age	1890	16.43	1.76	14	22
Male	1890	0.47	0.50	0	1
Study Time	1878	7.78	7.10	0	70
Educational Expectation - No Matric	1785	0.04	0.20	0	1
Educational Expectation – Matric	1785	0.31	0.46	0	1
Educational Expectation – Tertiary	1785	0.65	0.48	0	1

(Source: Cape Area Panel Study, 2002)

The sample comparison table below presents the weighted means and standard deviations for the individual level variables, separately for non-whites and whites. The table indicates differences between races for all variables. However, at the 5% level, only "age" shows statistically significant differences. The average age of non-whites is higher than that of whites, which is expected in South Africa as non-whites tend to stay in school well into their late teens and even early twenties (Lam, Ardington, and Leibbrandt, 2008).

Table 3. Sample Comparison: Individual Level Variables

	Non-white		*White*		
Variable	*Mean*	*Std. Deviation*	*Mean*	*Std. Deviation*	*Significance at 5% level*
Age	16.51	1.80	16.06	1.54	*
Male	0.47	0.50	0.48	0.50	-
Study Time	7.58	6.76	8.73	8.45	-
Educational Expectation - No Matric	0.04	0.20	0.04	0.19	-
Educational Expectation – Matric	0.32	0.47	0.27	0.44	-
Educational Expectation – Tertiary	0.63	0.48	0.70	0.46	-

(Source: Cape Area Panel Study, 2002)

5.2.3. Family Level Explanatory Variables

Family level variables included in studies of educational attainment primarily fall into three main categories; family structure, parental involvement and socio-economic status**.** This study includes "household size" and the "proportion lived with both parents" as measures of family structure, and "mother/father helps with homework" and parental educational

expectations as measures of parental involvement**.** The latter is measured using three dummy variables, complete schooling below matric "no matric", complete matric as the highest level of education "matric" and complete some tertiary education "tertiary". Family structure and parental involvement are often influenced by other factors such as parental education and family income (Davis-Kean, 2005; Kan and Tsai, 2005). This study uses 'log of per capita family income", mother's/father's years of completed education and a measure of household resources "household owns computer" to measure socio-economic status. The process of variable selection this study undertook is highlighted in Appendix A. The weighted descriptive statistics for the family level variables included in this study are provided below. There are a large number of missing observations for the variables parental educational expectations (141), mother's education (155), and father's education (605).

Table 4. Descriptive Statistics: Family Level Variables

Variables	*Obs*	*Mean*	*Std. Deviation*	*Min*	*Max*
Household Size	1890	5.53	2.36	1	20
Proportion lived with both Parents	1886	0.62	0.43	0	1
Mother Helps with Homework	1890	0.24	0.43	0	1
Father Helps with Homework	1890	0.15	0.35	0	1
Parental Expectations – No Matric	1749	0.08	0.27	0	1
Parental Expectations – Matric	1749	0.38	0.49	0	1
Parental Expectations – Tertiary	1749	0.54	0.50	0	1
HH Own Computer	1890	0.29	0.46	0	1
Log Per Capita Household Income	1890	6.62	1.18	2.13	9.53
Mother's Education Missing	1890	0.08	0.26	0	1
Mother's Education	1735	9.52	3.10	0	16
Father's Education Missing	1890	0.28	0.45	0	1
Father's Education	1285	9.75	3.61	0	16

(Source: Cape Area Panel Study, 2002)

The table below presents a comparison of the means and standard deviations between non-whites and whites for the family level variables. On both measures of family structure, white children appear to be more privileged than non-whites. On average whites come from smaller families and have lived with both parents for a larger proportion of their lives. Similarly, whites have higher levels of parental involvement than non-whites. More whites tend to get help from their parents with homework than non-whites. White parents on average also have higher educational expectations for their children than non-white

parents. The mean value for "household owns computer", and the "log of per capita household income" is very high for whites suggesting white families have higher income levels and higher accessibility to educational resources such as a computer. The mean values for both "mother's education" and "father's education" are much higher (approximately 4 years) for whites than non-whites. All the differences between races indicated are statistically significant at the 5% level.

Table 5. Sample Comparison: Family Level Variables

	Non-white		*White*		
Variable	*Mean*	*Std. Deviation*	*Mean*	*Std. Deviation*	*Significance at 5% level*
Household Size	5.86	2.45	4.06	1.02	*
Proportion lived with both Parents	0.57	0.44	0.84	0.30	*
Mother Helps with Homework	0.18	0.39	0.48	0.50	*
Father Helps with Homework	0.10	0.30	0.35	0.48	*
Parental Expectations - No Matric	0.09	0.29	0.03	0.16	*
Parental Expectations - Matric	0.40	0.49	0.30	0.46	*
Parental Expectations - Tertiary	0.51	0.50	0.67	0.47	*
HH Own Computer	0.17	0.38	0.85	0.36	*
Log Per Capita Household Income	6.29	0.99	8.15	0.70	*
Mother's Education Missing	0.09	0.28	0.02	0.14	*
Mother's Education	8.77	2.86	12.73	1.73	*
Father's Education Missing	0.33	0.47	0.06	0.24	*
Father's Education	8.72	3.34	13.11	2.10	*

(Source: Cape Area Panel Study, 2002)

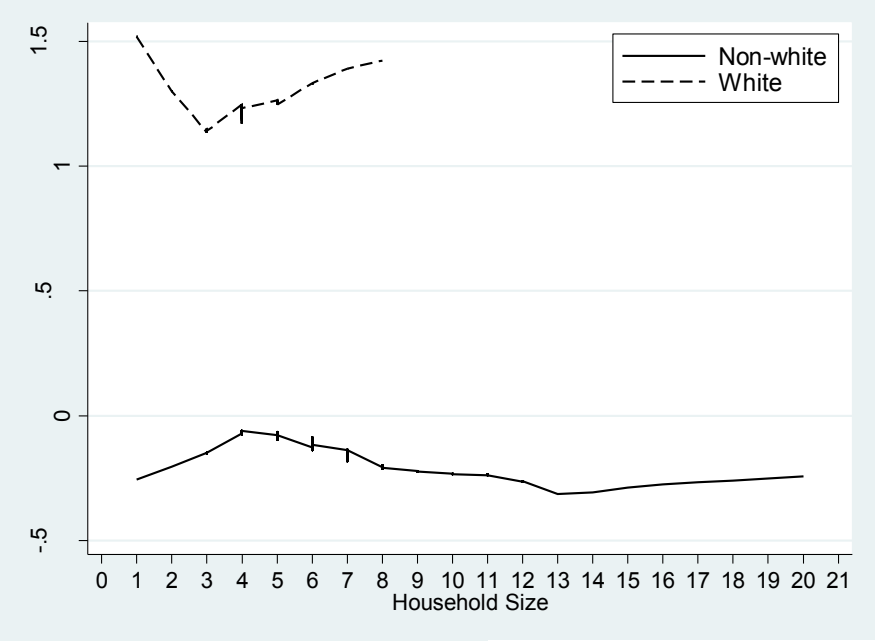
The relationship between each family level variables and the outcome variable the standardised LNE score is investigated below, and key trends are highlighted.

Household Size

Various studies suggest a negative relationship between household size and academic achievement (Lu, 2009; Downey, 1995; Ginther and Pollak, 2004; Anderson, 2000). The lowess (locally weighted regression) graph below illustrates the relationship between household size and LNE score, separately for non-whites and whites. The first important observation is that white households are much smaller than non-white households. The largest white household has 8 people (75.64% of whites have households with 4 members or less); whereas the largest non-whites household has 20 members (only 29.59% of non-whites have households with 4 members or less). South African studies show that non-white

families specifically African families, are often large extended families (Anderson, 2000). Secondly, the standardised LNE score for whites is greater than for non-whites for all household sizes. Thirdly, there is no clear relationship between household size and LNE score. For whites, LNE score's decrease steeply as household size increases to 3, thereafter increasing steadily. This suggests larger families may be beneficial in terms of educational support. For example in extended families grandparents may provide young adults with help on homework which positively influences learning and test scores (Anderson, 2000). For non-whites LNE scores increase as household size increases to 4, then slowly decreases as household size increases to 13 thereafter gradually increasing again. This suggests for non-whites in very large families additional members may cause little difference to how the family functions and hence have little impact on educational attainment.

Figure 2. Lowess: Household Size and Standardised LNE Score



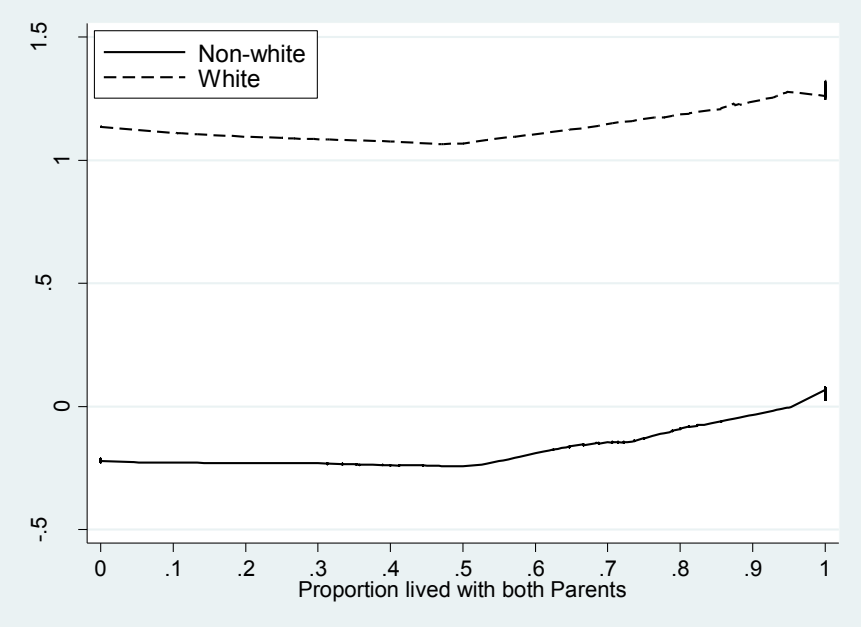
(Source: Cape Area Panel Study, 2002)

Proportion lived with both Parents

Studies suggest that children from nuclear families tend to perform better academically than children from other family arrangements (Ermisch and Francesconi, 2001; Biblarz and Raftery, 1999; Sandefur and Wells, 1999; Ginther and Pollak, 2004; Anderson, 2000). The graph below illustrates how the LNE score varies with increases in the proportion lived with both parents. For both non-whites and whites the LNE score stays fairly flat up until

proportion 0.5, thereafter gradually increasing. This suggests that increases in the proportion lived in a nuclear family are related to increases in LNE scores, provided that the child has already lived at least half of their lives with both parents. Increases for non-whites are larger than increases for whites, suggesting that the nuclear family structure may have a greater positive impact on non-whites. The standardised LNE score for whites is greater than non-whites for all proportions of living with both parents, reflecting the white advantage irrespective of living arrangements.

Figure 3. Lowess: Proportion lived with both Parents and Standardised LNE Score



(Source: Cape Area Panel Study, 2002)

Mother Helps with Homework

The literature review highlights the positive relationship between educational attainment and parental involvement measures such as help with homework (Hill and Tyson, 2009; Fan and Chen, 2001; Jeynes, 2005). Almost half of whites (48.7%) receive help from their mother for homework. This is contrasted sharply with only 16.5% of non-whites receiving help from their mother. The table below shows the mean LNE score when "mother helps with homework" and "mother does not help with homework", separately for non-whites and whites. Although, the differences are not significant, both non-whites and whites, present differences in mean LNE scores dependent on whether they receive their mother's help with

homework. For non-whites, receiving help is associated with higher scores while for whites somewhat surprisingly, receiving help is associated with lower LNE scores.

