The impact of teacher characteristics on student performance: An analysis using hierarchical linear modelling

PAULA ARMSTRONG

Stellenbosch Economic Working Papers: 25/14

KEYWORDS: EDUCATION; TEACHERS, HIERARCHICAL LINEAR MODELLING
JEL: I2; I21

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

UNIVERSITEIT
STELLENBOSCH
UNIVERSITY

BER
BUREAU FOR ECONOMIC RESEARCH

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

The impact of teacher characteristics on student performance: An analysis using hierarchical linear modelling

Abstract

This paper makes use of hierarchical linear modelling to investigate which teacher characteristics impact significantly on student performance. Using data from the SACMEQIII study of 2007, an interesting and potentially important finding is that younger teachers are better able to improve the mean mathematics performance of their students. Furthermore, younger teachers themselves performance better on subject tests than do their older counterparts. Changes in teacher education in the late 1990s and early 2000s may explain the differences in the performance of younger teachers relative to their older counterparts. However, further investigation is required to fully understand these differences.

Keywords: Education; teachers, hierarchical linear modelling

JEL codes: I2; I21

1. Introduction

The impact of teacher characteristics (both qualifications and demographic characteristics) is important for education policy. Ensuring that teachers best suited and most able to enhance student performance are employed is a key responsibility for policymakers. Wayne and Youngs (2003: 89) explain that a large body of literature about teacher characteristics and education outcomes exists. The focus on the studies vary between questions about teacher quantity and turnover and issues surrounding teacher quality. In many countries (South Africa included) certain qualifications need to be obtained before teachers are permitted to enter the teaching force. Much of the literature surrounding teacher characteristics and student performance is comprised of analyses of the impact of these and other qualifications. Attempts have been made to identify trends in the quality of teachers, and the question whether characteristics of teachers in different parts of the schooling system exist is often investigated (Wayne & Youngs, 2003: 90).

The relationship between teacher characteristics and student performance is surprisingly elusive, however. Researchers have found it difficult to find aspects of teacher training that correlate with student performance in a statistically significant way (Chingos & Peterson, 2011: 449). Conflicting or indeterminate results occur often. Summers and Wolfe (1977) investigated the impact of teacher scores on “Philadelphia’s National Teacher Evaluations” on performance amongst primary schools students in that state, finding a negative relationship between teacher performance and student scores on standardised tests. Anderson (2000) investigates the determinants of student performance in mathematics and language in Mexico and finds a positive and statistically significant impact in both mathematics and language for teachers making use of a more interactive approach to teaching as opposed to a traditional approach in which lessons are dominated by teachers talking and instructing (Anderson, 2000: 144). She also finds evidence of a positive relationship between hours spent teaching and performance in both subjects[1] (Anderson, 2000: 145). Teacher effort variables therefore impact positively and statistically significantly on student performance. An interesting and important result is the positive and significant impact on both language and mathematics observed for teacher training during the year in which the study was conducted (Anderson, 2000: 146). Angrist and Lavy (2001) find positive estimates of the impact of in-service teacher training on both mathematics and language in secular primary schools in Jerusalem. They report that their results are robust to a number of estimation techniques, namely regression, difference-in-difference techniques as well as matching techniques. The fact that the effect is only observed in secular schools may be due to the fact that the training programme was introduced later and on a smaller scale in religious schools (Angrist & Lavy, 2001: 365).

Ferguson (1998) used data from the “Texas Examination of Current Administrators and Teachers” to evaluate the impact of student performance at all levels of the schooling system. Contrary to the results obtained by Summers and Wolfe, Ferguson found a positive correlation between student performance and teacher test scores.[2] The relationship between teacher performance on tests in the subject they teach and student performance in that subject has also been tested extensively. Positive associations between teacher test score and student

[1] Anderson notes that this variable is self-reported (Anderson, 2000: 145) and may well be over-reported. However, if this is the case, it likely that the coefficient on this variables is a lower bound of the effect of time on task of student performance.

[2] Important to note is that Ferguson’s study aggregated data to the district level. Hanushek, Rivkin and Taylor (1996: 616) explain that aggregating data to a “higher” level (i.e. school, district or state level) increases the likelihood of obtaining positive results.

performance are observed in some studies across a range of subjects (Ehrenberg & Brewer, 1995; Hanushek, 1992; Rowan, Chiang & Miller, 1997), while others find a negative impact of teacher test scores on student outcomes (Murnane & Phillips, 1981). It seems then that the evidence regarding the impact of teacher content knowledge on student outcomes is mixed. Results obtained for formal teacher qualifications were also mixed, with the majority of studies conducted returning indeterminate results. Amongst those that did return results, both negative and positive impacts were observed (Wayne & Youngs, 2003: 101-103). The existing research therefore leaves us with few answers to questions about the relationship between teacher qualifications and student performance. Indeed, are teacher qualifications important at all?

Evidence from Pakistan suggests that teacher qualifications are indeed important for student performance. Arif and Saqib (2003) control for the individual and family characteristics of students, the characteristics of the schools they attend, geographic characteristics as well as a range of teacher characteristics and find that whether a teacher has a bachelor's degree or higher is positively and statistically significantly associated with student performance in language, mathematics and general knowledge as well as a measure capturing performance in all three (Arif & Saqib, 2003: 19-20). An earlier study conducted in Pakistan (Behrman, Kahn, Ross & Sabot, 1997) construct teacher quality indices for language and mathematics. These indices are linear functions of teacher performance on literacy or numeracy tests, educational attainment, and teaching experience and its squared term (Behrman et al., 1997: 131). Controlling for student demographic characteristics and family background, school characteristics, student-teacher ratios and student ability, they find a positive and statistically significant relationship between the teacher quality index and student performance in both numeracy and literacy (although the effect seems to be larger in literacy – an interesting result, since an effect, if observed at all, is usually stronger in the case of mathematics) (Behrman et al., 1997: 133).

Another study that finds a relationship between observable teacher characteristics and student performance was conducted by Slater, Davies and Burgess (2009) using UK data for 7 000 students (14 year olds) writing GCSE Keystage 4 examinations.[3] Slater et al. (2009)

[3] Keystage 4 examinations are compulsory examinations dictating entrance to post-secondary education. These are written at age 16. Keystage 3 examinations are written at the beginning of Keystage 4 programme during the year that students turn 14 (Slater, Davies & Burgess, 2009: 4). Keystage 3 examinations are often used as a "pre-test" measure in education research, or an as indication of prior attainment.

investigate whether the observable characteristics of teachers are correlated with measures of teacher effectiveness. Teacher effectiveness is measured as the effect that teachers have on student performance on the examinations. The observable characteristics available are teacher gender, age, educational attainment and teaching experience. None of these characteristics are statistically significant in explaining teacher effectiveness (Slater et al., 2009: 12). Interesting to note, however, is that Slater et al. (2009: 13) find a correlation (albeit weak) between the ability of students and teacher effectiveness, suggesting non-random allocation of students within a school. Allocating students to teachers in such a way that places less able student with more effective teachers may well enhance the positive impact of teacher effectiveness.

Raudenbush, Eamsukkawat, Di-Ibor, Kamali and Taoklam (1993) investigate whether in-service training affect student performance significantly. They measure in-service training by including a variable capturing the amount of exposure (in terms of days) of in-service training as well as a variable controlling for the number of times that teachers received internal supervision (Raudenbush et al., 1993: 286). They also include a measure of whether a teacher has a bachelor's degree. They come up with a very interesting result: although in-service training does not appear to have any significant effect on student performance, internal supervision (by the school principal or another teacher at the school)[4] has a large and significant effect. They explain the effect of intensive internal supervision as being as large as a teacher obtaining a bachelor's degree (Raudenbush et al., 1993: 294). It appears then that although formal in-service training does not appear to improve teacher quality, a type of mentoring and "coaching" approach does. Results from a study conducted using Cambodian data (Marshall, Chinna, Nessay, Hok, Savoeun, Tinon & Vaesna, 2009: 406) show positive and significant effects (as well as inequality reducing effects) on the performance of grade 6 students on language tests. High levels of mathematical content knowledge amongst teachers also showed a positive and significant effect on grade 6 mathematics performance and high levels of mathematics pedagogical content knowledge had a significant impact on grade 3 mathematics performance (Marshall et al., 2009: 406). The authors did not control for formal teacher qualifications or teaching experience separate to content knowledge. Luschei and Carnoy (2010: 175) find no significant impact for teachers' postgraduate education on student performance in mathematics or language in a study conducted using Uruguayan data.

[4] This is in contrast to external supervision by a district official (Raudenbush et al., 1993: 294) which shows no significant impact on student performance.

Interestingly, however, high levels of teaching experience (10 years and above) are positively and significantly associated with both mathematics and language performance (Luschei & Carnoy, 2009: 175-176).

Another study that finds a statistically significant relationship between teaching experience and student performance is that of Clotfleter, Ladd and Vigdor (2007). These authors use North Carolina data to investigate the relationship between teacher characteristics and student performance. Since the early 1990s, the state of North Carolina has administered standardised mathematics and reading tests to all students between grades 3 and 8 (Clotfelter et al., 2007: 675). Furthermore, it is possible to match students to their teachers for each year. The authors are able to identify the teachers of at least 75% of grade 3, 4 and 5 students in the state's education system between 1993/1994 and 2003/2004, rendering it possible for them to conduct analysis on the impact of teacher characteristics on both the levels of mathematics and English performance and the gains in performance from year to year (and therefore controlling for various student and school-level effects, the gains that may be tentatively associated with the teacher) (Clotfelter et al., 2007: 675). The authors find a positive and statistically significant impact for teacher experience on student performance in both mathematics and English (Clotfelter et al., 2007: 676).[5] The size of the coefficients indicate that the majority (or more than half) of the returns to teaching experience occur within the first two years of teaching. An issue often raised when investigating returns to teaching experience is the possibility that positive returns to experience are overstated if it is likely that underperforming or weaker teachers will leave the profession after their initial year (Rockoff, 2004: 248). The authors test for this possibility by adding a variable controlling for whether a teacher remained in the profession in North Carolina for at least three years. They interact it with the categorical variables controlling teachers with 1 to 2 years of teaching experience. If weaker teachers leave the profession after their early years as teachers, a positive coefficient on the variable controlling for those who remain in the profession is expected. However, the opposite is observed. In the case of mathematics, a negative and statistically significant coefficient is observed in both the levels and gains model, suggesting that those who leave teaching are not less able than their counterparts who remain in the profession. Furthermore, the interaction term is not statistically significant in either subject,

[5] Teacher experience is captured by categorical variables denoting 1 to 2 years of experience, 3 to 5 years of experience, 6 to 12 years of experience, 13 to 20 years of experience, 21 to 27 years of experience and more than 27 years of experience. They therefore control for non-linear returns to teaching experience (Clotfelter et al., 2007: 676). The returns observed are higher for mathematics than they are for English – a finding largely in line with what is found in the literature about teaching experience and student performance.

suggesting that it is not differential attrition that drives the increasing returns to teaching experience observed in the data (Clotfelter et al., 2007: 676).

