HOME LANGUAGE AND SOCIO-ECONOMIC PEER EFFECTS IN SOUTH AFRICAN PRIMARY SCHOOLS

Rebecca Joy Selkirk [1] and Debra Lynne Shepherd [2]

Abstract

Since the end of apartheid in South Africa, there has been increased racial, lingual and socio-economic integration in the school system. This presents an opportunity to investigate whether or not home language and socio-economic peer effects exist, and how these peer effects affect the test scores of South African learners. While various peer effects, including socio-economic peer effects, have been studied in South Africa, no studies have investigated home language other than as a stand-alone explanatory variable. Internationally, the literature has been inconclusive as to the existence and directionality of the various peer effects. In this study, data from the 2007 National Systemic Test cycle for Grade 3 learners is used; this data has never been studied before, despite its rich contextual data. Unlike most other school datasets, the National Systemic contains information on both language of instruction and learner home language, which allows for a comprehensive investigation into language effects. Language and socio-economic fractionalisation indices, along with various other language and socio-economic variables, are created or obtained from the various parent, principal and learner responses to contextual questionnaires. Simple OLS estimates indicate a positive relationship between language fractionalisation and learner performance, and a strong negative relationship between socioeconomic fractionalisation and learner performance. Once the language and socioeconomic fractionalisation indices are interacted with each other, it is found that a language "penalty" for learners who do not learn in their home language is eliminated, or greatly reduced, through certain combinations of language and socio-economic fractionalisation. Inclusion of school fixed effects finds support for both positive language and socioeconomic peer effects. The evidence of peer effects found in this study provides motivation for further study into peer effects, and in particular home language peer effects, in South Africa, which up until now have remained largely unexplored.

[1] Economics Department, Stellenbosch University
[2] Economics Department, Stellenbosch University

1. Introduction

South African learners have consistently underachieved on international standardised tests, and as a result, numerous studies have investigated the effect of various inputs (as measured by school, learner and teacher characteristics) on educational outcomes in order to explain this performance. While peer effects have been investigated, albeit to a lesser extent, no study exists in the current South African education literature that incorporates home language other than as a stand-alone explanatory variable. Home language peer effects are therefore unexplored, presenting the opportunity to determine if classroom composition, in terms of home language and socio-economic status, has contributed at all to South African learners' achievements.

As a country with 11 official languages and high socio-economic inequality, ample opportunity exists for very diverse classes to arise in the South African schooling system. During apartheid, schools catered to designated racial groups. Since 1994, however, there has been increased diversification in the racial composition of schools and, by implication, in the home-language and socio-economic composition of classes. Soudien (2004) provides an overview of post-apartheid integration in South African schools, which identifies the flight of learners from former-Black and -homeland schools to English speaking schools, and also highlights the strong migration of Black learners to former-Indian and -Coloured schools. While this points to increased diversification of language, he notes that migration has occurred along socio-economic lines, following an assimilationist position in which social and cultural school contexts are determined by the values, traditions and customs of the dominant group (Soudien, 2004:95). Similarly, Israel's desegregation process of the 1970s saw desegregation in the school system along ethnic lines, resulting in long-term data for study. Studies using this data indicate a positive, but weak, composition effect, where a richer class composition has contributed to higher learner performance (Resh & Dar, 2012); this suggests that South African schools with more diverse classrooms would have higher performance on standardised tests than more homogenous schools.

School composition effects, such as home language and socio-economic peer effects, are important as they have implications for the role of schools. If large compositional effects exist, it implies that school management and practices aren't solely responsible for school performance, which makes the role of schools less clear. An ideal study would incorporate peer group, instructional, organisational, and management processes in its research design, which most studies neglect to do (Thrupp, Lauder & Robinson, 2002). Through the incorporation of peer group characteristics and proxies for instructional, organisational and management processes, this paper aims to fill a gap in the South African education literature by determining whether home language and socio-economic peer effects exist in South African primary schools at a Grade 3 level. Data

gathered in the 2007 National Systemic Tests is used, and the existence of peer effects is investigated through the inclusion of various fractionalisation indices and interaction terms in ordinary least squares (OLS) and fixed effects regressions. It is found that the performance of learners speaking a home language different from the school's language of learning and teaching (LOLT), hereafter referred to as non-LOLT home language learners, is significantly worse than that of their LOLT speaking peers, while the levels of language and socio-economic fractionalisation in schools have significant positive and negative effects on performance, respectively. Interacted together, language and socio-economic diversity is shown to have positive effects at certain levels of fractionalisation. The results obtained in this study support the existence of language and socio-economic peer effects, and provide a basis for further research into language and socio-economic peer effects in South Africa.

The paper begins with a literature review of the existing research that has been done on the different types of peer effects, both in South Africa and internationally. Section 2 provides a discussion of the data used in this study, along with the methodology employed to obtain the empirical results presented in section 3. Section 4 concludes the paper with an overview of the empirical findings, and a discussion of further areas of research.

2. Literature review

Peer effects are influences on an individual's performance derived as "spill over" effects from the characteristics of the individual's peers; there are a number of opportunities for peer effects to arise in a classroom setting. Soudien (1998) finds that minority learners experience 'othering' from their teachers and peers, which illustrates a potential negative peer effect that arises from speaking a different home language than one's peers or being of a different socio-economic status. Conversely, it is possible that a positive peer effect may arise for minority language learners through more frequent use of the language of instruction in a social context, if the majority language is the same as the language of instruction. This positive peer effect, however, may not occur if, as in Pijl, Frostad and Mjaatvan (2011), learners prefer to form groups and interact with similar peers. Wilkinson, Parr, Fung, Hattie and Townsend (2002) posit that different types of verbal interaction have been shown to support different types of learning, and that complex instruction affects group dynamics and willingness to participate in discussions. This type of peer effect is particularly relevant when studying home languages, as it is possible that learners with a minority home language could have their learning hindered by different verbal interactions with their teachers and peers.

A number of studies in South Africa have investigated the effect of various learner, teacher, and school characteristics on learner performance in standardised tests. Some studies (DBE, 2008;

Shepherd, 2011) have found home language to be a factor that influences learner performance; however, language effects are usually not clear as a result of the strong correlations between language, socio-economic status (SES) and school choice (Spaull, 2013; Taylor & Coetzee, 2013). While language peer effects for immigrants have been investigated in countries such as Canada (Friesen & Krauth, 2008) and Chile (McEwan, 2007), no similar studies involving home language peer effects have been conducted in South Africa.