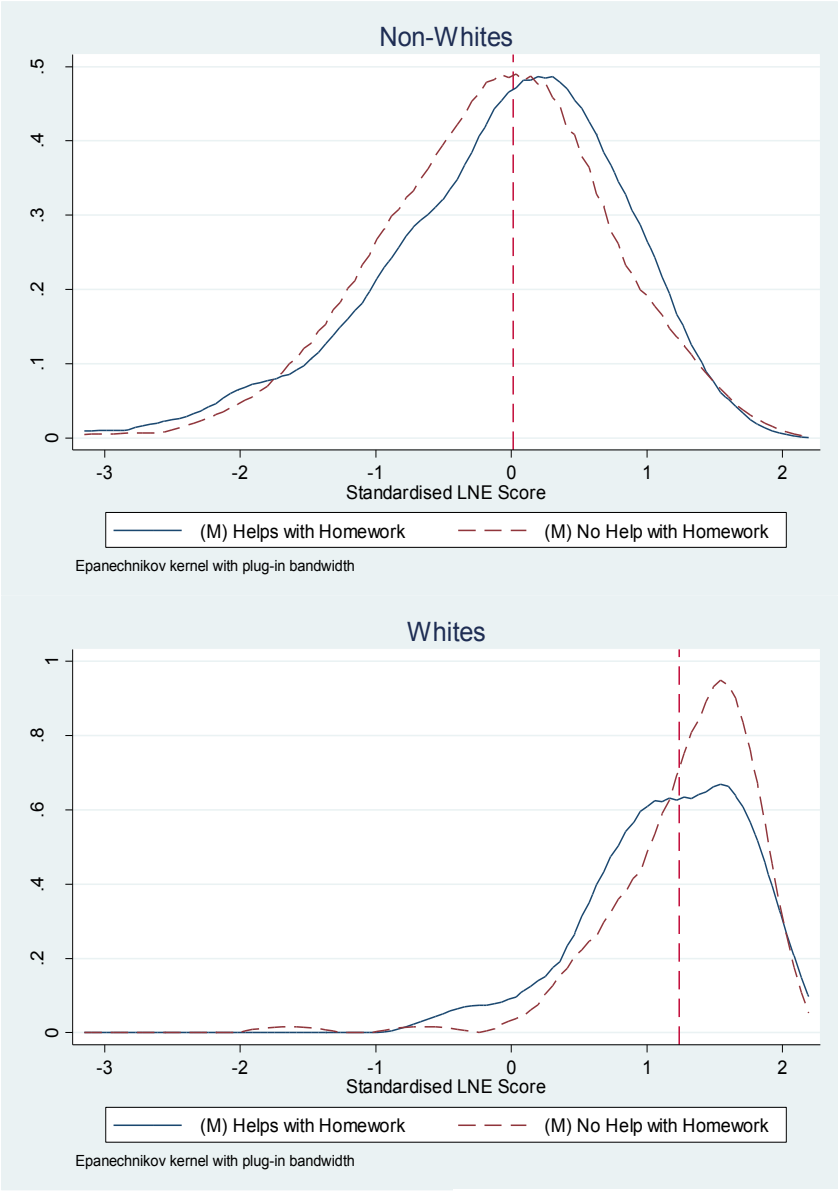
Table 6: Mean Standardised LNE score when Mother helps with/does not help with Homework

	Mother Helps with Homework	*Mother Does Not Help with Homework*	*Significance at 5% level*
Non-whites	0.06	0.001	-
Whites	1.20	1.27	-

(Source: Cape Area Panel Study, 2002)

The kernel density graphs below illustrate the variation in LNE score separately when a mother helps and does not help with their child's homework. For non-whites, the positive relationship between a mother helping with homework and the LNE score is illustrated by the curve for those who receive help from their mother being to the right of the curve of those who do not receive help. The vertical lines in the graphs depict the mean LNE score for the full sample of non-whites (0.01) and whites (1.24) respectively as depicted in Figure 1. For non-whites, the mean LNE score for those who receive help from their mother's is much higher than the mean score for the full non-white sample (the graph is to the right of the vertical line). For whites, the mean score for the full sample is higher than the mean scores for those who receive help from their mother's, but this could be the result of a small white sample. This analysis suggests that non-whites may benefit substantially from a mother's help with homework, while it is unclear whether or not this is true for whites.

Figure 4: Kernal Density Graphs: Standardised LNE scores and Mother's Help with Homework



(Source: Cape Area Panel Study, 2002)

Father Helps with Homework

A much larger percentage of whites (33.8%) receive help from their father for homework than non-whites (only 7.9%). This could be a result of the high percentage of whites co-residing with their fathers (74.4%) as compared to 46.1% of non-whites co-residing with their fathers. The table below shows the mean LNE score when a father helps with homework and when a father does not help with homework, separately for non-whites and whites. For both non-whites and whites, the mean standardised LNE scores are higher for those who receive help from their father for homework. This difference is larger and

statistically significant at the 5% level for non-whites. This suggests that for non-whites, the positive impact of "father helps with homework" on academic achievement may be larger.

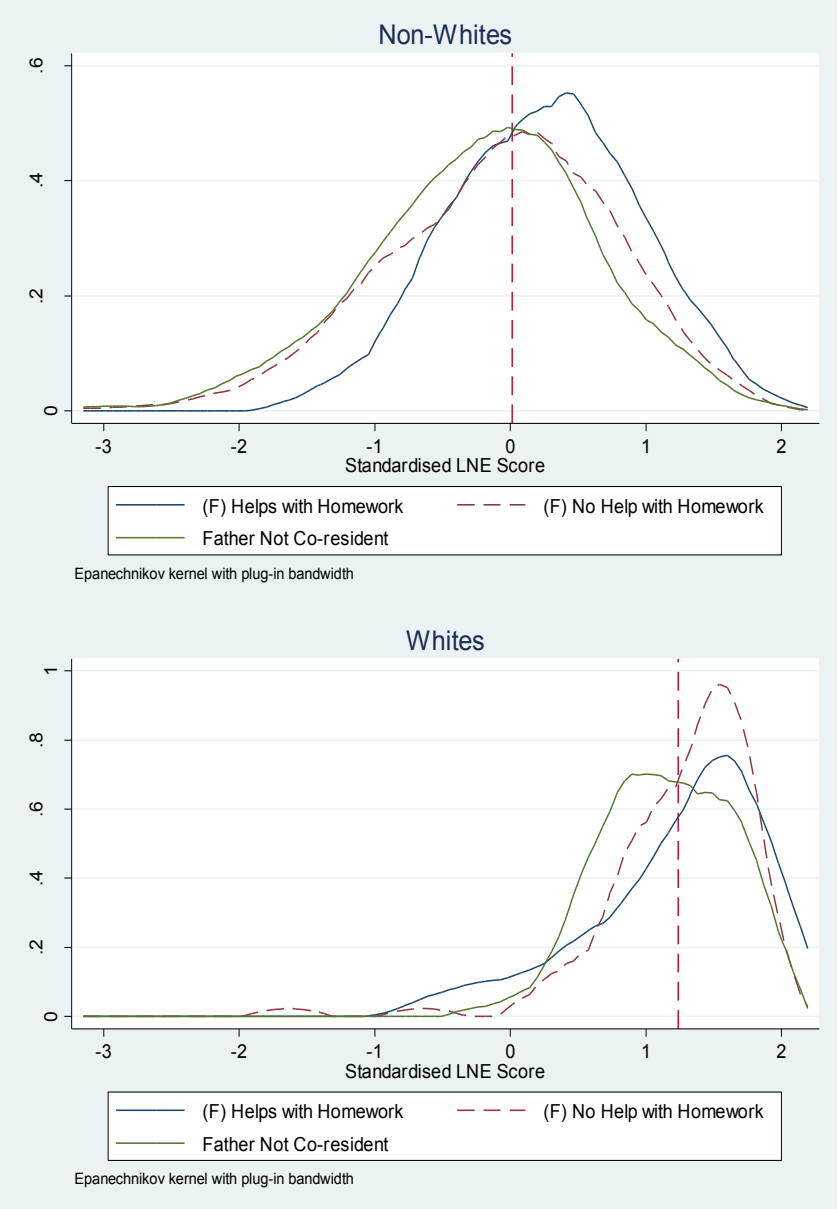
Table 7: Mean Standardised LNE score when Father helps with/does not help with Homework

	Father Helps with Homework	*Father does not Help with Homework*	*Significance at 5% level*
Non-whites	0.31	-0.02	*
Whites	1.27	1.22	-

(Source: Cape Area Panel Study, 2002)

In order to take into account a father's co-residence, the kernal density graphs below illustrate the variation in standardised LNE scores between those who do not co-reside with their fathers, those who co-reside with their fathers and receive help with homework and those co-residing but not receiving help from their fathers. For both non-whites and whites, those who do not co-reside with their fathers have the lowest mean LNE scores. For non-whites, those who do not co-reside with their father and those who do not receive help from their fathers with homework (but co-reside with their fathers) have very similar distributions on LNE scores; with those who receive help from their fathers having a distribution with a higher mean (to the right). For whites, there is little difference between the distributions of those who receive help from their fathers and those who do not, which are both to the right of the graph for those who do not co-reside with their fathers. This suggests that receiving help from a father may have a much larger impact on non-whites, whereas for whites co-residence with a father seems more important. As in Figures 1 and 4, the vertical lines in the graphs depict the mean LNE scores for the full sample of non-whites (0.01) and whites (1.24) respectively. For non-whites and whites, the graphs for those who receive help from their father for homework are far to the right of the vertical lines i.e. they have higher mean scores than the full samples of their race groups.

Figure 5: Kernal Density Graphs: Standardised LNE scores and Father's Help with Homework



(Source: Cape Area Panel Study, 2002)

Parental Educational Expectations

Studies suggest parents with higher educational expectations for their children tend to provide better financial and interpersonal resources for education, consequently in improving their child's educational outcomes (Fan and Chen, 2001; Jeynes, 2005; Hoover-Dempsey and Sandler, 1995). The table below presents mean LNE scores for each category of parental educational expectations "no matric", "matric" and "tertiary". White parents tend to have higher educational expectations for their children, 67.3% expect their children to complete tertiary compared to 53.3% of non-whites. For both groups, the mean LNE

score is lowest for those with parents who have expectations that they will not complete matric and highest for those with parents who expect their children to attend tertiary education. The LNE score differences are statistically significant at the 5% level for non-whites; however, for whites, only the difference between matric and tertiary expectations is significant. These results may suggest LNE scores increase as parental educational expectations increase for both non-whites and whites. Alternately these results may indicate the presence of a variable that jointly influences parental expectations and LNE scores. For example, genetic endowments may both raise the LNE score and influence parental expectations, where higher levels of ability translate into a parent having higher educational expectations for their child. In this case, the higher mean LNE scores is due to higher innate ability and not the parental educational expectations.

Table 8: Mean Standardised LNE Scores by Parental Educational Expectations

	No Matric	*Matric*	*Tertiary*	*Significance at 5% level*
Non-whites	-0.41	-0.02	0.14	*
Whites	0.946	0.953	1.42	-

(Source: Cape Area Panel Study, 2002)

Household Owns Computer

Computer ownership is used as a proxy measure for access to educational and financial resources. The next section on per capita household income highlights the positive relationship between family income and educational attainment. It follows that computer ownership is assumed to have a positive relationship with LNE scores. The table below presents the mean standardised LNE score separately for those from households who own a computer versus those from households who do not. For both non-whites and whites the mean LNE score is lower for young adults who belong to households who do not own a computer. The increase in LNE score associated with computer ownership is statistically significant at the 5% level and larger in the non-white sample, suggesting the impact of computer ownership on educational outcomes may be larger for non-whites. This may be because most white households (82.1%) own a computer compared to a mere 11.8% of non-white families. Whites have higher mean LNE scores than non-whites irrespective of computer ownership status of the non-white sample. Computer ownership is correlated to

household income and potentially parental education suggesting the relationship between LNE scores and computer ownership may indicate the effect of these variables.

Table 9: Mean Standardised LNE score when Household does/does not own a Computer

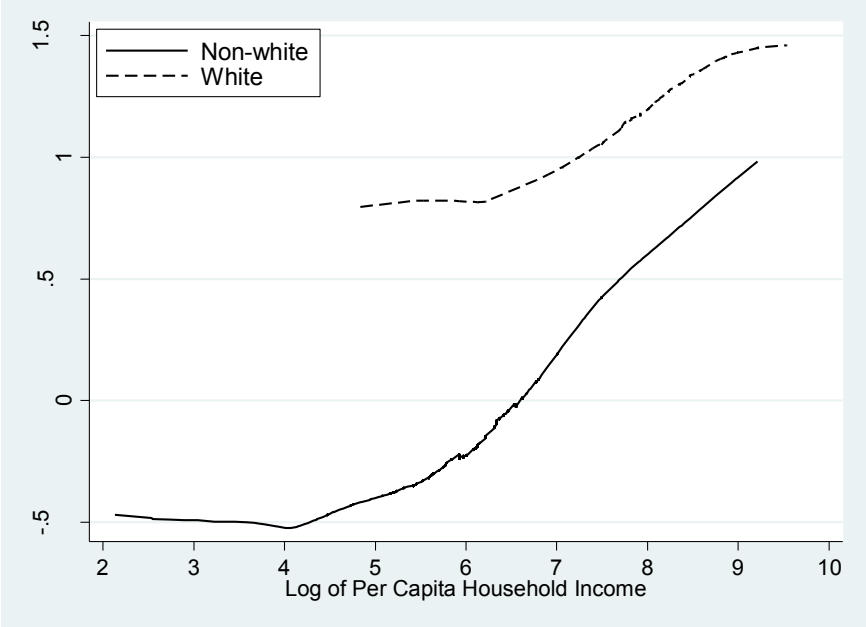
	Household Owns Computer	*Household Does Not Own Computer*	*Significance at 5% level*
Non-whites	0.62	-0.11	*
Whites	1.27	1.07	-

(Source: Cape Area Panel Study, 2002)

Per Capita Household Income

Various studies highlight that increases in household income positively impact educational outcomes through their effect on increasing access to quality educational resources (Blanden and Gregg, 2004; Davis-Kean, 2005; Lam, Ardington and Leibbrandt, 2008). The lowess graph below illustrates the relationship between the "log of per capita household income" and the LNE score. The upward sloping graphs show a positive relationship between the variables for both whites and non-whites. The graphs show that the minimum log per capita income for whites is higher (almost 5) than the minimum per capita income for non-whites (just over 2). For all levels of per capita household income, the LNE score is higher for whites than non-whites, but the difference between the race's decreases as per capita income increases, suggesting the disadvantages faced by non-whites fall as income rises. For both whites and non-whites, the graphs become steeper at higher levels of income, indicating that increases in per capita income are related to greater LNE score increases at higher levels of per capita income. The graph is, however, steeper for non-whites suggesting the relationship between per capita income and standardised LNE scores may be stronger for non-whites. For both race groups, at lower levels of per capita income, increases in household per capita income are not associated with significant increases in standardised LNE scores, suggesting there is a threshold that per capita income must reach before an associated increase in standardised LNE scores is seen. This threshold is much lower for non-whites (log of per capita household income = 4), than for whites (log of per capita household income = 6).