"By many accounts, the quality of teachers is the key element to improving student performance" (Hanushek, 2009: 171). The impact of being taught by a good teacher is quantified by Hanushek (2011: 42) where he estimates that students who perform a standard deviation above average (as measured by performance on high school tests) earn between 10 and 15 percent more per annum than average – an estimate he deems conservative as it is measured in the early years of their career (before they have reached their full earning potential) and it does not account for the possibility that higher performance at high school level probably results in higher educational attainment (Hanushek, 2011: 42). The home background and motivation of the student obviously contribute significantly to the level of success that students are able to achieve, but rigorous research has isolated the impact of effective teaching on student performance. Hanushek (2011: 42) reports that studies have consistently shown that high-performing teachers (performing 1 standard deviation above the mean, or at the 84th percentile of the distribution) result in student grades that are at least 0.2 standard deviations higher at the end of a school year. Although these gains diminish over time, it is estimated (although somewhat less conclusively) that the long term benefit of being taught by an effective teacher is 70 percent of the immediate gain, and so consecutive years of high quality teachers result in student outcomes markedly higher than they would have been had students been taught by teachers at the 50th percentile of the distribution (Hanushek, 2011: 42). It is clear then teacher quality and teacher effectiveness have a considerable effect on the lifetime earnings of students.

Evidence of the impact of teacher quality in later life also exists. Chetty, Friedman and Rockoff (2011) find evidence of fairly sizeable impacts of teacher quality on adult earnings of their students. Teacher quality (measured by value added) improves the probability of college attendance, the quality of college attended by students (measured by the earnings of former students of colleges) as well as future earnings of students (Chetty et al., 2011: 2).

How then should we measure teacher quality? To what extent are we "missing the point?" An important aspect of teacher quality and teacher effectiveness to consider is the extent to which the education received by teachers is well-suited to enabling them to teach. A significant literature (some of which is discussed above) exists around whether teaching is an attractive profession to highly able individuals endowed with skills that fetch a high price in the labour market. It is important to understand whether or not those skills are likely to translate into positive outcomes for students or whether there is "something else" required of teachers that does not necessarily guarantee that highly able individuals will be effective teachers. One way to approach this question is to investigate the specific knowledge requirements of teachers.

The National Council of Teachers of Mathematics in US (NCTM) refers to teacher's knowledge of their students as students as being central to their ability to influence their performance (NCTM, 2000: 17). This broadly refers to teachers being able to identify "preconceptions and background knowledge that students typically bring to each subject" (National Board for Professional Teaching Standards (NBPTS), 2012: vi). This is essentially what is referred to as Pedagogical Content Knowledge (PCK) (Hill, Loewenberg Ball & Schilling, 2008: 373). Although its importance in improving student outcomes is widely acknowledged, very little exists in the way of empirical evidence and understanding of this relationship. Hill et al. (2008: 373) believe that this results from two factors. Firstly, there is an absence of studies that are able to prove that teachers possess such knowledge, and secondly, measures to assess programmes which aim to develop this knowledge and its impact on student achievement have not yet been developed. In the absence of such measures, it may be difficult to measure the aspect of teacher quality that truly affects student performance.

Research that does investigate the type and depth of subject (and other) knowledge required to teach presents some very important results. The mathematical knowledge required of mathematics teachers is extensive (Ball, Thames & Phelps, 2008: 399). The tasks involved in teaching mathematics require "significant mathematical knowledge, skill, habits of mind and insight" (Ball et al., 2008: 399). What is referred to as *common content knowledge* is the mathematical knowledge that teachers require to perform their job. Teachers also require *specialised content knowledge* – mathematical knowledge and skills particular to teaching. This type of mathematical knowledge is not particularly useful (or even desirable) outside the context of teaching and requires a certain "unpacking" of mathematical knowledge. Examples of this kind of mathematical content knowledge would be the analysis of student errors or evaluating whether a nonstandard approach to calculation would work in general (Ball et al., 2008: 400). A third domain, *knowledge of content and students*, involves understanding and therefore anticipating how students will interpret and understand the work and where they will experience difficulty (Ball et al., 2008: 401). The fourth domain, *knowledge of* *mathematics and teaching*, refers to an understanding of how mathematics should be taught. For example, the sequencing of topics and examples would fall under this category of mathematical knowledge (Ball et al., 2008: 401). The authors point out that the mathematical knowledge required of teachers (and indeed teachers across different fields and subjects) includes and extends beyond that of other professions requiring mathematical knowledge. This is important to acknowledge this when evaluating the importance of the profession in society.

A rare study in which the impact of different kinds of mathematics knowledge amongst teachers (based to a large extent on the findings of Ball et al. discussed above) was tested amongst students attending schools in rural Guatemala (Marshall & Sorto, 2012) presented encouraging results. Using hierarchical linear modelling, they test the impact of different kinds of teacher knowledge in different areas of mathematics performance. Interestingly, they find coefficients of very similar size to those observed in US studies. Marshall and Sorto (2012: 188) find significant results for what they call "mathematics knowledge for teaching" (as opposed to common content knowledge and specialised content knowledge). Interestingly and importantly, the coefficients for mathematics knowledge for teaching are largest and most significant for areas of the mathematics test that have the highest degree of cognitive demand required of students (Marshall & Sorto, 2012: 191). This makes intuitive sense – the more difficult the content, the more specialised a teacher needs to be to ensure that student learning takes place.

This chapter aims to investigate which characteristics of South African teachers, both demographic and in terms of qualifications and teaching experience, impact on student performance. The chapter is organised as follows: section 2 defines the research question, introduces the dataset that will be used in the analysis, SACMEQ III, and provides the descriptive statistics of the variables that will be included in the model. Section 3 discusses the necessity for hierarchical linear modelling, while section 4 presents the model that will be specified in attempting to answer the research question. Section 5 presents the results obtained from the model, and section 6 concludes with a discussion of the possible driving factors behind these results.

2. Research question and data

2.1 Defining the research question

As indicated, this research aims to answer the question of whether teacher characteristics (both demographic and human capital) impact student performance. As explained, South Africa's educational performance is weak. The question we attempt to answer in this chapter is whether this weak performance can be explained by observable teacher characteristics. In order to measure the impact of these characteristics, the fact that students share "teacher characteristics" with the students in the same class means that the assumptions that would render ordinary least squares (OLS) regression coefficients accurate (i.e. that students are drawn from a random sample) are violated. The multi-level nature of the data requires that this element be controlled for and modelled in the investigation. This is discussed at length in section 3. In summary, the confidence intervals that would result from OLS would be deceptively narrow as a result of inaccurately small standard errors (Arnold, 1992: 62). Students being taught by the same teacher not only share "teacher characteristics", but are also more likely to be more similar to one another than to students taught by different teachers. This further violates the assumption of students being drawn at random (Arnold, 1992: 62).

The following subsection explains the data used to conduct the analysis – the third study conducted, in 2007, by the Southern and Eastern African Consortium for Monitoring Educational Quality (SACMEQ III).

2.2 Data: SACMEQ III

The paper makes use of data collected by the third study conducted by SACMEQ in 2007. SACMEQ was launched in 1995 with the objective of conducting research and providing training that enables policy makers to monitor and improve their education systems (Moloi and Strauss, 2005: 12). SACEMQ undertook 3 major surveys (referred to as SACMEQ I, II and III) in 1995, 1998 and 2007 respectively. 15 countries participated in SACMEQ III, namely Botswana, Kenya, Lesotho, Malawi, Mauritius, Mozambique, Namibia, Seychelles, South Africa, Tanzania (Mainland and Zanzibar), Uganda, Zambia and Zimbabwe (Spaull, 2011: 4).

SACMEQ III involved administering 3 tests to grade 6 students - a reading test, a mathematics test and a health test (aimed largely at measuring the level of knowledge about HIV/AIDS). In South Africa, 9 038 grade 6 students in 392 schools were tested, along with 498 mathematics teachers, 498 reading teachers and 492 health teachers (totalling 1 488). All the teachers completed a health test, and reading and mathematics teachers completed a test in the subject that they taught (Spaull, 2011: 5).

The data obtained from SACMEQ III comprise the most extensive nationally representative sample available for the South African education system.[6] Importantly, the testing was only conducted in English and Afrikaans. It is therefore highly likely (if not certain) that a significant proportion of the students writing the tests were disadvantaged in terms of understanding the mathematics questions, given that neither English nor Afrikaans was their first language. The extent to which English is spoken outside of school is controlled for at the student level but the dataset did not contain the corresponding variable for Afrikaans. It is worth noting, however, that the aforementioned language disadvantage applies to the majority of students tested in South Africa (Moloi and Strauss, 2005: 67).

Importantly, in any analysis of performance in education making use of cross-sectional data that does not contain a pre-test score, unobservable characteristics of students (such as motivation or intelligence) which influence their performance on mathematics tests are therefore not controlled for. It is also important to bear in mind that the impact of teachers on students' education is cumulative. The results observed in grade 6 therefore reflect the impact of teachers throughout students' educational "career" and cannot be attributed only to the teachers by whom students are taught in that year. Having said that, we do not have a pre-test score and we are therefore not able to control for students' ability or level of performance before their exposure to their current teacher.

2.3 **Variables included in the model**

Table 5 below provides a brief explanation of the variables included in the investigation as well as the means and standard deviations. The dependent variable, $ZMAT_{ij}$, is the z-scored

[6] Mullens, Murnane and Willett (1996: 140) explain the need for longitudinal data in assessing the impact of teachers on student learning. In the majority of studies investigating this topic in the developing world, longitudinal data are not available and so researchers have no choice but to use cross-sectional data. Cross-sectional data can only tell us about the level of student achievement and not about the progress that takes place (i.e. the actual learning). However, data on changes in achievement are necessary to truly evaluate the effectiveness of teachers (Mullens et al., 1996: 140).

(standardised) mathematics score of student *i* in classroom *j*. Z-scoring the dependent variable centres the variable around a mean value of 0 and gives the variable a standard deviation of 1. The interpretation of coefficients on independent variables for z-scored dependent variables is the standard deviation change in students' mathematics performance.