It has been hypothesized (Friesen & Krauth, 2008) that language concentration can positively affect learner performance through better support networks, or negatively affect learner performance if slow language acquisition exists among the group of learners with a different home language to that of the language of instruction or of the majority group. However, studies investigating this have been inconclusive; McEwan (2007) finds that while it is possible that within-school segregation decreases learner performance, it is unlikely, and Friesen and Krauth (2008) find that peer achievement has more important effects on learner performance than home language when measuring the peer effects for English second language (ESL) speakers across multiple cohorts and using school-by-grade fixed effects. Gabriel, Lilla, Zander and Hannover (2014) study immigrant learners' language usage and how this develops learners' self-perceptions, to show that the use of different languages within an academic setting has a negative impact on test scores, even though bi- or multilingualism has a positive effect on learners' performance of executive control processes. They also find that there appears to be some importance as to where the second language is learned, as it is theorised that individuals change their conduct in accordance with the culture that is associated with the language being spoken.

While language peer effects have been relatively unexplored, there exists a large literature on the existence and directionality of peer effects associated with other characteristics, such as ability, race, gender and socio-economic status (SES) (Contini, 2013). As with the existing literature on language peer effects, the direction and size of non-language peer effects have not been conclusively determined, although there are some effects which are found to have a significant impact on academic achievement. Low SES and minority learners have both been found to perform more poorly than their peers in standardised assessments, and peer effects are generally found to be negative in classrooms with high proportions of minority or low-SES learners (Luyten, Schildkamp & Folmer, 2009). However, Contini (2013) finds that in fifth and sixth grade Italian classrooms, the proportion of immigrants has no significant effect on learner achievement. In terms of ethnicity, Hall and Leeson (2010), using an "overall performance" effect, find that racial diversity is negatively associated with performance, where racial diversity is measured by a racial fractionalisation formula. By investigating the mathematics achievement scores for Belgian learners

at the end of second grade, Boonen, Speybroeck, de Bilde, Lamote, Van Damme and Onghena,(2014) found no direct school composition effects, but small differential effects with regards to high achievers' prior achievement. They also found, consistent with other studies, a negative effect for non-European language speakers in classes with a high proportion of minority learners. Vandenberghe (2002) states that previous literature agrees that academic peer effects do exist at a primary and early secondary education level, as classes with a higher proportion of high-achieving pupils leads to an increase in everyone's achievement. He substantiates this with his own quantitative study, and also investigates quadratic peer effects. Through this he finds that, in some cases, high performing or high-SES learners are less sensitive to peer effects.

Wilkinson et al (2002) propose that compositional effects are probabilistic in nature as a result of their small magnitude and variability across schools, which confounds identification. Additionally, compositional and peer effects at school level are difficult to identify as they are a combination of direct and indirect effects. School mix can also be changed through the flight of learners as a result of parents' perceived classroom-level effects. In their study, they obtain results that are inconsistent with those of descriptive studies, which they attribute to the fact that most outcomes-based research studies a higher level of aggregation than that at which peer effects take place, or that often peer effects are not modelled in theoretically appropriate ways. They conclude that in order to promote peer effects, it is not sufficient to solely alter class composition – higher expectations of learners' ability, along with improved curricular specifications and teacher quality, are needed for this. This is supported by Luyten, Schildkamp and Folmer (2009), who hypothesize that changes in instructional methods can arise from changes in how teachers perceive the learners' ability – these changes in perceived ability can result from classroom composition if minority learners are perceived to be weaker than their peers. In addition to this, they hypothesize that classroom composition may also affect academic self-perceptions which in turn affects academic performance. In their own study using data from the Netherlands, they find that classroom composition has a significant negative effect on spelling performance when considering the proportion of low-SES minority learners in a class, but from Grade 3 this effect tapered quickly until, by Grade 7, no achievement gap remained. Guldemond and Bosker (2005) also find this decreasing achievement gap between low-SES minority learners and their peers, but they found no significant composition effects. Hattie (2002) explores the peer effects that result from tracking of learners into classes according to ability, gender or ethnicity, as well as peer effects arising from class size and multi-age classes. He concludes, similarly to Luyten et al (2009), that there is very little evidence to suggest that peer effects arise from any form of tracking, and it is rather changes in instructional methods that have the potential to improve learner performance. As instructional methods do not appear to, in practice, vary according to different tracking methods, potential

positive peer effects are negated.

There are a variety of other peer effects that have been hypothesised to affect performance. A learner's peer group's average level of motivation in terms of both academic motivation and class engagement, for example, may affect achievement, as learners often mimic the behaviour of their peers (Masland and Lease, 2013). Von Suchodoletz, Larsen, Gunzenhauser and Fäsche (2015) examine the effects of class-average intelligence and class-average self-control on German 3rd grade learners' performance on spelling and reading tests. They find that the average intelligence level of the class has a significant positive effect on spelling, but no effect on reading, while the average self-control level of the class has no main effect, but could have a classroom interaction effect. Marks (2002) fails to find composition and peer effects in high schools according to the school organisational setting, possibly as a result of the high overlap between different types of organisational settings, while the role of expectations from parents, peers and oneself on educational attainment was investigated by Sommerfeld (2015), all of which were found to be significant. As higher achievers are more likely to attain higher levels of education, it could therefore be hypothesised that expectations can have an effect on academic achievement.

There is some debate as to, if peer effects exist, why school composition creates these effects (Boonen, et al., 2014), and thus far no significant effect has been found between social identity, or the feeling of belonging to one's peer group, and academic conformity (Masland and Lease, 2013). Some academics are of the opinion that, as found by Lee and Croninger (1994), home and school supports account for a far greater share of learner achievement than peer effects do, and therefore peer effects are seen to be of little importance. Vardardottir (2015), however, proposes that there may be asymmetric peer effects that occur, which could explain the lack of significant peer effects being found at an aggregate level. In his own study, he finds that average peer effects are insignificant for both reading and science, but significant for mathematics and problem solving, and also that class heterogeneity created a positive effect on mathematics performance.