Figure 6: Lowess Standardised LNE Score and Log of Per capita Household Income



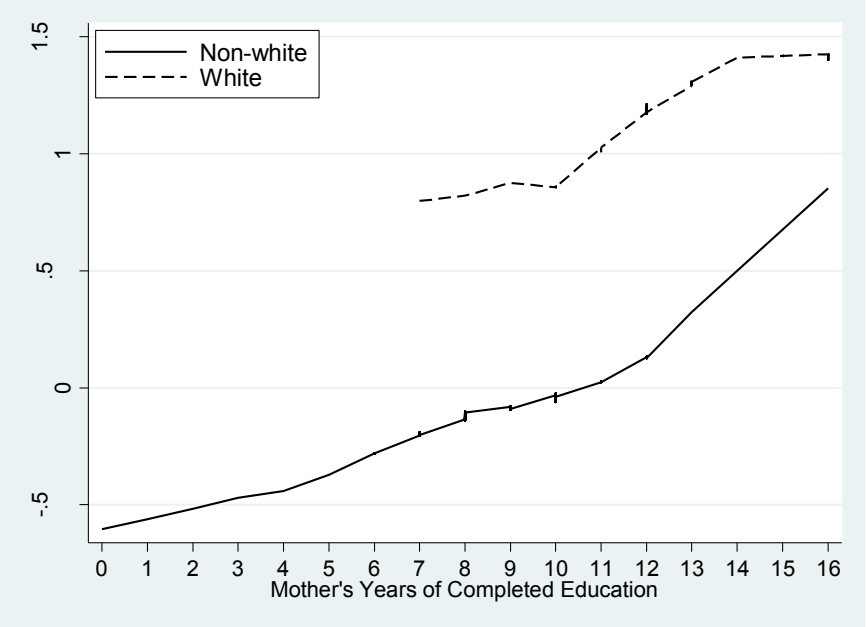
(Source: Cape Area Panel Study, 2002)

Mother's Education

It is widely accepted that a parents education plays an important role in a child's educational attainment, where higher parental education is associated with improved child educational outcomes (Anderson, Case and Lam, 2001; Davis-Kean, 2005; Dubow, Boxer, and Huesmann, 2009). South African studies have suggested that the effect of a mother's education on their child's educational outcome is greater than the effect of a father's education (Nimubona and Vencatachellum, 2007). The lowess graph below illustrates the relationship between a mother's years of completed education and standardised LNE scores, separately for whites and non-whites. The graphs are upward sloping for both whites and non-whites illustrating a positive relationship between mother's education and standardised LNE scores. All mother's of white young adults in the sample have at least 7 years of completed education, whereas in the non-white sample 26% of mothers' have less than 7 years of education. Similarly a large percentage of white mothers' (41.87%) have a tertiary education, compared to only 4.28% of non-white mothers'. For non-whites, increases in mother's education are related to increases in the standardised LNE score for all levels of mother's education. The graph, however, becomes steeper from 12 to 16 years of education, suggesting that between these levels of mother's education there is a stronger relationship between increases in mother's education and increases in LNE scores. The

graph for whites shows that increases in mother's education between 7 and 10 years are associated with almost no increases in standardised LNE scores. This suggests for whites there is a minimum threshold (i.e. 10 years) for a mother's education level before it relates to an increase in standardised LNE scores. Alternately, as only 32 mothers from the white sample of 234 have an education level between 7 and 10 years, this result may be a result of the small sample size. The graph also shows no associated increases in standardised LNE scores with increases in mother's education from 14 to 16 years, suggesting there is a plateau (i.e. 14 years) beyond which increases in mother's education have little association with increases in standardised LNE scores. Whites display higher standardised LNE scores than non-whites for all levels of mother's education.

Figure 7: Lowess Standardised LNE Score and Mothers Education



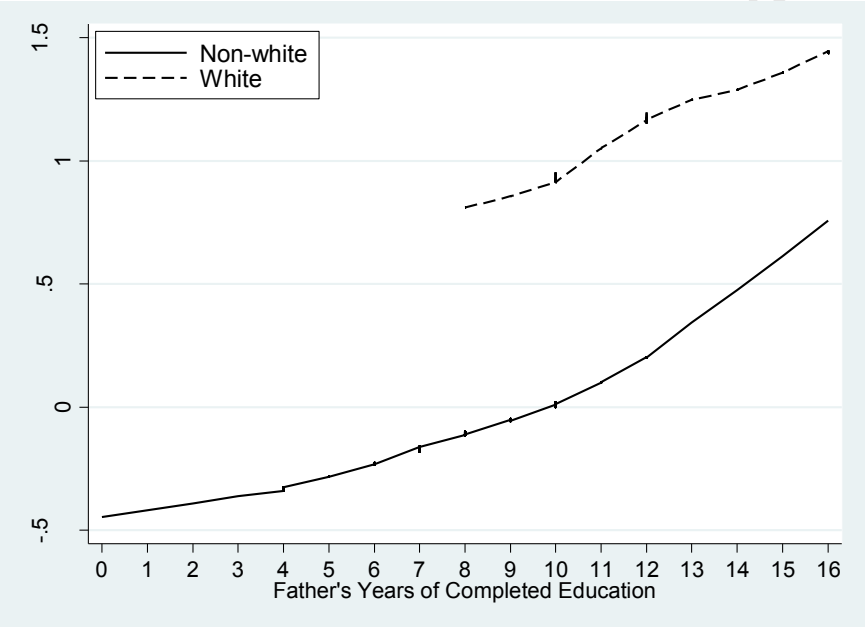
(Source: Cape Area Panel Study, 2002)

Father's Education

The graph below shows the relationship between a father's years of completed education and their child's corresponding standardised LNE score, separately for non-whites and whites. For both non-whites and whites, the graphs are upward sloping showing a positive relationship between father's education and child's standardised LNE score. White fathers' tend to have higher education levels than non-white fathers'. The minimum level of

education of a white father is 8 years whereas a staggering 60.3% of non-white fathers have an education lower than 8 years. With regards to tertiary education, 49.29% of white fathers have over 12 years of education as compared to only 3.26% of non-white fathers. For all levels of father's education, the standardised LNE score is higher for whites than for non-whites. The graphs for both non-whites and whites are quite smooth indicating increases in father's education are accompanied with gradual increases in standardised LNE scores across the whole spectrum of education levels. This is different to the finding for mother's education where the slope only increases past a minimum threshold.

Figure 8: Lowess Standardised LNE Score and Fathers Education



(Source: Cape Area Panel Study, 2002)

5.2.4. School Level Explanatory Variables

Surveying the literature and available data, this study uses three proxy measures for school and teacher quality. These are "pupil- teacher ratio", a composite variable that indicates "school problems" (teacher absenteeism, drugs and alcohol on school grounds and poor school facilities), and a series of dummy variables for self reported school quality. The latter is measured with the variables "excellent", "good", "acceptable", "poor" and "very poor". The variable selection process is highlighted in Appendix A. The weighted descriptive

statistics for the school level variables used in this study are presented below. The variable "school problems" has only 5 missing observations, the school quality dummy variables have 30 missing observations, and the pupil-teacher ratio has a very large number of missing observations (578).

Table 10. Descriptive Statistics: School Level Variables

Variables	*Obs*	*Mean*	*Std. Deviation*	*Min*	*Max*
Pupil-Teacher Ratio	1312	30.50	4.38	9	60
Pupil-Teacher Ratio Missing	1890	0.33	0.47	0	1
School Problems	1885	0.78	0.41	0	1
School Quality – Excellent	1860	0.22	0.41	0	1
School Quality – Good	1860	0.46	0.50	0	1
School Quality – Acceptable	1860	0.28	0.45	0	1
School Quality- Poor	1860	0.03	0.17	0	1
School Quality – Very Poor	1860	0.01	0.12	0	1

(Source: Cape Area Panel Study, 2002)

The table below shows the weighted means and standard deviations of the school level variables, separately for non-whites and whites. Differences between non-whites and whites are evident for all the variables. These differences are statistically significant at the 5% level for pupil-teacher ratios, school problems and four of the five school quality categories (including school quality very poor which has no white students). The means for pupil teacher ratio, school problems and poorer school quality dummy variables are higher for non-whites, suggesting poorer quality schools are attended by non-whites which is in lines with studies discussed earlier (Case and Deaton, 1999; Lam, Ardington, and Leibbrandt, 2008).

Table 11. Sample Comparison: School Level Variables

	Non-white		*White*		
Variable	*Mean*	*Std. Deviation*	*Mean*	*Std. Deviation*	*Significance at 5% level*
Pupil-Teacher Ratio	31.22	3.96	24.39	2.69	*
Pupil-Teacher Ratio Missing	0.27	0.44	0.60	0.49	*
School Problems	0.84	0.37	0.51	0.50	*
School Quality – Excellent	0.17	0.38	0.42	0.49	*
School Quality – Good	0.46	0.50	0.45	0.50	-
School Quality – Acceptable	0.31	0.46	0.13	0.33	*
School Quality- Poor	0.04	0.19	0.01	0.09	*
School Quality – Very Poor	0.02	0.13	0	0	*

(Source: Cape Area Panel Study, 2002)

5.2.5. Neighbourhood Level Explanatory Variables

Neighbourhood variables proxy socio-economic quality of a neighbourhood and include the racial breakdown of the neighbourhood, occupational and educational levels, income and the availability of facilities. Improved neighbourhood quality is typically associated with improved educational outcomes (Ellen and Turner, 1997; Wilson, 2001). This study includes measures of neighbourhood averages of unemployment "unemployment – subplace[30]", education "mean education – subplace" and income "log of mean income – subplace" as neighbourhood variables. The table below represents the weighted descriptive statistics for the neighbourhood level variables used in this study. There are no missing observations for the neighbourhood variables.

Table 12. Descriptive Statistics: Neighbourhood Level Variables

Variables	*Obs*	*Mean*	*Std. Deviation*	*Min*	*Max*
Unemployment – Subplace	1890	0.15	0.11	0	0.54
Mean Education – Subplace	1890	9.77	1.58	7.42	13.91
Log Mean Income – Subplace	1890	10.89	0.83	8.86	13.45

(Source: Cape Area Panel Study, 2002)

The table below presents weighted means and standard deviations, separately for non-whites and whites. All variables show differences between non-whites and whites which are statistically significant at the 5% level. The mean for the "unemployment –subplace" is higher for non-whites, whereas "mean education – subplace" and "log mean income-subplace" is lower for non-whites. This is expected in South African as apartheid segregated neighbourhoods on a racial basis and those from neighbourhoods that were predominantly non-white had access to fewer resources and opportunities resulting in lower neighbourhood aggregated socio-economic statuses (Burns, 2001; Fiske and Ladd, 2004).

Table 13. Sample Comparison: Neighbourhood Level Variables

	Non-white		*White*		
Variable	*Mean*	*Std. Deviation*	*Mean*	*Std. Deviation*	*Significance at 5% level*
Unemployment – Subplace	0.17	0.11	0.02	0.02	*
Mean Education – Subplace	9.18	1.00	12.43	0.82	*
Log Mean Income – Subplace	10.62	0.65	12.10	0.36	*

(Source: Cape Area Panel Study, 2002)

[30] A subplace is a measure of a geographical area or neighbourhood that is used in the South African census.

5.2.6. Value Added Variable

The value added variable "proportion of schooling passed" is included in this study to account for past achievement in terms of the value added specification of the production function. The weighted descriptive statistics for the variable is presented below. The variable has only 2 missing observations.

Table 14: Descriptive Statistics: Value Added Variable

Variables	*Obs*	*Mean*	*Std. Deviation*	*Min*	*Max*
Proportion of Schooling Passed	1888	0.54	0.06	0.25	0.8

(Source: Cape Area Panel Study, 2002)

The table below presents a sample comparison of weighted means and standard deviations between non-whites and whites, for the "proportion of schooling passed". The mean is marginally higher for whites and the difference between non-whites and whites is statistically significant at the 5% level.

Table 15: Sample Comparison: Value Added Variable

	Non-white		*White*		
Variable	*Mean*	*Std. Deviation*	*Mean*	*Std. Deviation*	*Significance at 5% level*
Proportion of Schooling Passed	0.53	0.06	0.57	0.05	*

(Source: Cape Area Panel Study, 2002)

5.3. Key Insights

There are several key points to highlight. Firstly, on average non-whites have much lower LNE scores than whites, with Africans performing the worse. This dramatic difference in LNE scores is seen across family structures and for all levels of parental involvement, parental education and family income. Secondly, for all family level variables statistically significant differences in means were seen between non-whites and whites. Non-whites on average come from larger families, are less likely to be living with both parents, have lower levels of parental involvement (such as help with homework) and have lower levels of parental education, family income and access to a computer, as compared to whites. Thirdly, it is clear that coming from a two parent family, having greater levels of parental involvement

and higher socio-economic status (parental education and family income) is positively related to LNE scores for both non-whites and whites. The relationship is particularly clear with respects to parental education and income with the lowess (locally weighted) regression graphs showing upward sloping graphs.

Section 6: Regression Analysis

Variations in the LNE score are analysed using ordinary least squares regression, adjusting for cluster sampling of the CAPS data. Weights have been used in all regressions to adjust for the survey design as well as household and individual non-response. Robust standard errors are displayed for all regressions. This section of the paper comprises of two sub sections: category specific regressions (i.e. individual, family, school and neighbourhood), and education production function regressions which combine the individual, family and school level variables.