TABLE 5: Description and descriptive statistics for variables included in the model

Variable	Mean and standard deviation	Standard deviation
STUDENT LEVEL VARIABLES		
Continuous variables:		
Mathematics score (z-scored; standardised to the mean within the South African dataset)	0.00	1.00
SES (z-scored; standardised to the mean within the South African dataset)	0.00	1.00
Dummy variables (takes a value of 1 if true; takes a value of 0 if not true)		
Overage (born earlier than 1994)	0.19	0.39
Female (reference value: 0)	0.51	0.50
Mother has completed matric	0.51	0.50
Attended less than 1 year of preschool	0.05	0.21
Attended 1 year of preschool	0.33	0.47
Attended 2 years of preschool	0.15	0.36
Attended 3 or more years of preschool	0.2	0.40
Speaks English at home sometimes	0.61	0.42
Speaks English at home most of the time	0.08	0.49
Speaks English at home always	0.07	0.26
Repeated a grade once	0.20	0.40
Repeated a grade twice		

	0.05	0.22
Repeated a grade three times	0.03	0.17
Repeated grade 6	0.09	0.29
Receives extra tuition	0.09	0.29
TEACHER LEVEL VARIABLES		
Continuous variables:		
Days of in-service training	13.04	46.04
Average class size (of the school)	40.79	12.6
Teacher maths score (z-scored; mean of 0 and standard deviation of 1)	0.00	1.00
Average classroom SES (z-scored; standardised to the mean within the South African dataset)	0.18	0.80
Dummy variables:		
30 to 39 years of age	0.39	0.49
40 to 49 years of age	0.44	0.50
50 to 59 years of age	0.14	0.34
60 years and older	0.01	0.09
School is in a rural area	0.38	0.49
Private school	0.05	0.22
Trained to teach mathematics	0.67	0.47
Parents sign students' homework	0.59	0.49
Test 2 to 3 times per term	0.52	0.50
Tests 2 to 3 times per month	0.22	0.42
Tests at least once per week	0.15	0.36
Completed junior secondary education	0.02	0.15
Completed senior secondary education	0.09	0.29

Completed A-levels[7]	0.16	0.37
Completed a degree	0.51	0.50
Received less than 1 year of teacher training	0.01	0.08
Received 1 year of teacher training	0.02	0.15
Received 2 years of teacher training	0.07	0.25
Received 3 years of teacher training	0.34	0.47
Received more than 3 years of teacher training	0.56	0.50
Experience: 6 to 10 years	0.11	0.31
Experience: 11 to 15 years	0.25	0.44
Experience: 16 to 20 years	0.18	0.39
Experience: 21 to 25 years	0.13	0.34
Experience: 26 to 30 years	0.05	0.22
Experience: 31 to 35 years	0.03	0.18
Experience: 36 to 40 years	0.01	0.09
Experience: 41 plus years	0.00	0.04

Source: SACMEQ III (SACMEQ, 2007).

3. Hierarchical linear modelling: The necessity of the method

Social science contains countless examples of hierarchical data structures. This means that although variables capture characteristics of individuals, these individuals also exist within larger groups and a set of variables describe the groups (Raudenbush & Bryk, 2002: xix). A classic example of hierarchical data structure is education data. Students are grouped according to the schools they attend, so individual or learner-level variables describe individual students, and school-level variables describe schools. Although school-level variables may be independent of the students (for example, the type of buildings or the

[7] A-levels is not available in the South African education system. It is likely that teachers misunderstood the question and equated A-levels with having completed matric. The variable is retained for the sake of completeness since 16% of teachers reported having completed A-levels.

geographical location of the school), school-level variables may also represent aggregated learner-level data (for example, the racial or gender composition of the school or the average socioeconomic status of the students attending the school). The school probably consists of smaller groups such as classrooms, which have their own characteristics captured by classroom-level variables. Schools may also form the smaller groups contained in school districts (Raudenbush & Bryk, 2002: xix).

In this chapter we are interested in understanding how teacher characteristics influence student performance. As described above, students are grouped within classrooms which in turn are grouped within schools. In education the context in which students are educated is immensely influential in determining their performance. In other words, characteristics of the school classroom significantly influence the level of learning that takes place for individual students and therefore their performance on standardised tests (Luke, 2004: 1). Relationships and occurrences at the higher level of analysis affect what happens at the lower level of analysis. In South Africa the context in which learning takes place differs dramatically across the school system and so the variables describing characteristics at the classroom and school level reflect large differences between schools within the country. We are interested in how these differences at the higher level impact on lower level performance (Luke, 2004: 4-5). For example, how do differences in school management characteristics translate into differences in the performance of students on standardised mathematics and language tests? How does teacher training impact on student performance in mathematics and language tests?

The strongest motivation for making use of hierarchical linear modelling has to do with inaccuracies in the measurement of standard errors. If multi-level data are analysed solely at the level of the individual, two problems arise. The first of these is that the individual error term contains all the contextual information that has not been modelled (Duncan, Jones & Moon, 1998: 98). One of the basic assumptions of multiple regression is that there is no correlation between the error terms of individual observations – an assumption which is violated if individuals (students) share the same context (classroom or school) and the characteristics of this context are not modelled (Luke, 2004: 7). Students who attend the same school or who are taught in the same classroom will probably be more similar to one another than if they were selected at random. Secondly, if the context in which individuals find themselves is not explicitly acknowledged and modelled, regression coefficients is assumed to be equally relevant for all contexts (Duncan et al., 1998: 98). This would indicate that variables affect one another in the same way in all schools in the South African education system, for example – a notion that we know to be false.

How then does estimation in HLM differ from that in OLS? Furthermore, do the estimates and standard errors obtained using OLS and HLM differ substantially enough to warrant the use of HLM over OLS? It may be argued that making use of fixed effects in OLS circumvents the need for HLM. Chaplin (2011: 7) explains that fixed effects models in OLS are models in which the covariance between the error term and some of the explanatory variables is not constrained to be 0. Fixed effects may then control for the effects of characteristics not captured by explanatory variables (i.e. unobserved effects). Using fixed effects in OLS modelling would therefore allow the researcher to claim that the relationship of interest was not biased by unobservable characteristics in the data. For example, in order to observe the relationship between SES and student performance at an individual level and to be sure that the result obtained was not biased by unobservable characteristics at the level of the school, a fixed effects model would include individual school dummies to control for the impact of unobservable characteristics at the level of the school. Fixed effects do not, however, control for the strong possibility that students within a particular school are more similar to one another than students who have been randomly selected.

HLM is therefore often suggested as a safeguard against school effects biasing results obtained for individual level effects. However, this is only the case if multilevel models are actually modelled in the way in which they are presented: as multi-stage models (Chaplin, 2011: 8). Estimating HLM in this way would require first running the individual level model for each school individually, followed by an estimation of second stage equations to investigate the impact of school-level factors on the relationship between individual-level characteristics. This two-stage estimation strategy would allow researchers to claim legitimately that their estimation of the relationship between individual level characteristics is not biased by school level factors (Chaplin, 2011: 8). However, HLM is not estimated in two stages. Coefficients are estimated using both within- and between-school variation and so any omitted variables at the level of the school will bias estimates of the relationship between variables at the level of the individual. HLM assumes zero covariance between explanatory variables and error terms, while OLS estimation using fixed effects allows for non-zero covariance. HLM estimates may therefore be biased (Chaplin, 2011: 8).

One possible way in which to avoid this bias is by centring second level variables (Bryk and Raudenbush, 1992: 23). Centring variables allows the researcher to investigate how the dependent variable responds when the value of explanatory variables change. By centring variables, the researcher is able to see what a standard deviation change in the explanatory variable does to the dependent variable. Centring involves subtracting the group mean (the average value within the group) from the individual values of the variable in order to capture the variation while getting rid of the "group effect". Goldberger (1991: 42) points out that group centring is one way in which fixed effects modelling is conducted in OLS. Chaplin (2011: 9) explains that the HLM estimates arrived at when group-mean centring is used are unbiased. Therefore, despite the fact that HLM controls for the possibility that students selected from the same school are more similar to one another than would be the case had they been selected at random, the fact that HLM models are not estimated in two stages means that variables must first be centred in order to ensure that the estimates obtained using HLM are unbiased.

A question often asked when considering whether to use HLM is whether similar results may not be achieved in OLS by making use of interaction effects. Newman, Newman and Salzman (2010: 5) point out that interaction terms are usually used to investigate the effect within a certain group of a given variable already included in the model over and above the main effect of that variable on the outcome of interest. That is, interaction effects are used to ascertain whether the effect of a particular variable on the outcome variable in one group differs significantly from its overall effect in the entire sample. They explain this as the "differential effect" across groups. HLM investigates the differential effects across groups. The second level of an HLM model therefore provides insights into differences between groups (slope differentials, for example). Interaction terms provide information on the differences over and above the main effect of the explanatory variable in question. Including interaction terms in an OLS model is therefore not the same as explicitly modelling multiple levels of data – the overall objective of HLM.

In summary then, estimates obtained using HLM are only unbiased if variables are. Furthermore, fixed effects estimation in OLS, while remedying the problem of biased estimates resulting from unobservable characteristics at the level of the school, do not control for the likelihood that students attending the same school are more similar than students selected at random. Finally, estimates produced using interaction terms in OLS are different from those obtained using HLM, as interaction effects capture "altered" effects within groups. In the case of HLM, different estimates are obtained for each group. Estimates obtained using HLM and OLS are likely to be similar, however. The estimates presented in this chapter were obtained using HLM, given that it appears to control for more of the complications associated with modelling multilevel data. However, as a robustness check and for the sake of completeness, models were estimated by using OLS and controlling for cluster effects at the level of the classroom. These estimates are presented in Appendix C. The results are similar in size and significance to those obtained using HLM.

3.1 **Hierarchical linear modelling: The analytical method**

Hierarchical linear modelling is a method that effectively runs regressions of regressions. As explained above, multilevel modelling aims to predict outcomes based on variables from multiple levels (Luke, 2004: 9). In this chapter we investigate student performance in mathematics as a function of both student characteristics (e.g. age, gender, socioeconomic status) and characteristics of teachers (e.g. levels of educational attainment, experience, age). Students are therefore nested within classes[8]. The structure of the model is presented in equations 1 and 2 below.

$Level$ 1: $Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + r_{ij}$ **(1a)**

$Level$ 2: $\beta_{0j} = \gamma_{00} + \gamma_{01} W_j + u_{0j}$ **(1b)**

$\beta_{1j} = \gamma_{10} + \gamma_{11} W_j + u_{1j}$ **(1c)**

The subscript j in the equation for level 1 indicates that the model is being estimated j times, once for each of the j groups in the sample (Luke, 2004: 10). It is therefore possible (and indeed likely) that each of the j groups will have a different mean mathematics score (β_{0j}) and that the effects of individual level characteristics (for example, student socioeconomic status) on the outcome variable (β_{1j}) will differ for students taught by different teachers.

In equation 1 the intercept (β_{0j}) and slope (β_{1j}) as outcomes in the group model is straightforward. In equation 1b, the value of β_0 for group j is a function of the overall mean for the sample (γ_{00}) and the effect of the group-level characteristic W_j on the group average (γ_{01}). The additional variability in the average of group j is captured in the error term u_{0j}.

[8] Students are organised into classrooms, each of which is taught by a particular teacher. For the sake of this analysis, within-classroom differences in fact refer to within-teacher differences. The remainder of the paper will refer to within-classroom elements for the sake of brevity.

Similarly, the value of $\beta_1$in group j is modelled as a function of the overall mean impact of individual level characteristic (X_{ij}) on student outcomes (γ_{10}) and the effect of the group-level characteristic W_j on this relationship (γ_{11}). The variability in this relationship not accounted for in the model is captured by the error term u_{1j} (Luke, 2004: 10).