2. Data and methodological approach

2.1 National Systemic dataset

Data from the 2007 cycle of the National Systemic Tests is used. These tests were conducted in one of South Africa's 11 official languages, according to the school specific LOLT, on a random and nationally representative sample of Grade 3 South African learners attending mainstream public schools. Literacy and numeracy levels in respect of grade appropriate curriculum outcomes were measured using test instruments (DBE, 2008:1). In addition to the test scores, contextual questionnaires captured information from parents, principals, teachers, and the learners themselves. Owing to problems in matching the datasets,[3] however, only parent, learner and principal responses are contained in the final dataset used in this study.

The original sample consisted of 54 704 observations from 2 376 schools. Of this sample, 2 188 and 587 observations were missing learner home language and LOLT, respectively. These observations were dropped from the sample, along with 1 565 observations from 67 dual medium schools.[4] Dual medium schools with English and Afrikaans as the LOLT were retained in the sample, as these schools' test performances differed from the rest of the sample. After dropping the aforementioned observations, a sample of 50 550 observations from 2 198 schools remained; this represents approximately 93 percent of the original sample. Literacy and numeracy scores, as measured on a Rasch scale[5], were standardised to have a mean of zero and a standard deviation of one. Figure 1 shows the distributions of these standardised numeracy and literacy scores for the entire sample, compared to the standard normal distribution. There does not appear to be much difference between the distributions of the numeracy and the literacy scores, nor does either differ significantly from the standard normal distribution. For the purposes of this study, only standardised numeracy scores are examined, and where the term "performance" is used, it refers to numeracy test scores.

In South Africa, schools are able to teach in any of the country's 11 official languages in the initial three years of schooling (foundation phase), with most schools switching to English as the main language of instruction by Grade 4 (Draper & Spaull, 2015:4). This allows for a diverse set of

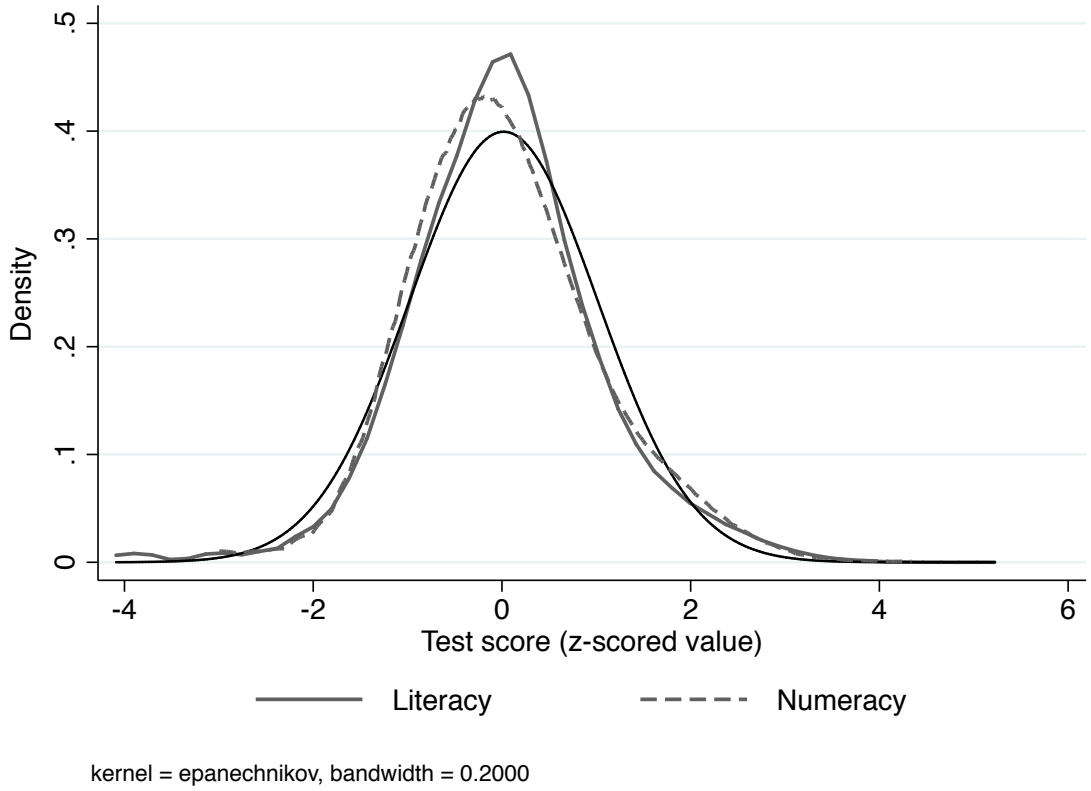
[3] Not all subject teachers were questioned, and therefore not all students could not be linked to their respective mathematics and reading teacher.

[4] These schools were observed to teach in a combination of two or three African languages; this made it very difficult to assess the language fractionalisation of these schools. The test performance of learners attending these schools did not differ significantly from the rest of the sample, and therefore the decision to drop them from the sample is not likely to create bias in the results.

[5] Rasch scaling uses raw learner responses to create a rescaled test score measure that accounts for question difficulty. Using this scaling, two learners who received the same raw test score will receive different Rasch scores if one learner answered very difficult questions correctly and the other learner answered easier questions correctly, with the former learner receiving a higher Rasch score.

LOLT and home language combinations, and language-related variables can be created for a number of different categorisations, such as having a different home language from one's peers or having a different home language from the school's LOLT, indicated by the dummy variable "difflang", which takes on a value of one if home language and LOLT are different, and zero otherwise. Where parent and learner-reported home language did not agree across the contextual questionnaires, the parent-reported home language was assumed to be correct. Figure 2 provides an indication of how many learners have a home language that is different from the school LOLT. While close to seventy percent of schools have fewer than 20 percent of their learners with a different home language from the LOLT, nearly 6 percent of schools have no learners who speak the LOLT at home.