6.1. Category Specific Regressions

Regressions for each category (individual, family, school and neighbourhood) are run separately in order to highlight category specific correlations between the explanatory variables and the standardised LNE score. Two separate regressions are run for each category. The first regression combines Africans and coloureds (non-whites), and controls for race. The second regression includes only whites from the sample. This racial separation is justified by the lack of common support to combine non-whites and whites, discussed in section 5.

Value added and contemporaneous specification regressions are run in each category and key differences between the two are highlighted. Value added regressions are presented in the main body of this study as various authors suggest the specification is a better approximation of reality as it controls for past achievement (Hanushek, 1979, 2003; Todd and Wolpin, 2006; Lamdin, 1996). The contemporaneous regressions are presented in Appendix B. The focus of this study is in understanding the role of the family within an education production function that controls for other variables that may impact educational outcome. The individual, school and neighbourhood specific regressions are therefore not the focus of this study and have been included in order to build the education production function. Furthermore, to emphasise the family, three additional sets of regressions for each component of family variables (i.e. family structure, parental involvement, family income and parental education) were run separately for non-whites and whites. This allows insight into relationships between the sub-categories.

6.1.1. Individual Level Regressions

The two regressions below show how the LNE score varies with respect to individual level determinants of education, separately for non-whites and whites. The regressions are defined in terms of the value added specification and control for past achievement[31] through the inclusion of the variable "proportion of schooling passed".

The control variable 'African' is negatively correlated to LNE scores and statistically significant. Given that the reference category is coloured this suggests that African learners have worse LNE scores than coloured learners on average, even when individual level variables, genetic endowments and historical inputs (represented by past achievement) are controlled for. For both non-whites and whites, "male", "educational expectation tertiary" and "proportion of schooling passed" are positive and statistically significant; "educational expectations matric" is positive but not significant. The coefficient for "male" is marginally smaller in the white regression than in the non-white regression, while the coefficients for "educational expectations tertiary" and "educational expectations matric" are fairly larger in the white regression than the non-white. For non-whites, "study time" is also positive and significant. It is expected that studying harder and having higher educational expectations for oneself should be positively related to test scores, hence these results are in line with other studies (Beal and Crockett, 2010; Deary, Strand, Smith and Fernandes, 2007).

These results suggest various things. The individual variables as a whole are significant, indicating that individual level variables have a role in explaining the variations in the LNE score. Although there seems an advantage to being male for both race groups, this advantage is larger for non-whites. Higher educational expectations have a positive role to play for both non-whites and whites, however the strength of this role may be higher for whites.

[31] Past achievement is a proxy for the combination of genetic endowments and all historical inputs into the education process (Todd and Wolpin, 2003).

Table 16: Individual Level Regressions for Standardised LNE Score - Value Added Specification

Variable	*Non-white*	*White*
African *(ref = Coloured)*	-0.556***	
	(0.053)	
Age	0.0112	-0.0332
	(0.0114)	(0.0235)
Male	0.192***	0.140**
	(0.0387)	(0.0684)
Study Time	0.0105***	0.0036
	(0.00262)	(0.004)
Educational Expectation Matric	0.0881	0.338
	(0.0806)	(0.24)
Educational Expectation Tertiary	0.377***	0.684***
	(0.0798)	(0.242)
Proportion of Schooling Passed	4.713***	3.205***
	(0.337)	(0.944)
Constant	-2.927***	-0.688
	(0.26)	(0.645)
Observations	1547	224
R-squared	0.344	0.253

Robust standard errors in parentheses. *** Significant at 1% level ** Significant at 5% level
*Significant at 10% level

There are two key differences between regressions presented above and the contemporaneous regressions presented in Appendix B. In the contemporaneous specification, "educational expectations matric" becomes statistically significant at the 10% level in the non-white regression. In the white regression "male" is no longer statistically significant although it is still positively related to LNE score.

6.1.2. Family Level Regressions

Family level characteristics have been divided into three main sub-categories, family structure, parental involvement and family income and parental education. Each sub-category is investigated separately using OLS regression analysis. The variables from these sub-categories are then combined in a family specific regression in terms of the value added and contemporaneous specifications of the education production function. The sub-category regressions have the advantage of showing how a small set of variables are related to LNE scores and this is helpful in interpreting the more complicated, full family level model.

6.1.2.1. Family Structure

The two regressions below show how the standardised values of the LNE score vary with respect to family structure, separately for non–whites and whites. Similarly, to the individual level non-white regression "African" is negative and statistically significant, indicating the African disadvantage when compared to coloureds. In the non-white regression both family structure variables are statistically significant when controlling for race, "household size" is negatively related to LNE scores, and fairly small (-0.0233); "proportion lived with both parents" is positively related to LNE scores. For whites only the "proportion lived with both parents" is significant; the coefficient on the variable is positive and larger for whites (0.229) than non-whites (0.129). These results suggest that household size has no impact on whites, but negatively impacts the educational outcomes of non-whites. Furthermore it appears that living in a nuclear family has a positive impact on both race groups, however the effect is larger for whites.

These results are in line with international studies showing the negative impact of larger families on educational outcomes and positive impacts of living in a nuclear family (Ginther and Pollak, 2004; van Eijck and de Graaf, 1995; Biblarz and Raftery, 1999). However, the negative relationship of household size and educational outcomes for non-whites is different compared to other South African studies, which have shown a positive relationship (Case and Deaton, 1999) or no relationship (Lu, 1999).

Table 17: Family Structure Regression for Standardised LNE Score

Variable	*Non-white*	*White*
African *(ref = Coloured)*	-0.643***	
	(0.0560)	
Household Size	-0.0233***	0.0357
	(0.00825)	(0.0309)
Proportion lived with both Parents	0.129**	0.229**
	(0.0538)	(0.109)
Constant	0.292***	0.909***
	(0.0708)	(0.171)
Observations	1655	233
R-squared	0.164	0.022

Robust standard errors in parentheses. *** Significant at 1% level ** Significant at 5% level
*Significant at 10% level

6.1.2.2. Parental Involvement

The regressions below show the variation of LNE scores with respect to measures of parental involvement, separately for non-whites and whites. The sample size for both regressions is smaller than that of the family structure regression due to the large number of missing values for the parental involvement variables. Similarly to the family structure regression, "African" in the non-white regression is negative and statistically significant.

Studies suggest that any positive parental involvement activity such as assisting the child with homework and higher educational expectations for the child should be positively correlated to educational outcomes (Fan and Chen, 2001; Jeynes, 2005; Tan and Goldberg, 2009). This is reflected in the non-white regression where all variables are positive and statistically significant, except "mother helps with homework" which is negative and significant. This suggests that a father helping with homework has the expected result of improving a child's educational outcomes, whereas it is not clear that a mother's help yields a positive impact. However, the coefficient for "mother helps with homework" is not large (-0.101). The coefficient for "parental expectations tertiary" is larger (0.567) than "parental expectations matric" (0.233). This suggests that parental expectations may positively impact LNE scores, and higher expectations have a greater impact. Alternatively, parents with higher expectations may base their expectations on their child's superior genetic endowments or the family's greater access to financial resources (Hoover-Dempsey and Sandler, 1995; Miller, Mulvey and Martin, 2001). In this case the positive relationship between parental expectations and LNE scores may actually be representing the potential relationship between genetic endowments and financial resources on the one hand and LNE scores on the other.

In the white regression all variables are statistically insignificant. This does not mean that parental involvement does not matter, as the model as a whole is significant. Rather this suggests that for whites, as levels of parental involvement are high for most children, it is not a key factor in explaining the variations in outcomes among whites. As illustrated in the descriptive statistics the level of parental help with homework as well as parental expectations to complete matric and tertiary are uniformly high in the white sample. This

lack of variation in the white sample suggests that parental involvement measures are not able to explain variations in education attainment amongst whites because they are uniformly high as opposed to being unimportant. One could further still suggest that the higher educational outcomes of whites on average, as compared to non-whites could in part be linked to the higher levels of parental involvement for whites as compared to non-whites (highlighted in the descriptive statistics).

Table 18: Parental Involvement Regression for Standardised LNE Score

Variable	*Non-white*	*White*
African *(ref = Coloured)*	-0.759***	
	(0.0551)	
Mother Helps with Homework	-0.101*	-0.0989
	(0.0543)	(0.0752)
Father Helps with Homework	0.134**	0.0993
	(0.0678)	(0.0939)
Parental Expectations – Matric	0.233***	0.0283
	(0.0774)	(0.439)
Parental Expectations – Tertiary	0.567***	0.479
	(0.0757)	(0.433)
Constant	-0.102	0.947**
	(0.0822)	(0.444)
Observations	1532	217
R-squared	0.221	0.177
Robust standard errors in parentheses. *** Significant at 1% level ** Significant at 5% level *Significant at 10% level		

6.1.2.3. Socio-economic Status

The regressions below show correlations between various measures of socio-economic status ("household owns computer", "log of per capita family income" and parental education) and standardised LNE scores, separately for non-whites and whites. As presented in the previous non-white regressions "African" is negative and statistically significant. Given that the reference category is coloured this suggests that African learners have worse LNE scores than coloured learners on average, even when socio-economic status is controlled for.

The variable "household owns computer" is a proxy measure for resources available in the household for educational purposes. It is assumed that household's with a computer have a socio-economic advantage and including greater resources for education, which is

suggested to be positively related to educational outcomes (Blanden and Gregg, 2004; Sirin, 2005). This is evident in the non-white regression as the variable household owns a computer is positive and statistically significant. In the white regression "household owns computer" is negative but not significant, probably because there is little variation in the white sample (i.e. 82.1% of whites have a computer in their household).

Studies suggest a positive relationship between higher levels of family income and parental education on the one hand and improved educational outcomes on the other hand (Davis-Kean, 2005; Lam, Ardington and Leibbrandt, 2008; Aakvik, Vaage, and Salvanes, 2005; Dubow, Boxer, and Huesmann, 2009). For both non-whites and whites the "log of per capita family income" and "mother's education" are positive and statistically significant. The coefficients of "log of per capita family income" and "mother's education are reasonably small for both race groups. For both variables, the coefficients are smaller for non-whites than whites, (0.0867) compared to (0.175) for "log of per capita family income" and (0.0344) compared to (0.0619) for "mother's education". Although "father's education" is also positive in both regressions it is not significant. These results suggest income constraints play an important role in the educational outcomes of all races in South Africa, with lower incomes translating to lower LNE scores on average. Furthermore, in the South African context a mother's education may be more important than a father's education in a child's educational outcomes. This has in fact been suggested by Nimubona and Vencatachellum (2007), in their study of intergenerational education mobility in South Africa. The coefficients of "log of per capita family income", "mother's education" and "father's education" are marginally larger for whites than non-whites, suggesting that socio-economic status may have a stronger relationship to educational outcomes for whites.

Table 19: Socio-economic Status Regression for Standardised LNE Score

Variable	*Non-white*	*White*
African *(ref = Coloured)*	-0.468***	
	(0.0562)	
HH Own Computer	0.331***	-0.0317
	(0.0676)	(0.0951)
Log Per Capita Household Income	0.0867***	0.175***
	(0.0264)	(0.0553)
Mother's Education Missing	0.189*	0.441
	(0.102)	(0.371)
Mother's Education	0.0344***	0.0619***
	(0.00902)	(0.0221)
Father's Education Missing	0.0394	0.517*
	(0.0882)	(0.297)
Father's Education	0.0116	0.0348
	(0.009)	(0.024)
Constant	-0.803***	-1.402***
	(0.161)	(0.427)
Observations	1656	234
R-squared	0.244	0.174

Robust standard errors in parentheses *** Significant at 1% level ** Significant at 5% level
*Significant at 10% level

The explanatory variables from the previous three family sub-category regressions are combined below to provide a full family level category regression, separately for non-whites and whites. In the non-white regression the family structure variables "household size" and "proportion lived with both parents" lose their statistical significance when combined with other family variables. A test of joint significance further indicates these variables are not significant as part of the full family level model. Studies suggest that the role of family structure on educational outcomes is mediated through the effect of financial and interpersonal resources on educational outcomes. Therefore, although a negative effect of family structure may exist, it largely operates through limiting the income and parental time available to provide educational financial and interpersonal resources (van Eijck and de Graaf, 1995; Downey 1995). The inclusion of "log per capita family income" and "mother/father helps with homework" controls for the financial and interpersonal resources provided to a child in the current model. The loss of significance of the family structure variables could therefore suggest that in the South African context, the effect of family structure operates largely indirectly through limiting the resources available for education. For whites, "household size" which is positive becomes statistically significant in the full family level regression. This suggests that when controlling for the effect of family size on

limiting the per capita family income, larger families may actually provide support and encouragement thereby improving educational outcomes. This is surprising as previous South African studies have associated such results solely to Africans (Case and Deaton, 1999).

For non-whites, "father helps with homework" loses its statistical significance, while "mother helps with homework" is still negative and significant. The parental expectations variables retain their statistical significance, although their coefficients are smaller than the parental involvement regression previously presented. These results reflect that for non-whites parental involvement has a role to play in explaining educational outcomes even after controlling for socio-economic status. Furthermore, socio-economic status also has a role to play in non-whites outcomes as "mother's education" and "household owns computer" are both positive and significant. For whites, results suggest that socio-economic status may play a greater role in educational outcomes as compared to parental involvement, as both "mother's education" and "log of per capita family income" are positive, significant and substantially larger than in the non-white regression; whereas the parental involvement measures are not individually significant. However, a test of joint significance of the parental involvement measures in the white regression, indicates the variables are jointly significant at the 1% level, supporting their inclusion in the model and indicating parental involvement may still play a role in white outcomes.