Equation 2 condenses the system of equations presented above into one prediction equation.

$$Y_{ij} = \left[\gamma_{00} + \gamma_{10}X_{ij} + \gamma_{01}W_j + \gamma_{11}W_jX_{ij}\right] + \left[u_{0j} + u_{ij}X_{ij} + r_{ij}\right] \qquad \textbf{(2)}$$

Equation 2 indicates that the level 1 parameters ($\beta_{0j}\ and\ \beta_{1j}$) are estimated indirectly through level 2, and the effects are given by the γs (Luke, 2004: 11). Equation 2 also indicates how the model is broken into fixed effects (the first set of brackets) and random effects (the second set of brackets). The random effects in multi-level modelling can be thought of as the variability that remains after level 1 and level 2 characteristics have been controlled for. This variation is comprised of classic individual level error (r_{ij}) as well as two error terms resulting specifically from the multi-level nature of the model. The first of these, u_{0j}, captures differences in the mean outcome between level 2 groups, and the second of these, u_{1j}, captures differences in the relationship coefficient between the level 1 characteristic and the outcome between level 2 groups (Luke, 2004: 11).

3.2 **Means-as-outcome regression**

For the purpose of this thesis, the hierarchical linear model that will be used is one which models the intercept term, or the average mathematics performance of students as a function of teacher characteristics. As mentioned before, this chapter aims to investigate the impact of teacher characteristics on student performance. Mean student performance within a school is therefore modelled at the second level. Relationships between student-level characteristics and the outcome variable will not be modelled as being functions of teacher-level characteristics. In terms of the model format presented in equation 1 above, then, the second level of the model is organised as shown in equations 3a to 3d below.

$$\beta_{0j} = \gamma_{00} + \gamma_{01}W_{1j} + \gamma_{02}W_{2j} + \ldots + \gamma_{0S}W_{Sj} + u_{0j} \qquad \textbf{(3a)}$$

$$\beta_{1j} = \gamma_{10} \qquad \textbf{(3b)}$$

$$\beta_{2j} = \gamma_{20} \qquad \textbf{(3c)}$$

$$\vdots$$

$$\beta_{Qj} = \gamma_{Q0} \quad \textbf{(3d)}$$

Where S = [1, 2, . . . , S] denotes the number of teacher-level characteristics included in the second level of the model. The combined model therefore takes the form of equation 4.

$$Y_{ij} = \gamma_{00} + \gamma_{10} + \gamma_{20} + \ldots + \gamma_{Q0} + \gamma_{01}W_{1j} + \gamma_{02}W_{2j} + \ldots + \gamma_{0S}W_{Sj} + u_{0j} \quad \textbf{(4)}$$

Where Q = [1, 2, . . . , Q] is the number of student level characteristics controlled for in the first level of the model.

4. **Modelling the impact of teacher characteristics on student performance**

Contextualising the research conducted in this paper in the model explained above requires first that we present the student-level or "within-classroom" model. This is the level 1 model explained in equation 1 above. This is presented in equation 5 below. Table 1 contains a description of the variables included in this equation.

$$\text{ZMAT}_{ij} = \beta_{0j} + \beta_{1j}(SES) + \beta_{2j}(Overage) + \beta_{3j}(Female) + \beta_{4j}(Mother\ matric) + \beta_{5j}(Father\ matric) + \beta_{6j}(Preschool\ less\ than\ 1) + \beta_{7j}(Preschool\ 1\ year) + \beta_{8j}(Preschool\ 2\ years) + \beta_{9j}(Preschool\ 3\ years\ plus) + \beta_{10j}(English\ sometimes) + \beta_{11j}(English\ most\ of\ the\ time) + \beta_{12j}(English\ always) + \beta_{13j}(Repeated\ once) + \beta_{14j}(Repeated\ twice) + \beta_{15j}(Repeated\ three\ times) + \beta_{16j}(Repeated\ grade\ 6) + \beta_{17j}(Extra\ tuition) + r_{ij} \quad \textbf{(5)}$$

Education production function theory suggests that student education outcomes are a function of both school-level (or "policy-controlled") characteristics and family- and peer-level (or "non-controlled") characteristics (Hanushek, 2007: 3). Family characteristics largely refer to socio-demographic characteristics and in equation 5 include socioeconomic status (*SES*), *Overage, Female, Mother matric* and *Father matric*. The relationship between SES and student performance is well-documented, particularly in the case of South Africa (Van der Berg et al., 2011). *SES* is included as a student-level explanatory variable to control for this relationship and to ensure that estimates observed for other explanatory variables – many of which are correlated with socioeconomic status – reflect the impact of those variables independently of the impact of SES. *Overage* and *Female* control for the possibility that children who are older than their appropriate age for their grade perform differently to those who are either the correct age for grade 6 or younger, and for the possibility that girls and boys perform differently. Students older than the grade-appropriate age seem likely to perform at a lower level than their peers given the possibility that they have repeated grades. However, dummy variables controlling for whether students have repeated a grade once, twice or three times and whether they are repeating their current grade (grade 6) are included to control for this possibility. As the results in section 6 indicate, the effect of being overage appears to work separately from the effect of repetition. Parental education is often included in the SES term in education production functions. The SES term in the SACMEQ III data was created using questions about assets in students' homes and did not include information on parental education. Parental education is an important socio-demographic indicator and whether or not a student's mother and father have attained matric are entered separately to investigate whether or not they have separate effects on student performance. As pointed out in section 2, testing in SACMEQ III in South Africa was conducted in English and Afrikaans. For the majority of South African students, neither of these is a first or home language. The frequency with which students speak English controls to some extent for this (*English sometimes, English most of the time, English always*), but the same variable does not exist for Afrikaans. *Extra tuition* controls for students receiving extra tuition but may well capture students with lower levels of ability rather than the effect of receiving instruction additional to that which they receive in the classroom. The number of years of preschool education is captured by four variables (*Preschool – less than 1 year, Preschool – 1 year, Preschool – 2 years and Preschool – 3 years plus*) in order to investigate whether investment in "school-readiness" has a significant impact on student performance.

The study investigates whether β_{0j} differs across teachers. The combined model of characteristics of both students and teachers is presented in equation 6.

$$\mathrm{ZMAT_{ij}} = \gamma_{00} + \gamma_{01}(Teacher\ is\ female) + \gamma_{02}(30\ to\ 39\ years\ old) + \gamma_{03}(40\ to\ 49\ years\ old) + \gamma_{04}(50\ to\ 59\ years\ old) + \gamma_{05}(60\ to\ 69\ years\ old) + \gamma_{06}(Teacher\ maths\ score) + \gamma_{07}(Experience) + \gamma_{08}(Days\ of\ training) + \gamma_{09}(Trained\ in\ mathematics) + \gamma_{010}(Trained\ to\ teach\ maths) + \gamma_{011}(Junior\ secondary\ education) + \gamma_{012}(Senior\ secondary\ education) + \gamma_{013}(A\ levels) + \gamma_{014}(Degree) +$$

$$\gamma_{015}(Less\ than\ 1\ year\ teacher\ training) + \gamma_{016}(1\ year\ of\ teacher\ training) + \gamma_{017}(2\ years\ of\ teacher\ training) + \gamma_{018}(3\ years\ of\ teacher\ training) + \gamma_{019}(more\ than\ 3\ years\ of\ teacher\ training) + \gamma_{020}(Parents\ sign\ homework) + \gamma_{021}(Test\ 2\ or\ 3\ times\ per\ terms) + \gamma_{022}(Test\ 2\ or\ 3\ times\ per\ month) + \gamma_{023}(Test\ at\ least\ weekly) + \gamma_{024}(Average\ class\ size) + \gamma_{025}(Rural) + \gamma_{026}(Average\ classroom\ SES) + \gamma_{027}(Private) + \beta_{1j}(SES) + \beta_{2j}(Overage) + \beta_{3j}(Female) + \beta_{4j}(Mother\ matric) + \beta_{5j}(Preschool\ less\ than\ 1) + \beta_{6j}(Preschool\ 1\ year) + \beta_{7j}(Preschool\ 2\ years) + \beta_{8j}(Preschool\ 3\ years\ plus) + \beta_{9j}(English\ sometimes) + \beta_{10j}(English\ most\ of\ the\ time) + \beta_{11j}(English\ always) + \beta_{12j}(Repeated\ once) + \beta_{13j}(Repeated\ twice) + \beta_{14j}(Repeated\ three\ times) + \beta_{15j}(Repeated\ grade\ 6) + \beta_{16j}(Extra\ tuition) + r_{ij} + u_{0j} \qquad \textbf{(6)}$$

The research question is whether or not teacher characteristics impact significantly on student performance. The variables at the teacher level in equation 6 are grouped according to four categories: demographic characteristics, education and experience characteristics, effort characteristics and school/classroom characteristics.

Demographic characteristics*:* Teacher gender may be important in explaining student performance if male and female teachers differ significantly from each other in terms of their ability to teach. *Teacher female* is included to control for whether a teacher is female and whether this has a statistically significant effect on mean student mathematics performance. Teacher age is controlled for using dummy variables for 10 year bands and the impact of teachers' age is measured relative to the youngest group of teachers (19 to 29 year olds). Significant coefficients on these variables may indicate either inherent differences in the ability of teachers to improve student performance associated with teacher age, or potentially differences in the training received by teachers trained at different times in South Africa.

Education and experience*: Experience*[9] is included to capture the number of years that teachers have been teaching. Literature on teacher experience suggests that beyond the initial years of teacher experience, the impact of having taught for longer periods of time becomes smaller. Teaching experience is rarely found to be statistically significant in its impact on

[9] Teaching experience and teacher age may have conflating effects on student performance. However, the model was run without controlling for teaching experience and this made very little difference to the age coefficients. Experience and age were asked separately in the teacher questionnaire. Both have been retained as they control for different characteristics, and both are necessary for the sake of this analysis.

student performance (Koedel, 2007). It is included in this analysis as dummy variables capturing experience in 5 year bands. Dummy variables capturing teachers' level of educational attainment are included to ascertain whether a certain level of education impact student performance significantly. Given the restructuring of teacher training with the closing of teacher training colleges in 2000, it is important to investigate the extent to which the attainment of a university degree impacts on student performance.

Days of training captures the time teachers spent participating in in-service training courses. In-service training programmes are perceived by researchers to be largely ineffective in affecting student performance (Taylor, 2012: 15).

Teacher training is captured by dummy variables reflecting whether teachers received less than 1 year, 1 year, 2 years or 3 years of teacher training. In the South African education system, teachers may qualify via various channels, an explanation of which is included in Appendix D. It is important to investigate the extent to which different avenues to teacher training impact on student performance.

Teacher maths score[10] is included to control for teachers' own mathematical content knowledge. The model is run including teacher maths score as well as excluding it. This is done in order to ensure that the impact of teacher training variables is separated from teachers' own performance in mathematics. Finally, dummy variables controlling for whether teachers are trained to teach (i.e. pedagogical training) and whether they are trained specifically to teach maths are included.