Figure 1: Standardised numeracy and literacy performance distribution

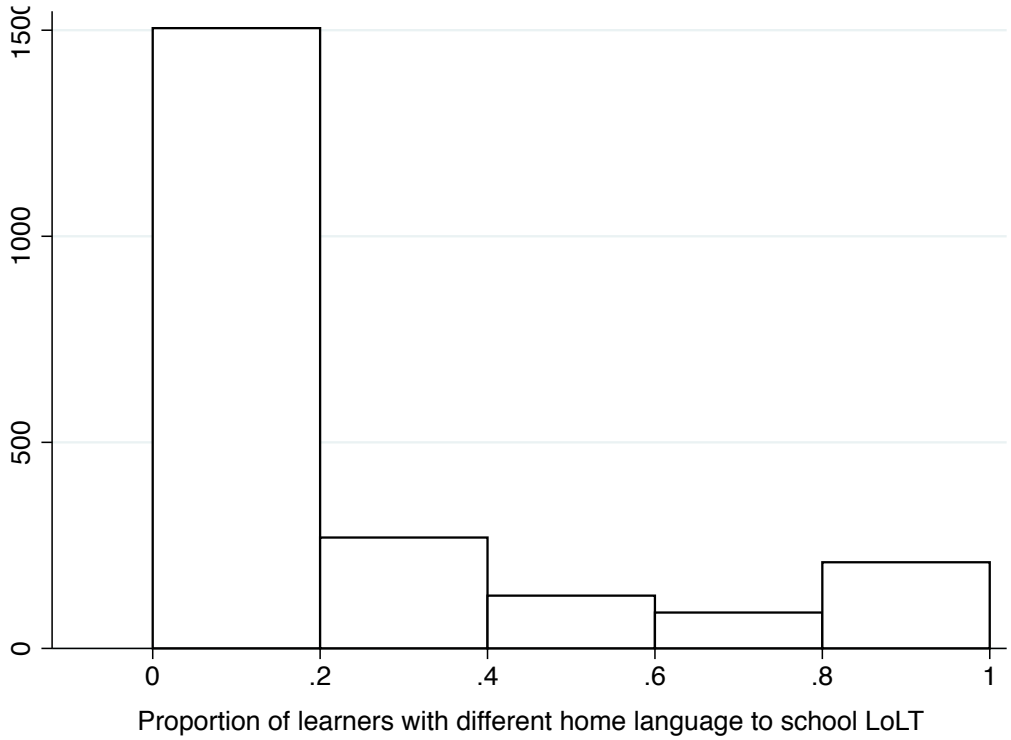


Note: own calculations using National Systemic (2007). Standard normal distribution indicated by the solid black line.

Higher SES learners are more likely to be in schools where a high proportion of the learners have different home languages from the LOLT; English LOLT schools had, on average, 72 percent of learners speaking a home language other than English. This is largely driven by the 118 out of 324 English schools in the sample that had 100 percent of their learners with a home language other than English. Even after excluding these schools, however, 56 percent of learners in English schools had a different home language from the LOLT, indicating a strong preference for English

instruction, regardless of home language. isiNdebele speaking learners were most likely to learn in a different language, with 73 percent of isiNdebele home language speakers attending a school with a LOLT other than isiNdebele. It must be noted, however, that very few isiNdebele schools existed in the sample, which could result in this large proportion of non-LOLT isiNdebele home language learners.

Figure 2: Distribution of non-LOLT home language learners by school



Note: own calculations using National Systemic (2007)

2.2. Identification strategy

2.2.1 Constructs of fractionalisation within a school

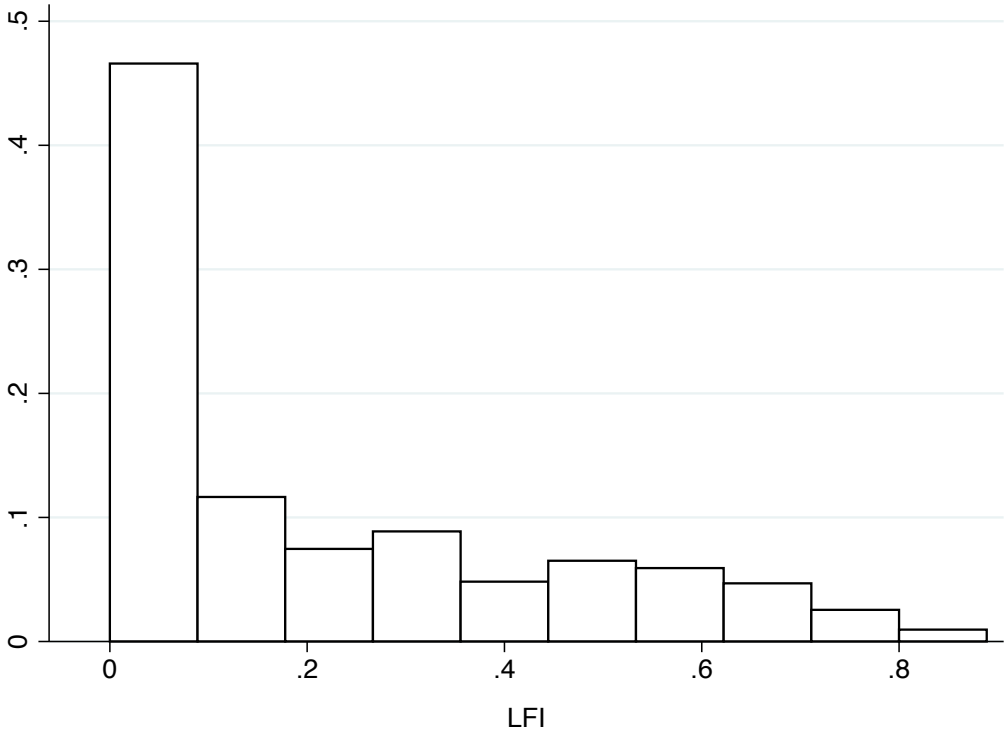
In order to examine peer effects through fractionalisation, a language fractionalisation index (LFI) score for each school was calculated. This index is similar to the racial fractionalisation index constructed by Hall and Leeson (2010:4), and uses the formula:

$$LFI = 1 - \sum_i HL_i^2, \quad [1]$$

where HL_i is the percentage of the school's learners that speak that particular home language. An LFI score of 0 represents a language homogenous school (that is, all learners speak the same language, regardless of the school's LOLT), while a score of 0.5 could indicate that the school's learners were equally divided into two home language groups. A higher score therefore indicates a

more fractionalised school in terms of learner home language. Figure 3 below illustrates that the majority of South African schools are quite language homogenous, with less than 20 percent of schools possessing an LFI score greater than 0.5.