Table 20: Family Level Regression for Standardised LNE score - Value Added Specification

Variable	*Non-white*	*White*
African *(ref = Coloured)*	-0.433***	
	(0.0569)	
Household Size	-0.000816	0.0503*
	(0.00734)	(0.0301)
Proportion lived with both Parents	0.0620	0.0159
	(0.0517)	(0.109)
Mother Helps with Homework	-0.0938*	-0.0412
	(0.0529)	(0.0671)
Father Helps with Homework	0.0304	0.105
	(0.0590)	(0.0789)
Parental Expectations – Matric	0.128**	0.00862
	(0.0609)	(0.366)
Parental Expectations – Tertiary	0.333***	0.308
	(0.0593)	(0.361)
Household Own Computer	0.281***	0.0170
	(0.0593)	(0.0979)
Log Per Capita Household Income	0.0331	0.102**
	(0.0272)	(0.0440)
Mother's Education Missing	0.217**	0.0611
	(0.0956)	(0.389)
Mother's Education	0.0274***	0.0524***
	(0.00874)	(0.0174)
Father's Education Missing	-0.0354	0.253
	(0.0902)	(0.207)
Father's Education	-0.000404	0.0121
	(0.00879)	(0.0157)
Proportion of Schooling Passed	4.178***	2.256**
	(0.324)	(0.947)
Constant	-2.789***	-2.115**
	(0.257)	(0.848)
Observations	1530	216
R-squared	0.360	0.336

Robust standard errors in parentheses. *** Significant at 1% level ** Significant at 5% level
*Significant at 10% level

It is noted that multicollinearity may be present, in such a model where family level variables are related to each other. As presented in correlation matrix in Appendix D parental education is correlated to family income. Appendix D shows that for both whites and non-whites there is a positive and significant correlation between both "fathers' education" and "mother's education" on the one hand and the "log of per capita family income" on the other hand. For whites the realtionship between mother's education and family income (0.156) is smaller as compared to non-whites (0.298), while the relationship between father's education and family income is marginally larger for whites (0.363) than for non-whites (0.331). Appendix D also illustrates a negative correlation between measures of socioeconomic status on the one hand (i.e. "log per capita household income"

and parental education) and household size on the other hand. For non-whites household size is negatively and statistically correlated to "log per capita household income" (-0.259), mother's education (0.137) and fathers education (0.082). For whites, only the correlation between fathers education and household size is significant (0.172).

These correlations are not unique and studies of this nature often present positive correlations between family level variables, especially with regards to parental education and family income (Aakvik, Vaage, and Salvanes, 2005; Dubow, Boxer, and Huesmann, 2009) and family socioeconomic status and household size (Downey, 1995; Ginther and Pollak, 2004; Anderson, 2000). It is for this reason that sound theoretical principles have guided variable selection and interpretation of results in this study (Davis-Kean, 2005; Ermisch and Francesconi, 2001).

There are two key differences between the family level value added regressions presented above and the contemporaneous regressions presented in Appendix B. Firstly, "log of per capita income" becomes statistically significant in the non-white contemporaneous regression. Secondly, "household size" loses its significance for whites, in the contemporaneous regression.

6.1.3. School Level Regressions

The two regressions below show how the standardised values of the LNE score vary with respect to the school level determinants of education, separately for non–whites and whites. These regressions are modelled in terms of the value added specification of the education production function and therefore control for past achievement through the inclusion of "proportion of schooling passed".

Similar to the individual and family level regressions, "African" is negative and statistically significant, indicating that Africans perform worse than coloureds on average even when school level variables and past achievement is controlled for. Studies suggest that higher pupil-teacher ratios and the presence of key school problems are a reflection of poor school quality which consequently negatively impacts educational outcomes (Case and Deaton,

1999; Anderson, Case, and Lam, 2001; Boissiere, 2004). In lines with these studies, for both non-whites and whites "pupil teacher ratio" and "school problems" are negatively related to LNE scores. Although "pupil teacher ratio" is statistically significant in both regressions, "school problems" is only significant for whites. The coefficient for "pupil-teacher ratio" is larger for whites than non-whites.

With "school quality – excellent" as the reference category, all the school quality dummy variables are statistically significant and negative for non-whites. This indicates the disadvantage of attending a poor quality school. For whites only "school quality – acceptable" is significant. It is important to note that the variables measuring school quality are self reported by young adults and therefore reflect an element of subjectivity and personal opinion.

These results suggest that for both non-whites and whites, school level variables have an important role in explaining variations in LNE scores. Both regressions are statistically significant as a whole. Although the R-squared is marginally smaller for whites than non-whites; the larger coefficient of "pupil-teacher ratio" and the significance of "school problems" suggest that school quality may potentially have a greater role (or at least the same role) to play in the educational outcomes of whites as compared to non-whites.

Table 21: School Level Regressions for Standardised LNE Score - Value Added Specification

Variable	*Non-white*	*White*
African *(ref = Coloured)*	-0.454***	
	(0.0549)	
Pupil-Teacher Ratio	-0.0202***	-0.0564***
	(0.00663)	(0.0175)
Pupil-Teacher Ratio Missing	-0.402*	-1.369***
	(0.213)	(0.443)
School Problems	-0.00994	-0.160**
	(0.0524)	(0.0720)
School Quality – Good	-0.142***	-0.0288
(ref - Excellent)	(0.0528)	(0.0679)
School Quality – Acceptable	-0.120**	-0.430***
(ref - Excellent)	(0.0573)	(0.117)
School Quality – Poor	-0.300***	-0.0170
(ref - Excellent)	(0.0907)	(0.0823)
School Quality - Very Poor	-0.288**	0
(ref - Excellent)	(0.119)	(0)
Proportion of Schooling Passed	4.719***	2.641**
	(0.339)	(1.039)
Constant	-1.659***	1.270*
	(0.291)	(0.730)
Observations	1623	230
R-squared	0.313	0.252

Robust standard errors in parentheses. *** Significant at 1% level ** Significant at 5% level
*Significant at 10% level

There are no significant differences in the school level contemporaneous specification regressions in Appendix B and the value added specification above.

6.1.4. Neighbourhood Level Regressions

The regressions below show how the LNE score varies with respect to the neighbourhood level variables, separately for non–whites and whites. Similarly to the other category specific regressions, these regressions control for past achievement. Although not statistically significant, in both the non-white and white regressions "unemployment – subplace" is positively related to the LNE score. This is not expected as studies suggest that higher neighbourhood unemployment is associated with exposure to negative influences which in turn reduces educational attainment (Ellen and Turner, 1997; Jencks and Mayer, 1990). This may be the because "unemployment-subplace" and "log mean income – subplace are highly correlated to each other as indicated in Appendix D. For both whites and non-whites the variables are significantly negatively correlated, with the correlation larger for non-whites (0.877) than whites (0.504). The variable "mean education –

subplace" is positive and statistically significant for non-whites, while "log mean income – subplace" is positive and statistically significant for whites. This suggests that the education of a neighbourhood has a greater role to play in impacting the educational outcomes of non-whites, while for whites the income of a neighbourhood is the most important measure of socio-economic advantage. This may be because white adults on average have relatively higher level of education than their non-white counterparts (Fiske and Ladd, 2004; Van der Berg and Burger, 2002). The neighbourhood level regressions are both significant as a whole. However, R-squared for non-whites (0.319) is substantially larger than for whites (0.188), suggesting that neighbourhood variables as a whole may play a bigger role in explaining the variation in LNE scores for non-whites.

Table 22: Neighbourhood Level Regressions for Standardised LNE Score - Value Added Specification

Variable	*Non-white*	*White*
African *(ref = Coloured)*	-0.431***	
	(0.103)	
Unemployment – Subplace	0.202	3.240
	(0.529)	(2.286)
Mean Education – Subplace	0.127***	-0.0260
	(0.0457)	(0.0864)
Log Mean Income – Subplace	0.0799	0.555***
	(0.0983)	(0.186)
Proportion of Schooling Passed	4.742***	2.523***
	(0.317)	(0.935)
Constant	-4.430***	-6.657***
	(0.878)	(1.581)
Observations	1655	233
R-squared	0.319	0.188

Robust standard errors in parentheses. *** Significant at 1% level ** Significant at 5% level
*Significant at 10% level

The only significant difference between the neighbourhood level contemporaneous specification in Appendix B and the value added regression above is the statistical significance of "unemployment – subplace" in the contemporaneous white regression

The neighbourhood variables were regressed against race to ascertain the degree of explanatory power that race has in each case. The variable "unemployment – subplace" had the highest correlation with race (R-squared 0.767), followed by the "log mean income – subplace" (R-squared 0.679) and "mean education – subplace" (R-squared 0.627). These regressions are presented in Appendix C. This correlation is expected as the legacy of

apartheid has resulted in most South African neighbourhoods being racially driven (Fiske and Ladd, 2004). The education production function controls for race in the non-white regression and further runs a separate regression for whites. As such, no neighbourhood variables are included in the education production function in section 6.2, as their inclusion would not provide much explanatory power when race is accounted for.

6.2. Education Production Function Regressions

This section presents an education production function that jointly regresses the individual, family and school level variables on the standardised LNE score, in terms of the value added specification. Neighbourhood variables have not been included in these regressions due to the high level of correlation with race, as discussed in section 6.1.4. Similar to the other regressions, non-whites (Africans and coloureds) and whites are presented separately.

As was seen in the previous regressions "African" is negative and statistically significant. Given that the reference category is coloured this suggests that African learners have worse LNE scores than coloured learners on average, even when individual, family and school variables as well as genetic endowments and historical inputs (represented by past achievement) are controlled for. In both the non-white and white regressions, "proportion of schooling passed" is positive and significant. The coefficient for the variable is much smaller for whites suggesting that past achievement may have a greater role in determining future achievement for non-whites as compared to whites. This result is seen accross all the value added regressions in this study.

The individual level variables "male", "study time" and "educational expectation – tertiary" are positive and statistically significant for non-whites. For whites only "educational expectation – tertiary" is positive and significant, and the coefficient is larger than in the case for non-whites. A test of joint significance showed that as a whole the individual level variables are significant for both non-whites (1% level) and whites (10% level), indicating the importance of their inclusion in the model.

For non-whites, it is interesting that there appears to be an advantage to being male, as males have on average higher LNE scores. Although this is in line with what is seen in other African countries, South African studies have typically showed either no gender differences (Anderson, Case, and Lam, 2001; Case and Deaton, 1999) or a female advantage Nimubona and Vencatachellum, 2007). For non-whites, the positive importance of academic effort captured in "study time" is in lines with the results from Van der Berg (2008). For whites "study time" is still positive although not significant, indicating that academic effort has a greater role to play in explaining variations in educational outcomes amongst non-whites as compared to whites. It is interesting that "study time" is significantly and positively correlated to parental education for non-whites, but not so for whites as presented in the correlation matrix in Appendix D. This illustrates the indirect effects of parental education in influencing the behaviour of an individual. The will to achieve proxied by educational expectations is positively related to outcomes for both race groups, largely in lines with results from international studies (Beal and Crockett, 2010; Poropat, 2009).

The family structure variables are not significant for either non-whites or whites. As discussed above this reflects that family structure seems to operate primarily through limiting financial and interpersonal resources for education. Hence, once these factors are controlled for (through measures of family income and parental involvement) family structure no longer plays a significant role in educational achievement. Regarding measures of parental involvement for non-whites the parental involvement measures "mother helps with homework" and "parental expectations – tertiary" are both significant. As expected, expectations are positively related to LNE scores, however "mother helps with homework" is still surprisingly negative. The persistence of the statistical significance and negative relationship between "mother helps with homework" and LNE scores, even in the full education production function is interesting. From the descriptive statistics it is evident that parental education levels are much lower for non-whites than whites. This suggests that relatively uneducated mothers' may be helping their children with homework with little benefit to their children's educational outcomes. None of the parental involvement measures are significant for whites.