Effort characteristics*: Parents sign homework* is included as a dummy variable to capture the extent to which teachers ensure that students complete their assigned work. The variable is intended to proxy for teachers' interest in students' progress. Dummy variables controlling for the frequency of testing are included to measure teacher "engagement" with students' progress. Marking of tests is time-consuming and often tedious work for teachers. It is assumed that higher frequencies of testing indicate higher levels of effort. Important to note is that both variables are self-reported by teachers. It is likely therefore that the extent to which these activities occur is over-stated.

[10] Teacher maths score is missing for 98 teachers in the SACMEQ III dataset. Where possible, missing data were replaced with the mean mathematics score of teachers within the same school. Teachers from schools in which no teachers wrote the mathematics tests were excluded from the model in which teacher maths score was included as an explanatory variables. This meant that 29 teachers were dropped from this sample.

***School and classroom characteristics**:* A number of variables included in the teacher-level model are in fact school-level characteristics, but in the case of the SACMEQ data in a significant number of schools only one classroom was sampled. The classroom is therefore completely identified by the school and so for these variables (with the exception of *Classroom SES)* no variation occurs at the level of the school. The school-level variables, namely *Rural, Private school* and *Average class size* are therefore included to control for differences that are observed between students attending schools with these characteristics and those attending schools in which these characteristics are absent.

5. Results

The multi-level nature of education data necessitates hierarchical or multi-level modelling. The overall variation in student performance can be at the level of the student and the teacher. In other words, there are characteristics of both students and their teachers that influence student performance. A first step in performing hierarchical linear modelling is to ascertain whether or not any variation occurs at the higher level. The extent to which student performance is attributable to teacher characteristics therefore needs to be tested.

Formally partitioning the variance into the components that occur at the level of the student and the teacher is achieved by running a fully unconditional model in which students' mathematics performance is allowed to vary without including controls for any level 1 (student) or level 2 (teacher) characteristics. This is presented in equation 7 below.

$$Y_{\boldsymbol{ij}} = \beta_{0j} + r_{ij}$$

where $$\beta_{0j} = \gamma_{00} + u_{ij} \qquad \textbf{(7)}$$

The variance component associated with level 1 (i.e. the student level), r_{ij} (σ^2), is estimated at 0.452, while that associated with level 2 (i.e. the level of the teacher), $u_{ij}(\tau_{00})$, is estimated at 0.747. The intra-class correlation coefficient (ICC) is the variance at the level of the teacher as a proportion of overall variance. The ICC (ρ) is therefore calculated according to equation 8.

$$\rho = \frac{\tau_{00}}{\sigma^2+\tau_{00}} = \frac{0.738}{0.451+0.738} = 0.621 \qquad \textbf{(8)}$$

The variances presented above result in ρ = 0.621, indicating that just over 62% of the variation in students' mathematics performance is explained at the level of the teacher or school. There therefore seems to be a case for using multi-level modelling to explain the factors influencing student performance. The reliability estimate of the intercept term,[11] which measures the ratio of the variance of the parameter estimate to that of the sample mean for the intercept term, is 0.957, indicating that a large proportion of the variance in mean mathematics performance across teachers may potentially be explained at the level of the teacher.

This analysis of variance is conducted without including controls at either level. This may be problematic for two reasons. Firstly, it is possible that group-level predictors impact substantially on the outcome variable but that two variables have opposite effects with the result that they cancel each other out (Chaplin, 2011: 11). In this case, it may appear that no variation occurs at the level of the group when in fact group-level characteristics are significant in determining the outcome. Secondly, individual and group-level characteristics may offset each other, again masking sources of variation in the outcome variable and making it seem as if multi-level modelling is unnecessary when in fact significant variation occurs at the level of the group (Chaplin, 2011: 12). In both cases then the danger is that group-level variation is being masked. As shown below, it is unlikely that this is a problem in South Africa given the large proportion of variation in student mathematics explained at the level of the classroom.

The within-classroom model is presented in table 6 below.

TABLE 6: Student-level model

	Estimated Fixed Effects	
	Coefficients	**Standard Errors**
Intercept	0.105***	0.036
Student SES	0.132***	0.015
Overage	-0.130***	0.025
Female	-0.004	0.018
Mother completed matric	0.099***	0.020

[11] The reliability estimate is calculated as $\lambda_j = \frac{\tau_{00}}{\tau_{00}+\frac{\sigma^2}{n}}$

Father completed matric	0.051***	0.049
Less than 1 year preschool	0.015	0.042
1 year of preschool	0.036	0.024
2 years of preschool	0.060**	0.028
3 or more years of preschool	0.109***	0.029
Speaks English sometimes	0.166***	0.026
Speaks English most of the time	0.188***	0.042
Speaks English always	0.310***	0.059
Repeated a grade once	-0.215***	0.027
Repeated a grade twice	-0.210***	0.039
Repeated a grade three times	-0.250***	0.050
Repeated grade 6	-0.033	0.032
Receives extra tuition	-0.159***	0.044

Estimated Random Effects			
	Standard Deviation	**Variance**	**Chi-Squared**
Intercept	0.685	0.469	7711.760
Within-classroom	0.653	0.426	
Reliability of teacher-level random effects			
	Mean score	0.937	

Source: Own calculations from SACMEQ III (SACMEQ, 2007).

The results presented in table 6 above indicate that, predictably, socioeconomic status has a positive and significant impact on student mathematics performance. The coefficient in the table indicates that if student socioeconomic status increased by 1 standard deviation and the values of all other variables were held constant, student mathematics performance would improve by 0.130 standard deviations. Overage students perform 0.130 below their peers who are not overage (i.e. who are either the correct age for their grade or younger than the correct age for their grade) while students whose mothers completed matric outperform those whose mothers did not by 0.099 standard deviations. The impact of fathers having completed matric is positive and significant, but smaller than that observed for mothers at 0.051. This is in line with what is observed internationally. Students who have received 1, 2 and 3 years of preschooling outperform those who have had no preschooling by 0.035, 0.060 and 0.109 standard deviations respectively, while students who speak English outside the classroom sometimes, often and always outperform those who do not speak English outside the classroom by 0.166, 0.188 and 0.309 standard deviations respectively. Students who have repeated a grade once, twice or three times perform 0.22, 0.21 and 0.25 standard deviations below students who have not repeated a grade, respectively, indicating that there is no real difference in the performance amongst students who have repeated grades.[12] The coefficient for students repeating grade 6 is not statistically significantly different from that of students who are not repeating grade 6, suggesting that students repeating grade 6 do not perform differently from students repeating other grades[13]. Students receiving extra tuition are outperformed by their peers not receiving extra tuition by 0.159 standard deviations. This may well reflect a lower ability in the students receiving extra tuition rather than the extra tuition having a negative impact on their performance.

Controlling for the student level characteristics decreases the within-classroom variance by roughly 37% from 0.747 to 0.469. The fact that level 1 characteristics explain so little of the variance of the mean highlight the fact that a substantial portion of the variation in student performance is explained at a higher level or not at all. Most educational performance data for South Africa contain neither pre-test scores nor racial classification – both of which are highly correlated with educational performance. This is obvious in the case of pre-test scores. In the case of race, performance in the South African education system is still significantly correlated with performance of the system under apartheid, with the part of the schooling system historically serving South Africa's white population far outperforming the part of the schooling system historically serving South Africa's black population. The historically white part of the schooling system is now substantially more representative of South Africa's population than in previous years, while the historically black portion remains almost entirely black. Most white children find themselves in the historically white part of the schooling system and for this reason race is a significant determinant of schooling performance.

In addition, the variables included in the within-classroom model control for the home background and variables pertaining to previous education performance. Unobservable

[12] An F-test confirms this.
[13] An F-test confirms this.

characteristics such as intelligence or ambition play a significant role in school performance. However, it is impossible to measure and control for them.

In order to investigate the extent to which teacher-level characteristics impact on student performance, level 2 variables are added to the model. Table 7 below presents the results from the full multilevel model. Column 1 contains the results for the full teacher model including teacher maths score, while column 2 excludes teacher maths score.

TABLE 7: Full hierarchical linear model

Variable	**Model 1**		**Model 2**	
	Coefficient	**Std deviation**	**Coefficient**	**Std deviation**
Intercept	0.231	0.208	0.278	0.286
TEACHER DEMOGRAPHIC CHARACTERISTICS				
Female	0.071	0.046	0.063	0.045
30 to 39 years of age	-0.345***	0.130	-0.378***	0.131
40 to 49 years of age	-0.389***	0.132	-0.474***	0.132
50 to 59 years of age	-0.522***	0.160	-0.618***	0.161
60 years and older	-0.325***	0.296	-0.360*	0.301
TEACHER EDUCATION AND EXPERIENCE				
Teacher maths score	0.105***	0.024		
Experience: 6 to 10 years	0.150*	0.084	0.181**	0.084
Experience: 11 to 15 years	0.031	0.064	0.086	0.062
Experience: 16 to 20 years	-0.033	0.077	-0.022	0.075
Experience: 21 to 25 years	-0.027	0.083	-0.038	0.083
Experience: 26 to 30 years	0.170	0.141	0.226	0.138
Experience: 31 to 35 years	0.267*	0.162	0.323**	0.164
Experience: 36 to 40 years	0.042	0.266	0.071	0.270
Experience: 41 plus years	-0.412	0.624	-0.434	0.637
Number of days training received	-0.000	0.001	0.000	0.00
Trained in mathematics	0.093	0.303	0.086	0.308

Trained to teach mathematics	-0.213	0.302	-0.184	0.306
Completed jr secondary education	-0.029	0.164	0.006	0.166
Completed sr secondary education	0.058	0.086	0.064	0.087
Completed A-levels	0.002	0.072	0.033	0.071
Completed a degree	0.097*	0.059	0.111*	0.058
Received less than 1 year training	0.923	0.644	0.579	0.453
Received 1 year of training	0.029	0.306	0.011	0.308
Received 2 years of training	0.254	0.293	0.191	0.297
Received 3 years of training	0.169	0.280	0.112	0.284
Received 3 years plus of training	0.215	0.281	0.180	0.285
TEACHER EFFORT				
Parents sign students' homework	0.032	0.048	0.023	0.048
Test 2 to 3 times per term	0.020	0.075	0.034	0.076
Tests 2 to 3 times per month	0.025	0.080	0.015	0.081
Tests at least once per week	0.088	0.087	0.082	0.088
SCHOOL AND CLASSROOM CHARACTERISTICS				
Rural	-0.007	0.055	-0.001	0.054
Classroom SES	0.568***	0.040	0.683***	0.036
Private school	0.002	0.107	-0.024	0.108
Average class size (of the school)	-0.006***	0.002	-0.006***	0.002
STUDENT CHARACTERISTICS				
SES	0.063***	0.013	0.062***	0.012
Overage	-0.096***	0.022	-0.101***	0.021
Female	-0.007	0.015	-0.003	0.014
Mother completed matric	0.074***	0.017	0.072***	0.017
Father completed matric	0.048***	0.017	0.045***	0.017
Less than 1 year preschool	0.018	0.037	0.024	0.03

1 year of preschool	0.033	0.020	0.026	0.020
2 years of preschool	0.035	0.025	0.040	0.025
3 or more years of preschool	0.093***	0.024	0.094***	0.024
Speaks English sometimes	0.157***	0.020	0.157***	0.020
Speaks English most of the time	0.160***	0.034	0.158***	0.032
Speaks English always	0.271***	0.039	0.249***	0.038
Repeated a grade once	-0.204***	0.022	-0.206***	0.021
Repeated a grade twice	-0.229***	0.038	-0.211***	0.036
Repeated a grade three times	-0.249***	0.050	-0.218***	0.046
Repeated grade 6	-0.043	0.032	-0.052*	0.030
Receives extra tuition	-0.147***	0.034	-0.137***	0.032

Estimated Random Effects

	Standard Deviation	**Variance**	**Chi-Squared**
Intercept	0.416	0.173	3 468.531
Within-classroom	0.651	0.424	

Reliability of teacher-level random effects

Mean score	0.852

Source: Own calculations from SACMEQ III (SACMEQ, 2007).