Figure 3: Distribution of LFI scores by school



Note: own calculations using National Systemic (2007)

Poorer schools, according to governmental quintile classification (where quintile 1 represents the poorest schools and quintile 5 the wealthiest)[6] are more likely to have an LFI score of zero (complete homogeneity) than wealthier schools, with 42.1 percent of quintile 1 schools having an LFI of 0 compared to 12.3 percent of quintile 5 schools (see table 1). Conversely, only 8.2 percent of quintile 1 schools had an LFI of over 0.5, while 40.1 percent of quintile 5 schools had LFIs over 0.5. In order to identify potential non-linear effects of language fractionalisation in the regressions, schools were divided into four LFI groups:

$$LFIQ = \begin{cases} 0, & \text{LFI} = 0 \\ 1, & 0 < \text{LFI} \leq 0.2 \\ 2, & 0.2 < \text{LFI} \leq 0.5 \\ 3, & \text{LFI} > 0.5 \end{cases} \quad [2]$$

[6] It is relevant to note that school poverty quintiles are not calculated according to the conventional method whereby 20 percent of the sample is included in each quintile.

Table 1: Proportion of schools by language fractionalisation index within school quintile

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Other/ Unknown
LFI = 0	39.1	32.4	32.4	22.1	9.7	30.1
LFI = 1	25.5	29.7	22.2	22.1	14.0	19.0
LFI = 2	25.2	21.6	27.6	26.8	33.0	20.2
LFI = 3	10.3	16.3	17.7	29.1	43.3	30.7

Note: own calculations using National Systemic (2007)

In order to incorporate SES into the model, it was necessary to create a variable to represent learner and school wealth, as no explicit indication of either was provided in the dataset. Multiple component analysis (MCA) was therefore used to construct an asset wealth index to be used as an indication of learner SES.[7] School SES is then taken to be the average of its learners' SES scores, and both SES variables were standardised to have a mean of zero and a standard deviation of one.[8] Figure 4 illustrates the distribution of the generated learner SES values; this shape appears to relate well to the bimodal distribution of wealth in South Africa, with a large number of low-SES learners, and a smaller group of high-SES learners. As illustrated in figure 5, learner SES appears to matter only for the richest learners – quintile 1 to 4 learners have almost identical numeracy score distributions, while quintile 5 learners have a higher mean score, and a higher proportion of high achieving learners. It should be noted that here, learners are divided into quintiles according to their individual wealth, and not that of their school.

In order to include socio-economic fractionalisation in the model, a socio-economic fractionalisation index (SFI) was constructed by dividing learners into SES quintiles at a national level and subsequently using the formula:

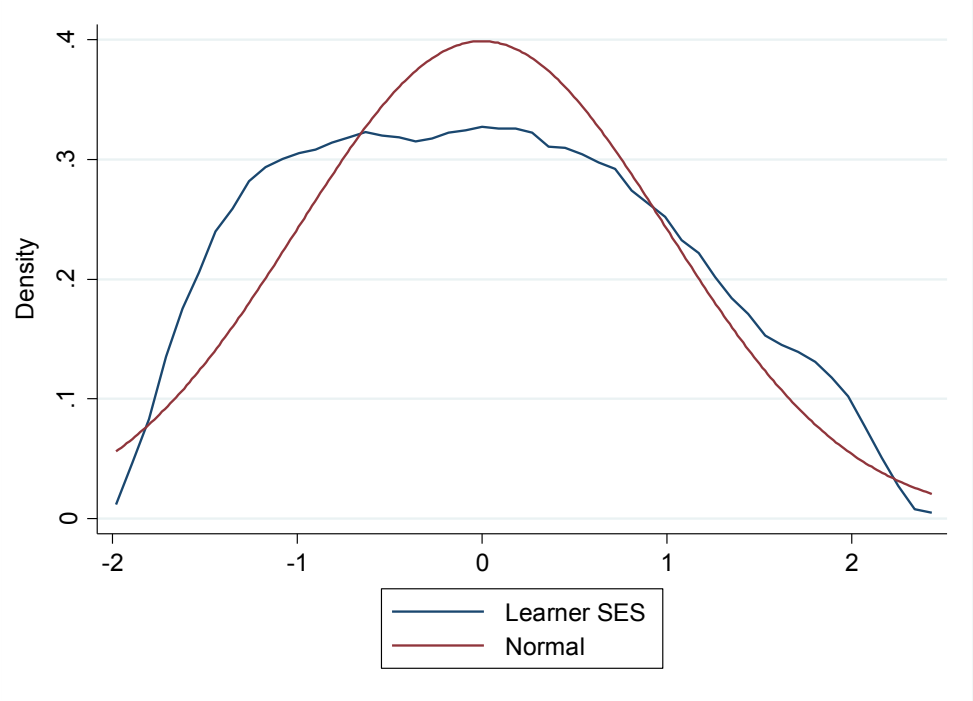
$$SFI = 1 - \sum_i SES_i^2, \qquad [3]$$

where SES_i represents the proportion of learners in the school in SES quintile i. The interpretation is identical to that of the LFI, but fractionalisation in this case refers to SES fractionalisation. A disadvantage of this formulation, which does not occur in the LFI, is that the SFI does not account for fractionalisation within learner SES quintiles; a school, for example, with an SFI of 0 is, by the above definition, completely socio-economically homogenous, but in reality there will be variation of learner SES within the school, even though all learners come from the same SES quintile according to national learner SES.

[7] As with home language, parent responses were used where parent and learner responses differed. In the case of missing responses to questions asking about the presence of a particular asset, this non-response was treated the same as if the respondent had answered that the asset in question was absent.