Table 23: Education Production Function Regressions for Standardised LNE Score - Value Added Specification

Variable	*Non-white*	*White*
African *(ref = Coloured)*	-0.466***	
	(0.0644)	
Age	0.00795	0.0134
	(0.0117)	(0.0293)
Male	0.196***	0.0629
	(0.0375)	(0.0673)
Study Time	0.00681***	0.000496
	(0.00260)	(0.00398)
Educational Expectation – Matric	0.0415	0.321
	(0.0805)	(0.281)
Educational Expectation – Tertiary	0.240***	0.556*
	(0.0756)	(0.295)
Household size	0.000132	0.0410
	(0.00790)	(0.0324)
Proportion lived with both Parents	0.0473	-0.0223
	(0.0491)	(0.106)
Mother Helps with Homework	-0.0948*	-0.00559
	(0.0539)	(0.0666)
Father Helps with Homework	0.0232	0.0673
	(0.0649)	(0.0832)
Parental Expectations – Matric	0.0823	-0.0624
	(0.0703)	(0.265)
Parental Expectations – Tertiary	0.193***	0.0199
	(0.0674)	(0.248)
Household Own Computer	0.229***	-0.0812
	(0.0590)	(0.107)
Log Per Capita Household Income	0.0111	0.103**
	(0.0280)	(0.0470)
Mother's Education Missing	0.161*	-0.0792
	(0.0945)	(0.391)
Mother's Education	0.0219**	0.0471**
	(0.00862)	(0.0196)
Father's Education Missing	-0.00860	0.380
	(0.0874)	(0.283)
Father's Education	0.000297	0.0103
	(0.00839)	(0.0188)
Pupil-Teacher Ratio	-0.0185***	-0.0516***
	(0.00622)	(0.0178)
Pupil-Teacher Ratio Missing	-0.431**	-1.330***
	(0.198)	(0.466)
School Problems	0.0390	-0.102
	(0.0535)	(0.0656)
School Quality – Good	-0.0400	-0.0291
(ref - Excellent)	(0.0554)	(0.0679)
School Quality – Acceptable	-0.0568	-0.242*
(ref - Excellent)	(0.0582)	(0.129)
School Quality – Poor	-0.131	0.0398
(ref - Excellent)	(0.0905)	(0.212)
School Quality - Very Poor	-0.106	0
(ref - Excellent)	(0.133)	(0)
Proportion of Schooling Passed	3.938***	2.273**
	(0.344)	(0.965)
Constant	-2.264***	-1.016
	(0.400)	(1.107)
Observations	1432	210
R-squared	0.397	0.426

Robust standard errors in parentheses. *** Significant at 1% level ** Significant at 5% level
*Significant at 10% level

This suggests that while parental involvement specifically parental educational expectations are an important variable in explaining variations in outcomes for non-whites, it is not the case for whites. A plausible explanation is that as levels of parental involvement are high for most white children, it is not a key factor in explaining the *variations* in outcomes among whites. Hango (2007) finds in the US that parental involvement has a greater role to play in explaining the educational outcomes of those from socio-economic disadvantage, and further serves to overcome the effects of socio-economic disadvantage. As non-whites have lower levels of socio-economic status reflected by lower levels of income and parental education, this study shows similar results in the South African context.

The socio-economic variables "household owns computer" is positive and statistically significant for non-whites, while "log per capita household income" is positive and significant for whites. For non-whites resources specifically related to education (computer ownership) seems to have a larger impact than income per capita. Mother's education is important for both non-whites and whites as the variable is positive and significant, the size of the coefficient is larger for whites than non-whites. Father's education is negative for non-whites which is surprising, but as expected it is positive for whites, however it is not significant in both cases. These results suggest that socio-economic status has a very important role to play in explaining educational attainment for both non-whites and whites. The importance of parental education has been established by various South African studies (Van der Berg, 2008; Case and Deaton, 1999; Anderson, Case, and Lam, 2001). Furthermore, Nimubona and Vencatachellum (2007) have found as highlighted in this study that mother's education may have a more importa role to play than a fathers' education.

Parental education and family income has a larger effect on white educational outcomes as compared to non-whites. This suggests that socio-economic status has a larger impact on whites, Van der Berg (2008) finds similarly. The author finds that socio-economic status has a larger effect above a threshold level of socio-economic status; furthermore those attending the richest schools (typically whites) have a larger sensitivity to changes in socio-economic status particularly income levels.

Tests of joint significance for the family level variables show that although they are not all individually significant as a group they are significant at the 1% level for both non-whites and whites. These results show that family level variables are important factors explaining the variations in educational outcomes for both non-whites and whites, even after controlling for individual, school and past achievement variables, and supports their inclusion in the education production function.

For both non-white and white groups "pupil-teacher ratio" is negative and significant, the coefficient is substantially larger for whites, suggesting whites may be more sensitive to changes in this ratio. For whites, the school level quality dummy "school quality – acceptable" is also negative and significant, when the default value is "school quality – excellent" As a whole a test of joint significance shows that the school level variables are significant for both non-whites (1% level) and whites (5% level).

These results indicate that reductions in school quality (where lower pupil-teacher ratios are associated with better quality schools) are correlated to lower LNE scores for both race groups as is seen in other South African studies (Anderson, Case, and Lam, 2001; Case and Yogo, 1999; Case and Deaton, 1999). This effect exists even after accounting for the effect of individual and family variables as well as genetic endowments and historical inputs (proxied by "proportion of schooling passed"). The "pupil-teacher ratio" results, the insignificance of "school problems[32]", and the fact that a large portion of variations in LNE scores remain unaccounted for in the education production function supports findings by Lam, Ardington, and Leibbrandt (2008) and Van der Berg (2008), that other less quantifiable measures of school quality such as school administration and teacher quality may have an important role in explaining education outcomes, especially among non-whites.

There are no significant differences between the value added regressions above and the contemporaneous education production function presented in Appendix B. That said, there

[32] School problems is a composite measure that measures if any of the following school problems exist, shortage of textbooks, dirty classrooms, crowded classrooms, teacher absence, teacher drunk, teacher threatened by students, noisy classrooms, drug dealing at school, unsafe environment at school, sexual harassment by teachers or students, bullying by students and violent teachers.

is one noteworthy difference. The size of most of the significant coefficients are slightly reduced when past achievement is controlled for. This is expected as past achievement measures the impact of historical inputs on LNE scores, and current input measures contain an element of historical input.

Section 7: Conclusions

In this study I set out to investigate the role of the family in determining in a child's educational attainment, and further highlight the differences between non-whites (Africans and coloureds) and whites. As there is no standard theoretical model or common set of explanatory variables used in studies of education progress, a thorough survey of the literature was required to establish the key determinants of education, and further integrate the various categories of determinants (individual, family, school and neighbourhood) in a coherent way. This was used to set out the holistic education production function model used in the study and select the most appropriate variables for use in the model. The holistic education production function is the strength of this study, as it enables the investigation into the role of the family while controlling for other determinants, especially those at the school level.

The CAPS provided a very rich data set that integrated information on the individual, family, school and neighbourhood. CAPS data set was supplemented with data from the School Registry of Needs and the Census, to provide school and neighbourhood information respectively. The process of variable selection required an in depth exploration of the data set, and variables were included in the empirical model primarily on the basis of their theoretical validity.

This study highlighted three key sub – categories of family level variables, family structure, parental involvement and socio-economic status. Key differences between non-whites and whites were seen with regards to all these family level variables, with whites on average coming from a more advantageous family background than non-whites. Whites tend to come from smaller families, are more likely to live with both parents, recieve higher levels of parental involvment and have higher levels of income and parental education as compared to non-whites.

Multiple regression analysis was used to further investigate these differences. Empirical models were defined in terms of both the contemporaneous and value added specifications of the education production function. The former specification includes only current

variables, while the latter controls for past achievement. Value added specifications are considered superior as past achievement proxies for genetic endowments and the historical influence of variables. Value added regressions were presented in the main text of the study, although a key result was that, when individual, family and school level variables are accounted for, there is no significant difference between the contemporaneous and value added specifications. Regressions were run separatly for non-whites and whites in order to highlight the key racial differences. Firstly, category specific (i.e. individual, family, school and neighbourhood) regressions were presented to identify the relationships between category specific variables and the outcome variable. The explanatory variables for the individual, family and school category regressions were then combined into an education production function regression that addresses the impact on educational attainment holistically. This holistic regression allowed investigation of the role of family level variables while controlling for variables across other categories.

The regression analysis finds that the family does indeed have a role to play in determining the educational outcomes of South African children, and further this role is different across race groups. For both race groups family structure consisiting of family size and family type does not seem to have a direct effect on educational outcomes. Rather it is suggested that these variables operate indirectly through their impact on limiting parental resources available to children, particularly financial resources. Parental involvement, especially parental educational expectations for their children plays a role in explaining variations in educational outcomes of non-whites, but plays a lesser role for whites. This is line with studies highlighting the increased importance of parental involvement in those of lower socio-economic status (Hango, 2007). The most interesting finding is the role of socio-economic status in terms of parental education and family income seems to be more important for whites than non-whites. This confirms findings from other studies that socio-economic status has a greater role to play for those from higher socio-economic statuses (Van der Berg, 2008).

Lastly, this study highlighted that each category of determinants, individual, family and school has a role to play in explaining variations in educational outcomes. Tests of joint significance for each category of inputs in the education production function regressions

showed each category to be statistically significant in both the non-white and white regression, supporting their inclusion in the model. It is therefore imperative that studies on education adequately account for the holistic process of education.

The results of this study highlight the need for educational policy in South Africa to encourage parental involvement specifically amongst non-white familes, as a means to improve educational outcomes. The current emphasis of educational policy has been on the improvement of school quality and access to education amongst non-whites (Van der Berg, 2008; Case and Yogo, 1999), without adequate emphasis on parental involvement. The results of this study indicate clearly that school quality matters has an important role to play on educational outcomes in the South African context. However, it is also evident that the positive effects of school quality can be supported through programs targeting parental involvement, specifically in communities where parental involvement levels are low (non-whites). The lack of parental involvement often stems from a lack of parental education impacting the ability to support the educational progress of children. Programs targeting adult literacy therefore may have a dual positive role, promoting both the current employability of the parents (improving socio-economic status), as well as promoting parental involvement and consequently the educational outcomes and employability of their children. Furthermore, programs that enable non-white parents to understand how the educational system works and how best to assist their children in school would also have lasting benefits by improving the long term prospects of their children. The effect of parental involvement in reducing the effects of socio-economic disadvantage and improving educational outcomes emphasises the need for a new dimension to educational policy.

Bibliography

Aakvik, A., Vaage, K., and Salvanes, K. G. (2005). Educational Attainment and Family Background. *German Economic Review , 6* (3), 377–394.

Anderson, K. G. (2000). *Family Structure, Parental Investment, and Educational Outcomes amoung Black South Africans.* Ann Arbor, Michigan: Population Studies Centre: University of Michigan.

Anderson, K. G., Case, A., and Lam, D. (2001). Causes and Consequences of Schooling Outcomes in South Africa: Evidence from Survey Data. *Social Dynamics , 27* (1), 37 — 59.

Ardington, C., and Leibbrandt, M. (2009). *Parental Loss and Schooling: Evidence from Metropolitan Cape Town.* Cape Town: SALDRU University of Cape Town: Southern Africa Labour and Development Research Unit, Working Paper Number 42.

Beal, S. J., and Crockett, L. J. (2010). Adolescents' Occupational and Educational Aspirations and Expectations: Links to High School Activities and Adult Educational Attainment. *Developmental Psychology , 46* (1), 258–265.

Becker, G. S. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.* New York: Columbia University Press.

Becker, G. S., and Tomes, N. (1986). Human Capital and the Rise and Fall of Families. *Journal of Labor Economics , 4* (3), 1-39.

Biblarz, T. J., and Raftery, A. E. (1999). Family Structure, Educational Attainment, and Socioeconomic Success: Rethinking the "Pathology of Matriarchy". *The American Journal of Sociology , 105* (2), 321-365.

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2005). The More The Merrier? The Effect of Family Size and Birth Order on Children's Education. *The Quarterly Journal of Economics , May*, 669 - 700.

Blanden, J., and Gregg, P. (2004). Family Income and Educational Attainment: A Review of Approaches and Evidence for Britain. *Oxford Review of Economic Policy , 20* (2), 245-263.

Boissiere, M. (2004). *Determinants of Primary Education: Outcomes in Developing Countries.* Washington DC: The World Bank.

Burns, J. (2001). *Inheriting the Future – the Role of Family Background and Neighbourhood Characteristics in Child Schooling Outcomes in South Africa.* Johannesburg, South Africa: DPRU/FES Conference.

Case, A., and Deaton, A. (1999). School Inputs and Educational Outcomes in South Africa. *The Quarterly Journal of Economics , 3* (114), 1047-1084.

Case, A., and Yogo, M. (1999). *Does School Quality Matter? Returns to Education and the Characteristics of Schools in South Africa.* Cambridge, Massachusetts: National Bureau of Economic Research: Working Paper 7399.

Davis-Kean, P. E. (2005). The Influence of Parent Education and Family Income on Child Achievement: The Indirect Role of Parental Expectations and the Home Environment. *Journal of Family Psychology , 19* (2), 294–304.

Davis-Kean, P. E., and Sexton, H. R. (2009). Race Differences in Parental Influences on Child Achievement. *Merrill-Palmer Quarterly , 55* (3), 285–318.

de Haan, M. (2009). Birth Order, Family Size and Educational Attainment. *Economics of Education Review , 29* (4), 576-588.

Deary, I. J., Strand, S., Smith, P., and Fernandes, C. (2007). Intelligence and Educational Achievement. *Intelligence , 35*, 13–21.

Dee, T. S. (2004). Teachers, Race, and Student Achievement in a Randomized Experiment. *The Review of Economics and Statistics , 86* (1), 195-210.

Department of Education. (2000). *Report on the School Register of Needs 2000 Survey.* Pretoria, South Africa.

Downey, D. B. (1995). When Bigger Is Not Better: Family Size, Parental Resources, and Children's Educational Performance. *American Sociological Review , 60* (5), 746-761.

Dubow, E. F., Boxer, P., and Huesmann, L. R. (2009). Long-term Effects of Parents' Education on Children's Educational and Occupational Success. *Merrill-Palmer Quarterly , 55* (3), 224–249.

Ellen, I. G., and Turner, M. A. (1997). Does Neighbourhood Matter? Assessing Recent Evidence. *Housing Policy Debate , 8* (4), 833 - 866.

Ermisch, J., and Francesconi, M. (2001). Family Matters: Impacts of Family Background on Educational Attainments. *Economica , 68*, 137-156.

Evans, W. N., Oates, W. E., and Schwab, R. M. (1992). Measuring Peer Group Effects: A Study of Teenage Behavior. *The Journal of Political Economy , 100* (5), 966-991.

Fan, X., and Chen, M. (2001). Parental Involvement and Students' Academic Achievement: A Meta-Analysis. *Educational Psychology Review , 13* (1), 1-22.