The results obtained from the full model are discussed for the model excluding teacher mathematics score as this model is run for a greater number of observations. The results obtained for both specifications are largely similar, however. Coefficients which differ markedly from each other will be discussed where relevant. For the most part, however, they are largely similar.

Teacher demographic characteristics*:* Whether a teacher is female does not have a statistically significant impact on student performance. An interesting result obtained is the effect of teacher age on mean student performance. The coefficients on age indicate that relative to the reference group (teacher age 19 to 29 years old – the youngest group of teachers in the sample), the mean mathematics score of students taught by teachers from all other age groups is lower. Furthermore, with the exception of the coefficients on SES and less than 1 year of teacher training, the coefficients on teacher age groups are the largest amongst the teacher level characteristics. Indeed the mean mathematics score of students taught by teachers who are 30 to 39 years old, 40 to 49 years old, 50 to 59 years old and older than 60 are respectively 0.378, 0.474, 0.618 and 0.360 standard deviations below that of students taught by teachers belonging to the youngest age group.[14] The size of the coefficient for the group of teachers aged 50 to 59 years old is slightly higher than for the other age groups, but other than this coefficients for different age groups seem consistent.[15] This may say something about teacher training, given the movement away from teacher training colleges in 2000. This is discussed in greater depth later.

Teacher education and experience: Some interesting results are observed for variables capturing teacher qualifications. The mean performance of students being taught by teachers who have obtained a university degree is 0.111 standard deviations higher than that of students taught by a teacher who has not obtained a university degree.

In terms of teaching experience, coefficients for two of the dummy variables are statistically significant – *Experience 6 to 10 years* and *Experience 31 to 35 years*. The coefficients on these variables indicate that relative to students being taught by teachers with 5 or less years of teaching experience, students being taught by a teacher with 6 to 10 years of teaching experience perform on average 0.181 standard deviations better, and students being taught by teachers with between 31 and 35 years of teaching experience perform 0.323 standard deviations above other students. Interestingly, in model 1 (which controls for teachers' performance on their mathematics tests), teachers' mathematics test performance results are statistically significantly positively related to mean student mathematics performance. As teachers' maths scores are z-scored, the coefficient of 0.105 indicates that an improvement of

[14] A possible explanation for the difference in the ability of younger teachers to elicit superior performance from their students is the fact that they themselves have a better grasp of the mathematical content which they are required to teach. An important part of understanding the differences illustrated by the coefficients above is investigating whether younger teachers are better at maths or whether they are better teachers. This is tested by interacting teacher test score with the dummy variables controlling for age. However, the coefficients are small and statistically insignificant. It does not appear therefore that this effect works through superior mathematical content knowledge amongst younger teachers.

[15] The model was re-run with different cohorts of teachers as the reference group. The results indicate that although the differences in the coefficients are smaller in size amongst groups older than the youngest group, the ability to elicit stronger performance from students does differ by teacher age, with younger teachers out-performing their older colleagues. This is confirmed by an F-test.

1 standard deviation in teacher maths performance results in an improvement of 0.105 standard deviations in mean mathematics performance amongst students. This is an important finding in a study in which few teacher characteristics appear to impact significantly on students.

Teacher effort: None of the teacher effort variables included in the model appears to impact on mean mathematics performance in a significant way. This may be due to the fact that these variables are self-reported by teachers. The frequency of testing as well as whether parents are required to sign homework may well be over-reported.

School and classroom characteristics: The large and statistically significant coefficient observed for classroom SES is to be expected. The coefficient of 0.627 indicates that a 1 standard deviation increase in classroom SES is associated with a 0.627 standard deviation increase in mean mathematics performance. The statistically significant negative coefficient for *Average class size (of the school)* is intuitive, suggesting that larger classes are associated with weaker performance. The size of the coefficient is very small, however. Increasing class size by one student decreases mean student performance by 0.006 of standard deviation. Despite the fact that it is statistically significant, it is not economically significant. It is too small to indicate any real relationship between the variables.

6. Discussion and conclusion

The results presented above are important in the context of South Africa's education system. Teachers are an important resource in education and it is necessary to understand how best to utilise the resource.

The results for the hierarchical linear model reveal that younger teachers are better able to increase the mean performance of students. In order to test whether this is a trend observed amongst teachers across different countries or whether this is a trend particular to South Africa, the identical HLM model was run for 3 other countries in the SACMEQ III dataset – two of South Africa's neighbouring countries, Botswana and Zimbabwe, and a high-performing East African country, Kenya. The coefficients on the teacher age variables are presented in table 8 below.

TABLE 8: HLM coefficients on teacher age variables for 4 SACMEQ countries

Teacher age	Botswana	Kenya	Zimbabwe	South Africa
30 to 39 years old	-0.075	0.062	0.005	-0.378***
	(0.078)	(0.109)	(0.103)	(0.131)
40 to 49 years old	-0.029	-0.232	-0.115	-0.474***
	(0.103)	(0.142)	(0.130)	(0.132)
50 to 59 years old	0.199	-0.561***	-0.287	-0.618***
	(0.152)	(0.191)	(0.201)	(0.161)
60 to 69 years old	-	-	-0.318	-0.360*
			(0.588)	(0.301)
Number of students	3 842	4 272	2 983	8 917
Number of teachers	342	259	273	498

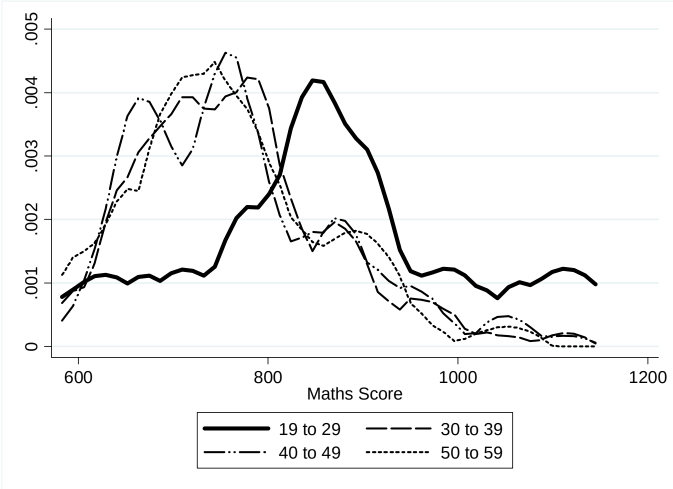
Source: Own calculations from SACMEQ III (SACMEQ, 2007).

The pattern for lower mean mathematics performance amongst students being taught by older teachers appears in both Kenya and Zimbabwe. The magnitude of these coefficients is comparable with those observed in South Africa. In fact in Kenya, the coefficient for teachers aged 50 to 59 years old is almost double that of South Africa's. However, this is the only coefficient which is statistically significant whereas in the case of South Africa, the coefficients for all teacher age groups are statistically significant relative to the reference group of teachers aged 19 to 29 years old.[16]

This discussion investigates why this may be the case. As described earlier, the studies conducted by SACMEQ in 2000 and 2007 included teacher tests. Due to union objections to teachers being tested, South African teachers participated only in the teacher test conducted in 2007 and were allowed to opt out of being tested. Interestingly, teacher performance on the mathematics test appears to differ according to age in the same way that teachers' ability to elicit test performance from their students does. Figure 23 below presents the distribution of teacher performance on mathematics tests for teachers of different ages.

[16] The coefficient for South African teachers aged 60 and older is not statistically significant. However, this group is comprised of just 4 teachers.

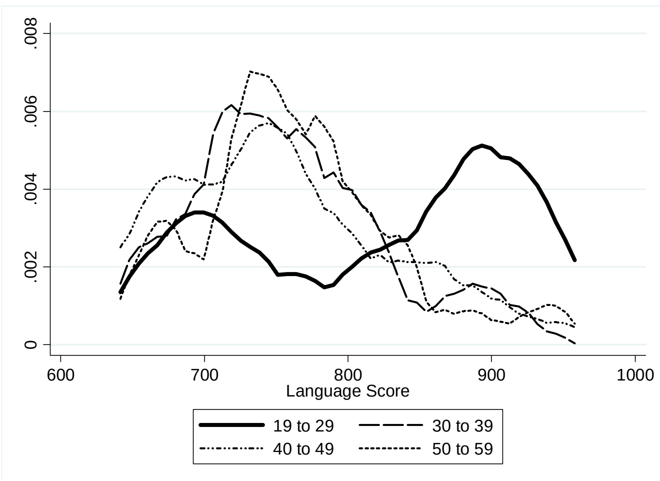
FIGURE 23: Teacher mathematics score by age group



Source: SACMEQIII, 2007.

The kernel density curves drawn in figure 23 demonstrate that younger teachers perform at a significantly higher level in the mathematics test than teachers in older age groups. Similar results are obtained with regards to teacher performance on language tests. Figure 24 presents the distribution of language performance results amongst teachers in different age groups. As seen in the mathematics test, teachers in the age group 19 to 29 perform better than their counterparts in older age groups in the language test.

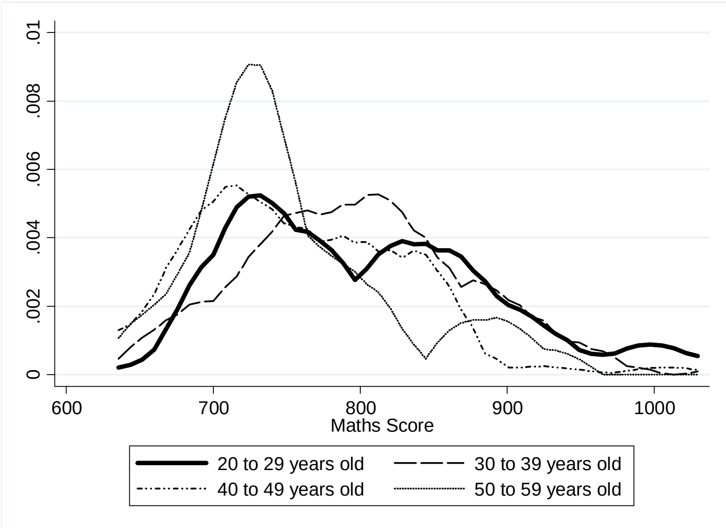
FIGURE 24: Teacher language score by age group



Source: SACMEQIII, 2007.