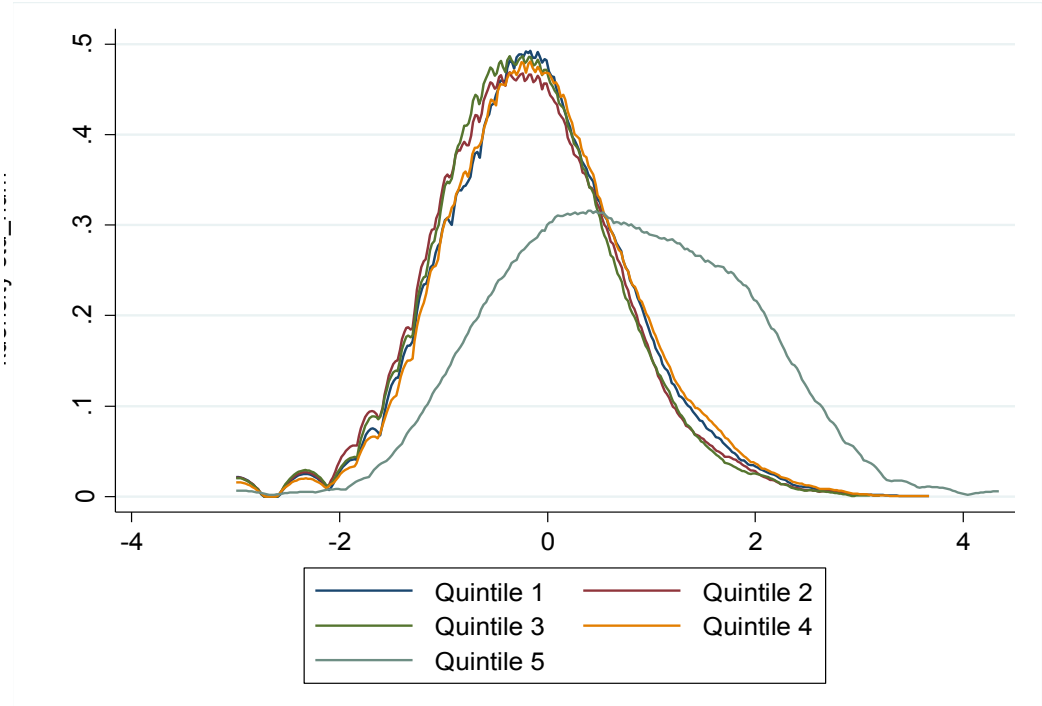
[8] While neither SES variable is completely accurate, learner SES correlated as expected with the household's main income source.

Figure 4: Distribution of learner SES



Note: own calculations using National Systemic (2007)

Figure 5: Distribution of numeracy performance by learner SES quintile



Note: own calculations using National Systemic (2007)

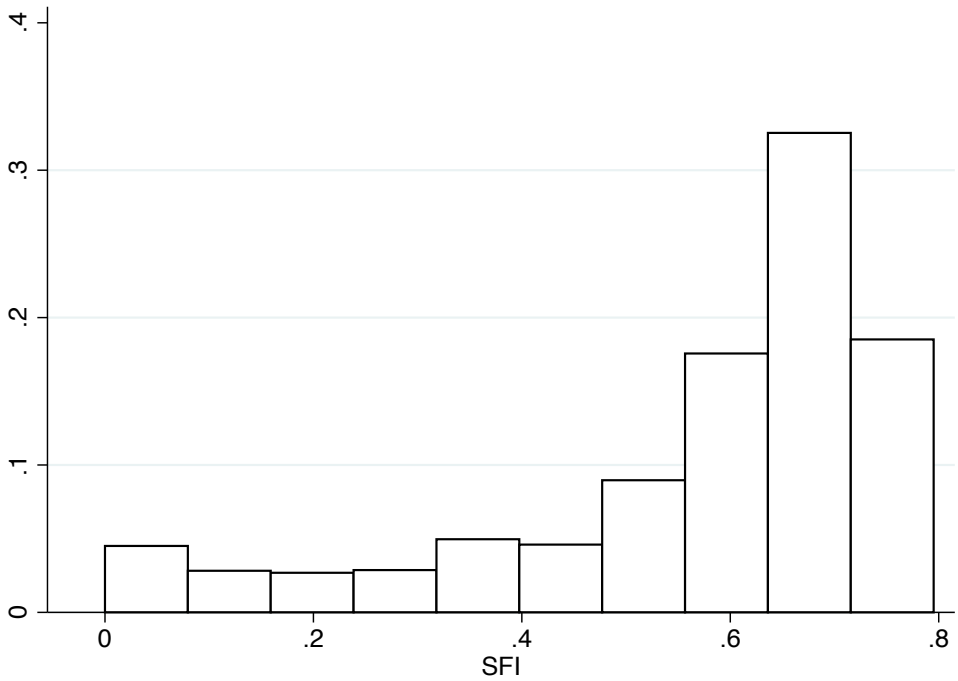
Figure 6 below shows the distribution of SFI. Unlike the LFI, the majority of schools have high SFI scores, indicating relatively high SES fractionalisation, with only 20 percent of schools possessing a SFI of less than 0.4, of which 55 schools (2.5 percent of the sampled schools) are completely homogenous with an SFI of 0. Quintile 5 schools were the most likely to be mostly homogenous (i.e. have a SFI of less than 0.4), with two-thirds of quintile 5 schools falling into this category, and only 7.5 percent of quintile 5 schools falling into the highest SFI grouping, compared to the average of 27.3 percent (see table 2). Quintile 1 schools were also slightly more likely to be homogenous and less likely to be highly fractionalised than average, but these deviations from the mean were not as large as for the quintile 5 schools. A total of 80 percent of completely SES homogenous schools were found in quintiles 1 and 5 (38.2 and 41.8 percent respectively).

Table 2: Proportion of schools by SES fractionalisation index within school quintile

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Other/ Unknown	All
SFI = 0	18.1	8.3	10.7	12.0	66.6	15.3	20.0
SFI = 1	22.8	15.5	13.2	19.8	11.5	21.1	17.0
SFI = 2	36.8	38.4	40.8	43.0	14.4	31.2	35.8
SFI = 3	22.3	37.8	35.3	25.3	7.5	34.5	27.3

Note: own calculations using National Systemic (2007)

Figure 6: Distribution of SFI scores by school



Note: own calculations using National Systemic (2007)

As with the LFI, schools were divided into different groups by SFI for use in the regressions. The

variable SFIQ took on the values of:

$$SFIQ = \begin{cases} 0, & \text{SFI} \leq 0.4 \\ 1, & 0.4 < \text{SFI} \leq 0.6 \\ 2, & 0.6 < \text{SFI} \leq 0.7 \\ 3, & \text{SFI} > 0.7 \end{cases} \quad [4]$$