Fiske, E., and Ladd, H. (2004). *Elusive Equity. Ediucation Reform in Post-Apartheid South Africa.* Washington DC: Brookings Institute Press.

Ginther, D. K., and Pollak, R. A. (2004). Family Structure and Children's Educational Outcomes: Blended Families, Stylized Facts, and Descriptive Regressions. *Demography , 49* (4), 671-696.

Glewwe, P., and Jacoby, H. (1994). Student Achievement and Schooling Choice in Low Income Countries: Evidence from Ghana. *The Journal of Human Resources , 29* (3), 843-864.

Hango, D. (2007). Parental Investment in Childhood and Educational Qualifcations: Can Greater Parental Involvement Mediate the Effects of Socioeconomic Disadvantage? *Social Science Research , 36*, 1371–1390.

Hanushek, E. A. (1979). Conceptual and Empirical Issues in the Estimation of Educational Production Functions. *The Journal of Human Resources , 14* (3), 351-388.

Hanushek, E. A. (1986). The Economics of Schooling: Production and Efficiency in Public Schools. *Journal of Economic Literature , 24* (3), 1141-1177.

Hanushek, E. A. (2003). The Failure of Input-Based Schooling Policies. *The Economic Journal , 113* (February), F64–F98.

Heckman, J. J., and Li, X. (2004). Selection Bias, Comparative Advantage and Heterogeneous Returns to Education: Evidence From China in 2000. *Pacific Economic Review , 9* (3), 155–171.

Hedges, L. V., Laine, R. D., and Greenwald, R. (1994). An Exchange: Part I: Does Money Matter? A Meta-Analysis of Studies of the Effects of Differential School Inputs on Student Outcomes. *Educational Researcher , 23* (3), 5-14.

Hill, N. E., and Tyson, D. F. (2009). Parental Involvement in Middle School: A Meta-Analytic Assessment of the Strategies That Promote Achievement. *Developmental Psychology , 45* (3), 740–763.

Hoover-Dempsey, K. V., and Sandler, H. M. (1995). Parental Involevement in Children's Education: Why Does it Make a Difference? *Teachers College Record , 97* (2), 310-331.

Jencks, C., and Mayer, S. (1990). The Social Consequences of Growing Up in a Poor Neighborhood. In J. L. Lynn, and M. G. McGeary, *Inner- city poverty in the United States* (pp. 111-186). Washington, DC: National Academy Press.

Jeynes, W. H. (2005). A Meta-Analysis of the Relation of Parental Involvement to Urban Elementary School Student Academic Achievement. *Urban Education , 40* (3), 237-269.

Kan, K., and Tsai, W.-D. (2005). Parenting Practices and Children’s Education Outcomes. *Economics of Education Review , 24*, 29–43.

Kodde, D. A., and Ritzen, J. M. (1988). Direct and Indirect Effects of Parental Education Level on the Demand for Higher Education. *The Journal of Human Resources , 23* (3), 356-371.

Krueger, A. B. (2003). Economic Considerations and Class Size. *The Economic Journal , 113* (February), F34-F63.

Lam, D., Ardington, C., and Leibbrandt, M. (2008). *Schooling as a Lottery: Racial Differences in School Advancement in Urban South Africa.* Ann Arbor: University of Michigan: Population Studies Center Research Report 08-632.

Lam, D., Ardington, C., Branson, N., Case, A., Leibbrandt, M., Menendez, A., et al. (2008a). *The Cape Area Panel Study: A Very Short Introduction to the Integrated Waves 1-2-3-4 Data.* The University of Cape Town.

Lam, D., Ardington, C., Branson, N., Case, A., Leibbrandt, M., Menendez, A., et al. (2008b). *The Cape Area Panel Study: Overview and Technical Documentation Waves 1-2-3-4 (2002-2006).* The University of Cape Town.

Lamdin, D. J. (1996). Evidence of Student Attendance as an Independent Variable in Education Production Functions. *The Journal of Educational Research , 89* (3), 155-162.

Lu, Y. (2009). Sibship Size and Education in South Africa: Black–White Variations. *Research in Social Stratification and Mobility , 27*, 110–125.

McNeal Jr., R. B. (2001). Differential Effects of Parental Involvement on Cognitive and Behavioral Outcomes by Socioeconomic Status. *Journal of Socio-Economics , 30*, 171–179.

Miller, P., Mulvey, C., and Martin, N. (2001). Genetic and Environmental Contributions to Educational Attainment in Australia. *Economics of Education Review , 20*, 211–224.

Nimubona, A.-D., and Vencatachellum, D. (2007). Intergenerational Education Mobility of Black and White South Africans,. *Journal of Population Economics* (20), 149-182.

Poropat, A. E. (2009). A Meta-Analysis of the Five-Factor Model of Personality and Academic Performance. *Psychological Bulletin , 135*, 322–338.

Sandefur, G. D., and Wells, T. (1999). Does Family Structure Really Influence Educational Attainment? *Social Science Research , 28*, 331–357.

Sirin, S. R. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research , 75* (3), 417-453.

Tan, E. T., and Goldberg, W. A. (2009). Parental School Involvement in Relation to Children's Grades and Adaptation to School. *Journal of Applied Developmental Psychology , 30*, 442–453.

Todd, P. E., and Wolpin, K. I. (2003). On the Specification and Estimation of the Production Function for Cognitive Achievement. *The Economic Journal , 113* (485), F3-F33.

Todd, P. E., and Wolpin, K. I. (2006). *The Production of Cognitive Achievement in Children: Home, School and Racial Test Score Gaps.* Philadelphia, PA: University of Pennsylvania.

Van der Berg, S. (2008). *How effective are poor schools? Poverty and educational outcomes in South Africa.* Göttingen: Centre for European Governance and Economic Development Reserach.

Van der Berg, S., and Burger, R. (2002). Education and Socio-Economic Differentials: A Study of School Performance in the Western Cape. *DPRU/FES Conference on Labour Markets and Poverty in South Africa.* Johanesburg, 22-23 October: University of Stellenbosch.

van Eijck, K., and de Graaf, P. M. (1995). The Effects of Family Structure on the Educational Attainment of Siblings in Hungary. *European Sociological Review , 11* (3), 273-292.

Vandenberghe, V. (1999). Economics of Education. The Need to go Beyond Human Capital Theory and Production-Function Analysis. *Educational Studies , 25* (1), 129-143.

Wilson, K. (2001). The Determinants of Educational Attainment: Modeling and Estimating the Human Capital Model and Education Production Functions. *Southern Economic Journal , 67* (3), 518-551.

Appendix A

Explanatory Variable Selection

The selection of explanatory variables was a key aspect of this study, and was done in a two phased approach. Firstly, the findings presented in the literature review were used to identify all possible measures for use in modelling from the dataset. Secondly, deductive reasoning and preliminary analysis was conducted to select the best variables for inclusion where it was evident that multiple variables captured a similar concept.

Individual Level Variables

Studies demonstrate that the most important individual level variables in the analysis of educational attainment are genetic endowments, conscientiousness and the will to achieve (commonly measured through measures of individual educational expectations) (Miller, Mulvey, and Martin, 2001; Poropat, 2009; Beal and Crockett, 2010).

The data set does not contain a measure of genetic endowments, based on various studies highlighted in section 3 of the study, past achievement and parental education was used as a proxy in the value added and contemporaneous specifications respectively. Will to achieve was measured using a series of dummy variables of individual level educational expectations. There was only one variable of individual expectations. Conscientiousness was measured through a variable "study time" measuring the effort of an individual. Various other measures of conscientiousness were considered, however these variables were not optimal for inclusion due to the minimal level of variation in the answers across the sample; as noted below.

Variable	Rationale For Exclusion
Time spent on extra lessons	79.95% of the sample spent no time on extra lessons, with 91.90% of the sample spending between 0 – 2 hours on extra lessons. The variable varies between 0 - 15.
How often did not do homework	87.57% of the sample seldom/never did not do their homework.
How often came to school late	91.75% of the sample seldom/never went to school late.

Family Level Variables

A survey of the literature illustrated that family level variables in studies of educational attainment primarily fall into three main categories; family structure, parental involvement and socio-economic status. Family structure typically includes variables that measure household size and family type. Although studies typically use the number of siblings to measure household size, this study uses a measure of household size which accounts for all those living in the household including extended family members. There are two reasons this was used, firstly it was the only measure of household size in the dataset and secondly, as highlighted in the literature review in the South African context extended families play an important role, especially amongst Africans. With respects to family type two alternate variables were available in the data set. The first was patterns of current co-residence between parent and child, which comprised a series of dummy variables indicating whether the child lived with both parents, with either parent, or with no parents. The second was a measure of the proportion of life lived in a specific family structure. As the education production function is cumulative and historical inputs matter, the proportion of life lived in a specific family structure was preferred to current co-residence patterns. The decision was made to include only the variable "proportion lived with both parents" and exclude the variables "proportion lived with mother only", "proportion lived with father only" and "proportion lived with neither parent". This was based on the findings of Downey (1995) and van Eijck and de Graaf (1995) emphasising that the key distinction is not between family types, but rather between a nuclear family comprising both parents and other family types.

Regarding parental involvement, the literature emphasised that although the personal relationship between parent and child may matter, what matters most in terms of educational outcomes is the ability of the parent to support the child in educational activities (Hill and Tyson, 2009; Hoover-Dempsey and Sandler, 1995; Fan and Chen, 2001; Tan and Goldberg, 2009). This observation led to the exclusion of the following from the analysis "how often has mother/father spent time with just you", and "how often has mother/father discussed personal matters". Four education related parental involvement measures were available in the data set, however only "mother/father helps with homework " and "parental educational expectations" were included in the analysis. "How

often mother/father talked to you about education" was excluded as it was only available for wave 3, which was conducted approximately two years after the literacy and numeracy evaluation used to derive the outcome variable. "Who was responsible for influencing school performance while growing up" was excluded due to a number of reasons. The variable has little variation in answers with over 75% of their sample stating their parents were the most influential. The variable also does not establish the mechanism of influence by specifying if the influence is through personal interaction or making decisions regarding which school will be attended. This variable simply indicates the importance of the role played by parents in determining their child's education, without highlighting how influence is established and whether it is a positive or negative influence.

This study uses 'log of per capita family income" and mother's/father's years of completed education as measures of socio-economic status as these are typically included as catch all variables that provide the socio-economic status of the family (Lam, Ardington, and Leibbrandt, 2008; Aakvik, Vaage and Salvanes, 2005; Ermisch and Francesconi, 2001). These are supplemented with the variable 'household owns a computer' to account for the level of educational resources available in the family, as this was suggested as an important component of socio-economic status in the South African context (Van der Berg, 2008). The variable "mother/father given money for school fees/tuition/books/uniform" was not included on the basis that it does not provide a quantitative measure of the money given, and consequently does not adequately account for the much higher levels of money provided to white children for educational purposes.

School Level Variables

Studies on educational attainment use proxy measures for school and teacher quality such as pupil teacher ratio, access to textbooks, classroom and school facilities and the average level of teacher education. There is no standard list of school level variables used across studies and ultimate selection depends largely on data availability (Boissiere, 2004; Lamdin, 1996).

Similarly, based on data availability, this study uses three proxy measures for school and teacher quality. These are "pupil- teacher ratio", a composite variable that indicates "school problems" (teacher absenteeism, drugs and alcohol on school grounds and poor school facilities), and a series of dummy variables for self reported school quality. Although the self reported measures of quality contain an element of subjectivity they were preferred to the alternate measure of school quality commonly used, the racially based categorisation of schools from the apartheid era. Racially based measures of quality are based on the assumption that previously African schools are of lower quality, however the large quality variations within African schools are not accounted for. Furthermore, as this analysis controls for race in the non-white regression and runs separate regressions for whites, racially based measures add little explanatory power in addition to race.

Although the literature highlights school facilities as an important aspect of school quality the variable "computer centre in school" was not included because of the high number of missing values (48.57%). Its inclusion would have substantially reduce the sample size without adding significantly to explaining variations in the outcome variable, when the other school quality proxy variables have been added.

Neighbourhood Level Variables

Neighbourhood level variables are concerned with measuring the socio-economic status of the neighbourhood (Ellen and Turner, 1997; Wilson, 2001). The two most commonly used measures of socio-economic status are typically education and income (Sirin, 2005). This study used census data on subplace education and income to cover the first to cater for these aspects of socio-economic status. In light of the high unemployment rates in South Africa, the subplace unemployment rate was also used.

Appendix B

Contemporaneous Specification Regressions

The contemporaneous regressions (parental education is used as a proxy for child genetic endowments) for the standardised LNE score, referred to in section 6 are provided below.