Kernel densities for Botswana, Kenya and Zimbabwe were drawn for teacher performance in mathematics tests in figures 25, 26 and 27 below and for teacher performance in language tests in figures 28, 29 and 30.

FIGURE 25: Teacher mathematics score by age (Botswana)

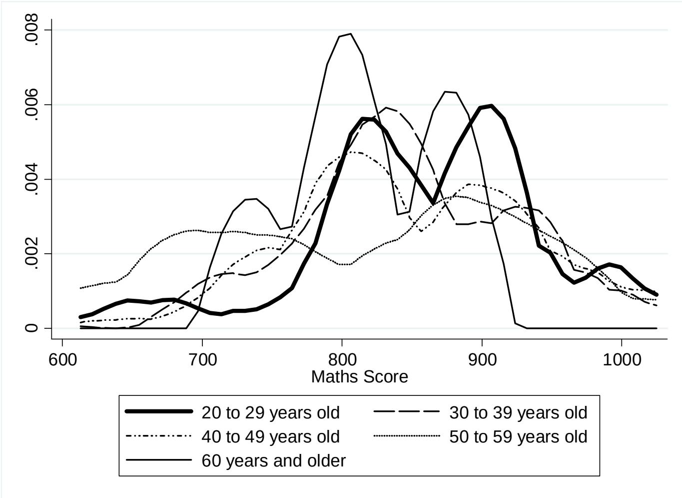


Source: SACMEQIII, 2007

FIGURE 26: Teacher mathematics score by age (Kenya)

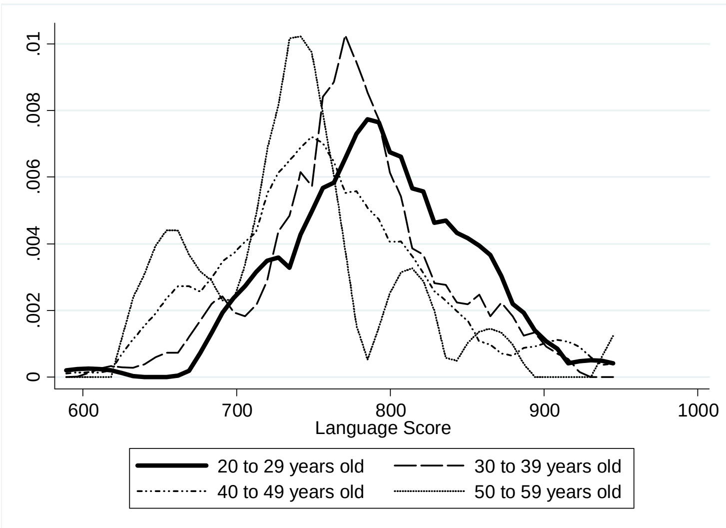
Source: SACMEQIII, 2007

FIGURE 27: Teacher mathematics score by age (Zimbabwe)



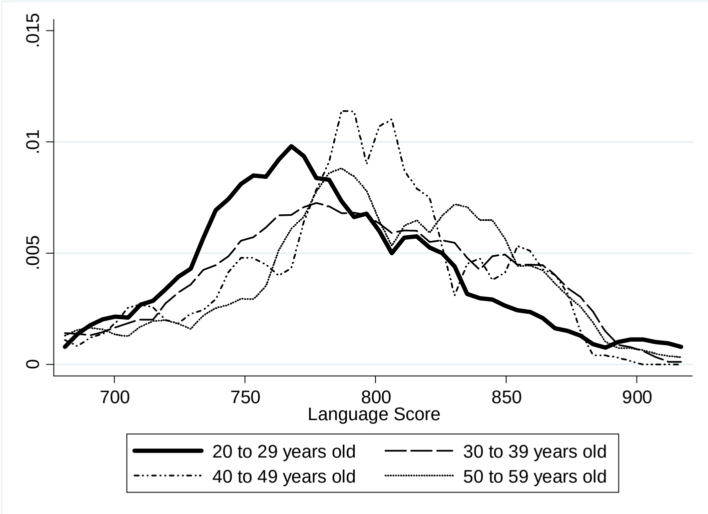
Source: SACMEQIII, 2007

FIGURE 28: Teacher language score by age (Botswana)



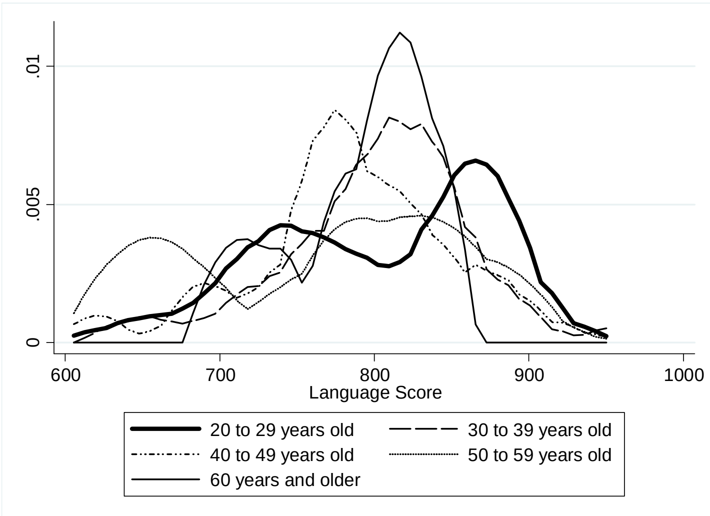
Source: SACMEQIII, 2007

FIGURE 29: Teacher language score by age (Kenya)



Source: SACMEQIII, 2007

FIGURE 30: Teacher language score by age (Zimbabwe)



Source: SACMEQIII, 2007

The differences in the performance of teachers of different ages in Botswana, Kenya and Zimbabwe are not as marked as they are in South Africa. It seems therefore that this is a phenomenon particular to South Africa.

A basic OLS regression was run to investigate whether the difference in performance between teachers is statistically significant. The results are presented in table 9 below.

TABLE 9: Regression of teacher test performance on teacher age

	Coefficient and standard deviation	
Variable	**Mathematics**	**Language**
30 to 39 years old	-0.997*	-0.715***
	(0.555)	(0.269)
40 to 49 years old	-1.586***	-0.701***
	(0.552)	(0.269)
50 to 59 years old	-1.237**	-0.738***
	(0.596)	(0.286)
60 years and older	-1.452	-0.330
	(1.243)	(0.408)
Constant	0.416	0.734***
	(0.530)	(0.256)
Sample size	497	415
R-squared	0.03	0.01

Source: Own calculations from SACMEQ III (SACMEQ, 2007).

It therefore appears that older teachers are outperformed by younger teachers in both mathematics and language. Younger mathematics teachers also seem better able to elicit better performance from their students. It is important to investigate the possible reasons for this pattern. Similar estimates were found by using data from PIRLS 2006 on reading and literacy amongst students of a similar age. Shepherd (2013: 31) used weighted least squares regression to investigate the determinants of student reading and literacy and found a large, positive and statistically significant coefficient for teachers who are 30 years old or younger. Interestingly, this is only observed amongst teachers of students who wrote the PIRLS test in an African language and who were therefore in the historically black part of the schooling system. Amongst students writing the test in English of Afrikaans, the coefficient was somewhat smaller, negative and statistically insignificant (Shepherd, 2013: 31). Interestingly, when the model is run for quintiles 1 to 4 for South Africa in the SACMEQ III dataset, the coefficients diminish in size and although still statistically significant, they are significant at a lower level. The results are presented in Appendix C.

More than one explanation may exist for the differential ability of younger teachers to elicit stronger performance from their students. Younger teachers may relate better to their students because they are closer in age than their older counterparts. Another possibility is that changes to teacher training may have left teachers trained under a new system better equipped to teach. We are able to test these hypotheses using data from the second SACMEQ survey conducted in 2000. As mentioned above, no teacher tests were conducted for South African teachers in 2000. Other than that, the questionnaires were almost identical, making it possible to compare the two surveys and so the same model can be run for SACMEQ II data. If younger teacher are inherently better at teaching (and not as a result of different teacher training) we expect to see similar coefficients to those observed using the SACMEQ III data for teacher age variables in similar models from different time periods.

The full HLM model was run using SACMEQ II data. The full results are presented in Appendix b. Table 10 presents the coefficients on the teacher age variables obtained when data from the 2000 study were used.

TABLE 10: HLM coefficients on teacher age variables using SACMEQ II (2000)

Teacher age	Coefficient (Std. Error)
30 to 39 years old	0.003 (0.120)
40 to 49 years old	0.315* (0.189)
50 years and older	0.671** (0.232)
Number of students	3 135
Number of teachers	187

Source: Own calculations from SACMEQ II (SACMEQ, 2000).

The coefficients in table 10 are quite different from those obtained from the 2007 data of the SACMEQ III survey. In fact, only the teachers aged 50 to 59 differ significantly from the youngest group of teachers and in this case, they seem to elicit better performance from their students. According to this data then, the statistically significant negative coefficients observed for teachers older than the 29 years of age (relative to the youngest group) are not explained by an inherent ability of younger teachers to positively influence mean student performance. It is possible then that differences in teacher training explain the differences in the student performance according to the age of their teacher.

As explained in the next subsection, teacher training is one of the few characteristics that may render younger teachers better able to impact positively on their students' performance. Changes in teacher training in the South African education system occurred in the late 1990s and early 2000s – the time in which the youngest cohort of teachers were trained. The following section discusses these changes.

6.1 **Differences in teacher training**[17]

An obvious avenue to pursue in understanding the differences that are observed in the performance of teachers of different ages is to investigate the extent to which the training received by teachers differed across years. A potential source of differences in teacher training is the shift from teacher training colleges as the institutions responsible for training teachers to the incorporation of teacher training within universities. Chisholm (2009: 9) explains that teacher training colleges expanded predominantly in the 1960s. The apartheid state located the majority of teacher training colleges in the "homeland" areas with the objective of staffing the institutions with the graduates. Chisholm (2009: 14) explains that enrolment in the teacher colleges was high due to the fact that opportunities in the formal economy were restricted for non-white South Africans, and entering a teacher training college was one of the very few ways in which people living in the homelands could enter higher education.

Teacher training colleges were expensive to run and were heavily subsidised by the state (Chisholm, 2009: 16). Because of a movement towards decreasing unit costs and enhancing productivity within the higher education sector, teacher colleges were offered the option of

[17] A brief explanation of the minimum requirements for the education of teachers is contained in Appendix C.

remaining open as independent institutions if they were able to enrol 2 000 full-time students in 1999, or becoming integrated as part of universities or universities of technology. Teacher training colleges were formally incorporated into universities and universities of technology from January 2001 (Chisholm, 2009: 16).