However, the apparent lack in overlap (common support) across school SES quintiles suggests a significant amount of non-random selection into schools according to own learner SES and the SES of the school. Quintile 5 schools are relatively more homogenous due to the fact that they are able to charge high school fees that are only accessible by a wealthier cohort of learners. In the regression analysis that follows, this lack of common support and school selection is likely to bias the results on the SFI variables. Therefore, we control for the SES fractionalisation of the school uisng the standard deviation of learner SES in a given school, with the variables SDQ taking value as follows:

$$SDQ = \begin{cases} 0, & \text{SFI} < 0.4 \\ 1, & 0.4 \leq \text{SFI} < 0.6 \\ 2, & \text{SFI} \geq 0.6 \end{cases} \quad [5]$$

2.2.1 Methodological approach

The main regression equation takes an education production function type form, as follows:

$$y_{is} = \delta \boldsymbol{f}_{l,s} + \beta \boldsymbol{d}_i + \gamma' \boldsymbol{x}_i + \alpha' \boldsymbol{z} + \varepsilon_{is} \quad [6]$$

In this equation, y represents learner test score, $\boldsymbol{f}$ is a vector of school fractionalisation index dummies, $\boldsymbol{d}$ is a dummy variable indicating that the learner speaks a home language that is different to the school LoLt, $\boldsymbol{x}$ is a vector of learner, school, and environmental characteristics, $\boldsymbol{z}$ is a vector of language dummies, and ε is a random error term, assumed to have a zero mean. Subscripts refer to learner i in school s. Interaction effects are introduced to the model through interactions between the $\boldsymbol{f}$ and $\boldsymbol{d}$, as well as between the indices themselves as follows:

$$y_{is} = \delta_1 \boldsymbol{lfi}_{l,s} + \delta_2 \boldsymbol{sd}_{l,s} + \beta \boldsymbol{d}_i + \theta_1 \boldsymbol{lfi}_i * \boldsymbol{d}_i + \theta_2 \boldsymbol{sd}_i * \boldsymbol{d}_i + \gamma\theta_3 \boldsymbol{sd}_i * \boldsymbol{lfi}_i + {}' \boldsymbol{x}_i + \alpha' \boldsymbol{z} + \varepsilon_{is} \quad [7]$$

where ***lfi*** and ***sd*** are categories of fractionalisation the language and socio-economic. The coefficients θ_1, θ_2 and θ_3 enable the investigation of how speaking a home language that is different to the LoLT interacts with language and socio-economic diversity within a school, and how different forms of fractionalisation work together to impact learners' test performance. For all regressions, the sample is limited to schools with 15 or more learners; this is to remove misleading SFI and LFI scores that occur in schools with very few learners. Once this restriction is

implemented and all controls are added to the model, a total of 43 946 observations remain in the regression sample. All regressions control for school clustered standard errors.

3. Empirical results

The results of basic OLS regressions as in equation [6] are presented in table 1, with only the relevant language and socio-economic estimation coefficients shown. In model 1, regressions of the LFIQ groups on the standardised numeracy scores indicates a non-linear relationship between LFI and performance, with highly fractionalised (i.e. language diverse LFIQ3) schools having the best performance and mostly homogeneous schools (LFIQ1) the worst performance, relative to homogenous schools (LFIQ0). Similarly, model 2 indicates a large, negative, non-linear effect of SFI on numeracy scores, suggesting that schools with learners from diverse socio-economic backgrounds perform more poorly than schools with learners from similar socio-economic circumstances, where mostly homogenous schools (SFIQ0) are the reference group. When both ***sd*** and ***lfi*** are controlled for, as in model 3, the coefficients on SES fractionalisation remains the same as before, whereas language fractionalisation has a much smaller, but still positive, effect for highly fractionalised schools. Controlling for the spread of socio-economic status within the school, only schools that experience a high level of language fractionalisation are found to perform significnatly better than more language homogenous school contexts.

The results of models 1-3 are likely to be influenced by the lack of other controls. From the previous section, it is evident that the relative distribution of language and socio-economic fracitonalisation differs across school wealth contexts. For exmaple, quintile 5 schools ar emore likely to be language diverse but socioeconomically homogenous. Model 4 therefore adds various learner, school, classroom, home and geographical controls, which removes the significance of the language fractionalisation indicators and approximately halves the size of the coefficients on socio-economic fractionalisation. Socio-economic fractionalisation continues to have large negative effects, with learners attending highly fractionalised schools performing approximately 40 percent of a standard deviation worse than learners attending mostly homogenous schools. The relatively small (but highly significant) coefficients on the standardised learner SES and standardised school SES indicate that the negative effect of high SES fractionalisation cannot be removed through either own or school SES alone.

Table 1: Regression results on standardised numeracy scores

	Model			
	(1)	(2)	(3)	(4)
	dependent variable: numeracy			
LFIQ = 1	-0.035***		-0.008	0.036
	(0.011)		(0.041)	(0.034)
LFIQ = 2	0.042***		0.043	0.030
	(0.012)		(0.044)	(0.035)
LFIQ = 3	0.250***		0.248***	0.070
	(0.014)		(0.051)	(0.043)
SDQ = 1		-0.457***	-0.453***	-0.222***
		(0.071)	(0.070)	(0.047)
SDQ = 2		-0.705***	-0.698***	-0.379***
		(0.069)	(0.068)	(0.015)
Different language				-0.221***
				(0.019)
Learner SES				0.019***
				(0.007)
School SES				0.157***
				(0.025)
Learner controls	N	N	N	Y
School controls	N	N	N	Y
Classroom controls	N	N	N	Y
Home controls	N	N	N	Y
Geographical controls	N	N	N	Y
Observations	48 185	48 185	48 185	43 946
R-squared	0.010	0.048	0.057	0.340

Notes: Robust standard errors in parentheses.*** p<0.01, ** p<0.05, * p<0.1

As with table 1, table 2 reports only the relevant language and socio-economic coefficient estimates. Here we introduce interaction effects between language fractionalisation, socioeconomic fractionalisation and the language of the student as in model [7]. Columns (1) – (3) and (4) – (6) indicate coefficients estimated from models excluding and including school fixed effects, respectively. The inclusion of school fixed effects removes potential bias in the coefficients driven specifically by school (quality) unobservables. Comparing the results of columns (1) – (3), it is evident that socioeconomic and language diversity in schools is related. Whilst language diversity is not estimated to be signifiantly related to numeracy scores (for both LoLT home-language and non-LoLT home-language speakers), socioeconomic diversity is negatively and significantly related with performance (with, again, no difference in the relationship for LoLT and non-LoLT speaking students). Once interactions between socioeconomic and language diversity is introduced to the model, it becomes evident that any level of socio-economic fractionalisation has a negative relationship with performance compared to a mostly socio-economically homogenous school, with performance generally decreasing as the level of socio-economic fractionalisation increases. The

effect of language fractionalisation, however, appears to depend on the level of socio-economic fractionalisation.