Table A1: Individual Level Regressions for Standardised LNE Score - Contemporaneous Specification

Variable	*Non-white*	*White*
African *(ref = Coloured)*	-0.705***	
	(0.0534)	
Age	0.0292**	0.0148
	(0.0114)	(0.0271)
Male	0.131***	0.0874
	(0.0403)	(0.0734)
Study Time	0.0127***	0.0029
	(0.00283)	(0.00353)
Educational Expectations Matric	0.151*	0.396
	(0.0802)	(0.282)
Educational Expectations Tertiary	0.479***	0.698**
	(0.0785)	(0.286)
Mother's Education Missing	0.280***	0.554
	(0.0974)	(0.382)
Mother's Education	0.0425***	0.0609***
	(0.0079)	(0.022)
Father's Education Missing	0.135	0.698**
	(0.0845)	(0.291)
Father's Education	0.0224***	0.0495**
	(0.00832)	(0.0191)
Constant	-1.267***	-1.068
	(0.211)	(0.651)
Observations	1548	225
R-squared	0.28	0.268

Robust standard errors in parentheses. *** Significant at 1% level ** Significant at 5% level
*Significant at 10% level

Table A2: Family Level Regressions for Standardised LNE Score - Contemporaneous Specification

Variable	*Non-white*	*White*
African *(ref = Coloured)*	-0.563***	
	(0.0596)	
Household Size	0.000272	0.0383
	(0.00756)	(0.0291)
Proportion lived with both Parents	0.0735	0.0434
	(0.0548)	(0.107)
Mother Helps with Homework	-0.163***	-0.0790
	(0.0548)	(0.0684)
Father Helps with Homework	-0.00137	0.0853
	(0.0614)	(0.0828)
Parental Expectations – Matric	0.208***	0.0343
	(0.0697)	(0.426)
Parental Expectations – Tertiary	0.449***	0.334
	(0.0659)	(0.417)
Household Own Computer	0.315***	0.0504
	(0.0626)	(0.104)
Log Per Capita Household Income	0.0652**	0.112**
	(0.0291)	(0.0501)
Mother's Education Missing	0.242**	0.185
	(0.105)	(0.416)
Mother's Education	0.0358***	0.0598***
	(0.00940)	(0.0197)
Father's Education Missing	0.00447	0.383*
	(0.0947)	(0.205)
Father's Education	0.00419	0.0197
	(0.00909)	(0.0162)
Constant	-0.916***	-1.120
	(0.208)	(0.682)
Observations	1530	216
R-squared	0.286	0.291

Robust standard errors in parentheses. *** Significant at 1% level ** Significant at 5% level
*Significant at 10% level

Table A3: School Level Regressions for Standardised LNE Score - Contemporaneous Specification

Variable	*Non-white*	*White*
African *(ref = Coloured)*	-0.583***	
	(0.0556)	
Pupil-Teacher Ratio	-0.0149**	-0.0404**
	(0.00637)	(0.0174)
Pupil-Teacher Ratio Missing	-0.152	-0.914**
	(0.204)	(0.420)
School Problems	-0.00550	-0.198***
	(0.0534)	(0.0707)
School Quality – Good	-0.130**	-0.0492
(ref - Excellent)	(0.0544)	(0.0632)
School Quality – Acceptable	-0.111*	-0.381***
(ref - Excellent)	(0.0596)	(0.142)
School Quality – Poor	-0.324***	-0.00203
(ref - Excellent)	(0.0868)	(0.112)
School Quality - Very Poor	-0.288**	0
(ref - Excellent)	(0.138)	(0)
Mother's Education Missing	0.256**	0.236
	(0.102)	(0.427)
Mother's Education	0.0390***	0.0541**
	(0.00839)	(0.0229)
Father's Education Missing	0.0963	0.502*
	(0.0876)	(0.280)
Father's Education	0.0230**	0.0292
	(0.00892)	(0.0217)
Constant	0.204	1.281**
	(0.230)	(0.576)
Observations	1624	231
R-squared	0.251	0.252

Robust standard errors in parentheses. *** Significant at 1% level ** Significant at 5% level
*Significant at 10% level

Table A4: Neighbourhood Level Regressions for Standardised LNE Score - Contemporaneous Specification

Variable	*Non-white*	*White*
African *(ref = Coloured)*	-0.515***	
	(0.111)	
Unemployment – Subplace	0.0980	3.689*
	(0.545)	(2.143)
Mean Education – Subplace	0.104**	-0.0791
	(0.0465)	(0.0894)
Mean Income – Subplace	0.108	0.570***
	(0.102)	(0.184)
Mother's Education Missing	0.212**	0.481
	(0.0977)	(0.399)
Mother's Education	0.0365***	0.0564**
	(0.00852)	(0.0230)
Father's Education Missing	0.0549	0.339
	(0.0873)	(0.318)
Father's Education	0.0177**	0.0322
	(0.00873)	(0.0231)
Constant	-2.369***	-5.892***
	(0.897)	(1.542)
Observations	1656	234
R-squared	0.241	0.192

Robust standard errors in parentheses. *** Significant at 1% level ** Significant at 5% level
*Significant at 10% level

Table A5: Education Production Function Regressions for Standardised LNE Score – Contemporaneous Specification

Variable	*Non-white*	*White*
African *(ref = Coloured)*	-0.616***	
	(0.0658)	
Age	0.00961	0.0526
	(0.0118)	(0.0376)
Male	0.156***	0.0510
	(0.0386)	(0.0696)
Study Time	0.00885***	0.00126
	(0.00275)	(0.00376)
Educational Expectation – Matric	0.107	0.336
	(0.0816)	(0.308)
Educational Expectation – Tertiary	0.339***	0.590*
	(0.0783)	(0.327)
Household size	0.00257	0.0345
	(0.00813)	(0.0322)
Proportion lived with both Parents	0.0568	0.00385
	(0.0526)	(0.106)
Mother Helps with Homework	-0.144**	-0.00682
	(0.0558)	(0.0704)
Father Helps with Homework	-0.0110	0.0757
	(0.0674)	(0.0855)
Parental Expectations – Matric	0.110	-0.0671
	(0.0770)	(0.284)
Parental Expectations – Tertiary	0.241***	-0.00102
	(0.0721)	(0.260)
Household Own Computer	0.238***	-0.0392
	(0.0619)	(0.114)
Log Per Capita Household Income	0.0397	0.119**
	(0.0292)	(0.0530)
Mother's Education Missing	0.183*	-0.135
	(0.103)	(0.439)
Mother's Education	0.0277***	0.0490**
	(0.00936)	(0.0215)
Father's Education Missing	0.0163	0.505*
	(0.0900)	(0.284)
Father's Education	0.00360	0.0144
	(0.00852)	(0.0191)
Pupil-Teacher Ratio	-0.0154**	-0.0446**
	(0.00610)	(0.0187)
Pupil-Teacher Ratio Missing	-0.228	-1.139**
	(0.195)	(0.476)
School Problems	0.0437	-0.0999
	(0.0552)	(0.0646)
School Quality – Good	-0.0201	-0.0468
(ref - Excellent)	(0.0575)	(0.0661)
School Quality – Acceptable	-0.0418	-0.314**
(ref - Excellent)	(0.0616)	(0.150)
School Quality – Poor	-0.159*	-0.00603
(ref - Excellent)	(0.0896)	(0.248)
School Quality - Very Poor	-0.114	0
(ref - Excellent)	(0.151)	(0)
Constant	-0.656*	-0.787
	(0.352)	(1.134)
Observations	1432	210
R-squared	0.338	0.395

Robust standard errors in parentheses. *** Significant at 1% level ** Significant at 5% level
*Significant at 10% level

Appendix C

In order to investigate the level of correlation between neighbourhood variables and race three separate regressions were run; each includes a different neighbourhood level variable as the outcome variable and race variables as the independent variables. These regressions are presented below.

Table A6: Regressions of Neighbourhood Variables and Race

Variable	***(1)*** *Unemployment (subplace)*	***(2)*** *Mean Education (subplace)*	***(3)*** *Mean Income (subplace)*
African	0.275***	-3.389***	-2.071***
	(0.00705)	(0.122)	-0.0615
Coloured	0.0851***	-3.184***	-1.179***
	(0.00486)	(0.13)	-0.056
Constant	0.0244***	12.43***	12.10***
	(0.00184)	(0.0887)	-0.0376
Observations	1890	1890	1890
R-squared	0.767	0.627	0.679
Robust standard errors in parentheses. *** Significant at 1% level ** Significant at 5% level *Significant at 10% level			

Appendix D

In lines with studies outlined in the literature review it is noted that many explanatory variables used in the education production function are correlated. For this reason a sound theoretical basis was established to guide the selection of explanatory variables, and interpret results in this study. The correlation matrices below illustrate the magnitude of some noted correlations between explanatory variables, separately for non –whites and whites.

Table A7: Quantitative Explanatory Variables Correlation Matrix for Non-whites

	Age	*Study Time*	*Household Size*	*Proportion lived with both Parents*	*Log Per Capita Household Income*	*Mother's Education*	*Father's Education*	*Pupil-Teacher Ratio*	*Unemploy-ment (Subplace)*	*Mean Education (Subplace)*	*Log Mean Income (Subplace)*
Age	1.00										
Study Time	-0.0494* (0.0453)	1.00									
Household size	-0.0684* (0.0054)	-0.0516* (0.0363)	1.00								
Proportion lived with both Parents	-0.0986* (0.0001)	0.0609* (0.0136)	-0.009 (0.7158)	1.00							
Log Per Capita Household Income	-0.1369* (0.00)	0.1178* (0.00)	-0.2593* (0.00)	0.1962* (0.00)	1.00						
Mother's Education	-0.1815* (0.00)	0.0937* (0.0001)	-0.1372* (0.00)	0.099* (0.0001)	0.2981* (0.00)	1.00					

	Age	*Study Time*	*Household Size*	*Proportion lived with both Parents*	*Log Per Capita Household Income*	*Mother's Education*	*Father's Education*	*Pupil-Teacher Ratio*	*Unemploy-ment (Subplace)*	*Mean Education (Subplace)*	*Log Mean Income (Subplace)*
Father's Education	-0.1103* (0.00)	0.1001* (0.00)	-0.0818* (0.0009)	0.4291* (0.00)	0.331* (0.00)	0.3224* (0.00)	1.00				
Pupil-Teacher Ratio	-0.1326* (0.00)	-0.1126* (0.00)	0.0738* (0.0027)	-0.0529* (0.0316)	-0.1687* (0.00)	-0.172* (0.00)	-0.1158* (0.00)	1.00			
Unemploy-ment (Subplace)	0.2903* (0.00)	-0.1055* (0.00)	0.0539* (0.0283)	-0.2168* (0.00)	-0.5295* (0.00)	-0.1936* (0.00)	-0.24* (0.00)	0.0412 (0.0941)	1.00		
Mean Education (Subplace)	-0.0466 (0.0582)	0.1076* (0.00)	-0.0787* (0.0013)	0.0878* (0.0004)	0.3402* (0.00)	0.2461* (0.00)	0.2013* (0.00)	-0.1659* (0.00)	-0.4453* (0.00)	1.00	
Log Mean Income (Subplace)	-0.2561* (0.00)	0.1218* (0.00)	-0.0643* (0.0088)	0.1931* (0.00)	0.5406* (0.00)	0.243* (0.00)	0.2496* (0.00)	-0.096* 0.0001	-0.877* (0.00)	0.6934* (0.00)	1.00

Matrix indicates correlations between pairs of quantitative variables. p - values in parentheses. * Significant at 5% level

Table A8: Quantitative Explanatory Variables Correlation Matrix for Whites

	Age	*Study Time*	*Household Size*	*Proportion lived with both Parents*	*Log Per Capita Household Income*	*Mother's Education*	*Father's Education*	*Pupil-Teacher Ratio*	*Unemploy-ment (Subplace)*	*Mean Education (Subplace)*	*Log Mean Income (Subplace)*
Age	1.00										
Study Time	0.1215 (0.0646)	1.00									
Household size	-0.1764* (0.0068)	-0.0259 (0.6948)	1.00								
Proportion lived with both Parents	-0.0849 (0.1966)	0.0516 (0.4349)	-0.0264 (0.6882)	1.00							
Log Per Capita Household Income	-0.0308 (0.6388)	-0.0626 (0.3427)	-0.0684 (0.2976)	0.1363* (0.0376)	1.00						
Mother's Education	0.0043 (0.9472)	-0.0230 (0.7279)	-0.0954 (0.1456)	0.2995* (0.00)	0.1555* (0.0173)	1.00					
Father's Education	0.0312 (0.6347)	0.0086 (0.8960)	-0.1724* (0.0082)	0.4848* (0.00)	0.3631* (0.00)	0.5551* (0.00)	1.00				
Pupil-Teacher Ratio	-0.4878* (0.00)	-0.0100 (0.8799)	0.1520* (0.0200)	0.0415 (0.5281)	-0.1703* (0.0091)	-0.1272 (0.0519)	-0.1011 (0.1231)	1.00			

	Age	*Study Time*	*Household Size*	*Proportion lived with both Parents*	*Log Per Capita Household Income*	*Mother's Education*	*Father's Education*	*Pupil-Teacher Ratio*	*Unemploy-ment (Subplace)*	*Mean Education (Subplace)*	*Log Mean Income (Subplace)*
Unemploy-ment (Subplace)	0.0436	0.0197	-0.1040	-0.1849*	-0.2978*	-0.1498*	-0.1610*	0.1412*	1.00		
	(0.5069)	(0.7658)	(0.1125)	(0.0046)	(0.00)	(0.0219)	(0.0137)	(0.0309)			
Mean Education (Subplace)	0.0313	-0.0189	0.0551	0.1693*	0.5104*	0.2641*	0.3681*	-0.2775*	-0.5407*	1.00	
	(0.6336)	(0.7750)	(0.4013)	(0.0096)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)		
Log Mean Income (Subplace)	0.0391	-0.0177	0.0658	0.2389*	0.4739*	0.2577*	0.3406*	-0.2619	-0.5038*	0.8552*	1.00
	(0.5515)	(0.7890)	(0.3161)	(0.0002)	(0.00)	(0.0001)	(0.00)	(0.0001)	(0.00)	(0.00)	
Matrix indicates correlations between pairs of quantitative variables. p - values in parentheses. * Significant at 5% level											