Teachers trained after the incorporation of teacher training colleges into universities or universities of technology would therefore have been 25 years old in 2007 when SACMEQ III was conducted[18] and allowing for some violations of the assumptions explained in footnote 1 below, the age group of 19 to 29 years old (the reference group in the analysis conducted above) captures teachers who are likely to have completed their teacher training at universities or universities of technology.[19]

If we assume that teacher training does in fact influence teacher performance, then it appears that teachers trained at universities and universities of technology are better able to teach than are teachers trained at teacher training colleges. If this is the correct interpretation of the results obtained in table 4, it has important implications for the teacher training landscape in South Africa. South African teacher unions have since 2002 called for the reopening of teacher training colleges (Chisholm, 2009: 17). The South African Democratic Teachers Union (SADTU), the biggest union as it represents two thirds of teachers (Wills: 2014: 4), is of the opinion that teacher shortages (particularly in the areas of mother tongue and foundation phase education) result in excessively large class sizes which interfere significantly with the ability of their members to provide quality education. Indeed, at SADTU's 2006 National Conference, there was a recommendation for setting a maximum acceptable class size of 30 students – a number which requires substantial increases in teacher supply in order to be achieved (Chisholm, 2009: 17). This resulted in SADTU's 2007/08 call for the reopening of teacher training colleges.

A second argument in favour of reopening teacher training colleges has to do with the quality of teacher training provided by universities and universities of technology. Patterson and Arends (2008: 85) are of the opinion that primary and secondary school teaching are not

[18] With the data available there is no way of knowing at what age teachers were trained. The age of 25 is based on the assumptions that teachers started higher education directly after finishing secondary school, and that teachers left secondary school at the grade appropriate age of 18, therefore turning 19 in their first year of tertiary education. In many instances these assumptions are most definitely violated. It is likely for example that individuals took longer than the prescribed amount of time to complete tertiary education, and that individuals started teacher training after having completed other courses of study.

[19] 73% of the teachers in this age group are younger than 25 years old.

given the attention they require in the higher education system. They consider university fees for primary education high enough to exclude candidates from the teaching profession. Finally, university education is considered by teachers already teaching in schools to be excessively theoretical and abstract relative to what is required to teach primary school (Patterson and Arends, 2008: 86). Teachers and lecturers trained in teacher training colleges feel that universities and universities of technology lack the "hands on" practical guidance that was provided by colleges. They are of the opinion that principals and experienced teachers do not have the same opportunities for involvement in training future teachers as had been available in teacher training colleges (Chisholm, 2009: 17).

For various reasons, therefore, there is a strong belief that re-opening teacher training colleges may improve the quality (and quantity) of teachers in general, and primary teachers in particular. The evidence above suggests that this may not be the case.

6.2 **Other sources of differentials by teacher age**

Other explanations for differences in the performance of older and younger teachers have less to do with the structures within which teacher training takes place and more with the nature of teaching itself. Anecdotal evidence from teachers suggests that younger teachers are better able to engage and build rapport with their students because they are closer in age to students and because successful teaching requires high levels of energy. Younger teachers are also likely to be more familiar with the current curriculum and may therefore be more familiar with the content they are required to teach to students (Education Forum, 2006). An unflattering view of the performance gap between older and younger teachers is the tendency or willingness of younger teachers to "cheat" or teach to the test in order to appear to be performing well, compared to older teachers who would probably be more intent on ensuring that students receive a broader, more complete education rather than to focus on what is prescribed by the curriculum (Education Forum, 2006). Literature on differences in performance of teachers by age is scarce in the area of primary education. Very little empirical evidence exists of such disparities, which renders the results obtained in this paper quite important.

The most important finding from this chapter then has been that younger teachers are better able to elicit performance from students in mathematics at a grade 6 level. Similar results are found by Shepherd (2013) using different data, also at a grade 6 level but for performance in reading literacy. More must be done to fully understand this finding and to further investigate the reasons for differences in the ability of teachers of different ages to affect student performance. Differences in the training received by teachers in universities and universities of technology and that received by teachers trained at teacher training colleges need to be understood. How exactly do these differences translate into student learning? Are their unobservable characteristics according to which teachers differ that are correlated with age? If so, what should be done to ensure that student have access to teachers with these characteristics?

References

Anderson, J.B. 2000. Factors affecting learning of Mexican primary school children. *Estudios Economicos.* 15(1):117-152.

Angrist, J.D. & Lavy, V. 2001. Does teacher training affect pupil learning? Evidence from matched comparisons in Jerusalem public schools. *Journal of Labor Economics.* 19(2):343-369.

Arif, G.M. & Saqib, N. 2003. Production of cognitive and life skills in public, private, and NGO schools in Pakistan. *Pakistan Development Review.* 42(1):1-28.

Ball, D., Thames, M.H. & Phelps, G. 2008. Content knowledge for teaching: What makes it special? *Journal of Teacher Education.* 59:389-408.

Behrman, J.R., Kahn, S., Ross, D. & Sabot, R. 1997. School quality and cognitive achievement production: A case study for rural Pakistan. *Economics of Education Review.* 16(2):127-142.

Chaplin, D. 2008. Hierarchical linear modelling: Strengths and weaknesses. Paper prepared for the conference of the Association of Public Policy Analysis and Management, November 2008.

Chetty, R., Friedman, J.N. & Rockoff, J.E. 2011. *Measuring the impact of teachers: Teacher value-added and student outcomes in adulthood.* (National Bureau of Economic Research (NBER) working paper no. 17699). Cambridge MA: NBER.

Chingos, M.M. & Peterson, P.E. 2011. It's easier to pick a good teacher than to train one: Familiar and new results on the correlates of teacher effectiveness. *Economics of Education Review.* 30:449-465.

Chisholm, L. 2009. *An overview of research, policy and practice in teacher supply and demand 1994-2008.* Cape Town: HSRC Press.

Clotfelter, C.T., Ladd, H.F. & Vigdor, J.L. 2007. Teacher credential and student achievement: Longitudinal analysis with student fixed effects. *Education of Education Review.* 26:673-682.

Duncan, C., Jones, K. & Moon, G. 1998. Context, composition and heterogeneity: Using multilevel models in health research. *Social Science and Medicine.* 46:97-117.

Education Forum. 2006. Teacher age and performance. Available: http://educationforum.ipbhost.com/index.php?showtopic=8063. [30 May 2014].

Ehrenberg, R.G. & Brewer, D.J. 1995. Did teachers' verbal ability and race matter in the 1960s? Coleman revisited. *Economics of Education Review.* 14:1-21.

Ferguson, R. F. 1998. Can schools narrow the Black-White test score gap? In C. Jencks & M. Phillips (Eds). *The Black-White test score gap.* Washington, DC: Brookings Institution. 318-374.

Goldberger, A. S. 1991. *A course in econometrics.* Cambridge: Harvard University Press.

Hanushek, E.A. 1992. The trade-off between child quantity and quality. *Journal of Political Economy.* 100:85-117.

Hanushek, E.A. 2007. The single salary structure and other issues of teacher pay. *Peabody Journal of Education.* 82(4).

Hanushek, E. 2009. "Teacher deselection" in D. Goldhaber and J. Hannaway (Eds). 2009. *Creating a new teaching profession*. Washington DC: The Urban Institute Press.

Hanushek, E. 2011. Valuing teachers. *Education Next.* 11(3):40-45.

Hill, H.C., Loewenberg Ball, D. & Schilling, S.G. 2008. Unpacking pedagogical content knowledge: Conceptualising and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education*. 39(4):372-400.

Luke, D.A. 2004. *Multilevel modeling*. Thousand Oaks, CA: Sage Publications.

Luschei, T.F., & Carnoy, M. 2010. Educational production and the distribution of teachers in Uruguay. *International Journal of Educational Development.* 30(2):169-181.

Marshall, J.H., Chinna, U., Nessay, P., Hok, U.N., Savoeun, V., Tinon, S., & Veasna, M. 2009. Student achievement and education policy in a period of rapid expansion: Assessment data evidence from Cambodia. *International Review of Education.* 55(4):393-413.

Marshall, J.H. & Sorto, A. 2012. The effects of teacher mathematics knowledge and pedagogy on student achievement in rural Guatemala. *International Review of Education.* 58:173-197.

Moloi, M. & Strauss, J. 2005. *The SACMEQ II project in South Africa: A study of the conditions of schooling and the quality of education – South Africa working report*. Available: http://www.sacmeq.org/education-south-africa.htm. [5 May 2014]

National Council of Teachers of Mathematics (NCTM). 2000. *Principles and standards for school mathematics.* Reston: NCTM.

Newman, D., Newman, I. & Salzman, J. 2010. Comparing OLS and HLM models and the questions they answer: Potential concerns for type VI error. *Multiple Linear Regression Viewpoints*. 36(1):1-8.

Patterson, A. & Arends, F. 2008. *Teacher graduate production in South Africa.* Cape Town: HSRC Press.

Raudenbush, S.W. & Bryk, A.S. 2002. *Hierarchical linear models: Applications and data analysis methods.* London: Sage Publications.

Raudenbush, S.W., Eamsukkawat, S., Di-Ibor, I., Kamali, M., & Taoklam, W. 1993. On-the-job improvements in teacher competence: Policy options and their effects on teaching and learning in Thailand. *Educational Evaluation and Policy Analysis.* 15(3):279-297.

Rockoff, J.E. 2004. The impact of individual teachers on student performance: Evidence using panel data. *American Economic Review.* 94(2):247-252.

Spaull, N. 2011. *A preliminary analysis of SACMEQ III South Africa*. Working paper number 1/11. Stellenbosch: Department of Economics, University of Stellenbosch.

Spaull, N. 2011. *Primary school performance in Botswana, Mozambique, Namibia and South Africa: A comparative analysis of SACMEQ III.* SACMEQ III Working paper No. 8. Available: http://www.sacmeq.org/reports.htm#working_papers. [7 April 2013].

Rowan, B., Chiang, F.-S., & Miller, R. J. 1997. Using research on employees' performance to study the effects of teachers on students' achievement. *Sociology of Education*. 70:256-284.

Shepherd, D. 2013. *A question of efficiency: Decomposing South African reading test scores using PIRLS 2006.* Working paper number 20/13. Stellenbosch: Department of Economics, University of Stellenbosch.

Slater, H., Davies, N. & Burgess, S. 2009. *Do teachers matter? Measuring the variation in teacher effectiveness in England*. (Centre for Market and Public Organisation (CMPO) Working Paper Series; number 09/212). Bristol: CMPO.

Summers, A.A., & Wolfe, B.L. 1977. Do schools make a difference? *American Economic Review*. 67:639-652.

Van der Berg, S., Burger, C., Burger, R., De Vos, M., Du Randt, G., Gustafsson, M., Moses, E., Shepherd et al. 2011. *Low quality education as a poverty trap.* Stellenbosch: University of Stellenbosch.

Wayne, J. & Youngs, P. 2003. Teacher characteristics and student gains: A review. *Review of Educational Research*. 73(1):89-122.

Wills, G. 2014. *The effects of teacher strike activity on student learning in South African primary schools.* (University of Stellenbosch Economic working papers; number 01/14). Stellenbosch: University of Stellenbosch.