Table 2: Regression results on standardised numeracy scores

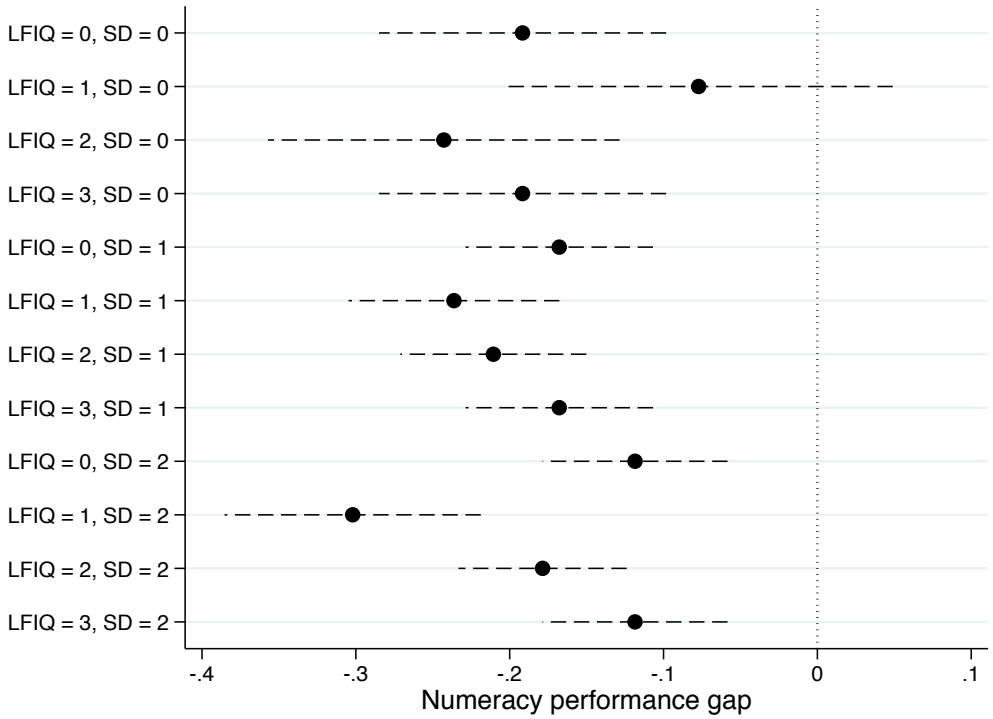
	(1)	(2)	(3)	(4)	(5)	(6)
	Dependent variable: numeracy score					
LFIQ1	0.039	0.036	0.222*			
	(0.035)	(0.034)	(0.117)			
LFIQ2	0.034	0.030	0.175			
	(0.036)	(0.035)	(0.115)			
LFIQ3	0.007	0.070	0.160			
	(0.046)	(0.043)	(0.120			
SD1	-0.217***	-0.221***	-0.090			
	(0.047)	(0.050)	(0.077)			
SD2	-0.373***	-0.382***	-0.286***			
	(0.048)	(0.051)	(0.074)			
difflang	-0.295**	-0.227***	-0.109***	-0.153***	-0.201***	-0.192***
	(0.128)	(0.078)	(0.040)	(0.021)	(0.036)	(0.048)
difflang×SD1		0.000	-0.077		-0.003	0.024
		(0.085)	(0.075)		(0.041)	(0.056)
difflang×SD2		0.014	-0.004		0.023	0.073
		(0.087)	(0.048)		(0.041)	(0.056)
difflang×LFIQ1	0.012		-0.190	-0.093***		0.114
	(0.130)		(0.131)	(0.035)		(0.081)
difflang×LFIQ2	0.040		-0.165***	-0.050*		-0.051
	(0.130)		(0.052)	(0.029)		(0.074)
difflang×LFIQ3	0.175		-0.133***			
	(0.130)		(0.045)			
SD1×LFIQ1[a]			-0.225*			-0.183**
			(0.129)			(0.091)
SD1×LFIQ2 [a]			-0.214*			0.008
			(0.124)			(0.089)
SD1×LFIQ3 [a]			-0.186			
			(0.125)			
SD2×LFIQ1 [a]			-0.185			-0.298***
			(0.125)			(0.095)
SD2×LFIQ2 [a]			-0.110			0.009
			(0.123)			(0.085)
SD2×LFIQ3 [a]			-0.168			
			(0.128)			
Learner controls	Y	Y	Y	Y	Y	Y
School controls	Y	Y	Y	N	N	N
Classroom controls	Y	Y	Y	Y	Y	Y
Home controls	Y	Y	Y	Y	Y	Y
School fixed effects	N	N	N	Y	Y	Y
Observations	43,946	43,946	43,946	43,946	43,946	43,946
R-squared	0.340	0.340	0.342	0.086	0.121	0.114

[a] In order to save space in this table, these coefficients are calculated from the interaction between difflang, LFIQ and SD.

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

However, as the pattern of diversity, particuarl socioeconomic diversity, does differ across school wealth quintiles, the results may be driven by student selection into sspecific schools as well as school unobseravbles. In order to better investiage the magnitude of the difference in performance between home language LoLT learners and learners speaking a different home language, conditional the language and socioeconomic diversity of the school, we construct a graphical representation of the partial effects (with 95% confidence intervals) that are derived by taking a linear combination of the coefficients from the variables of interest both in levels and as interactions (figure 7). These coefficients are interpreted as the *expected performance gap* between students who speak the LoLT of the school as a home-language and students that do not speak the LoLT of the school as a home-language, for a given level of socioeconomic and linguistic diversity at school. It is immediately evident from figure 7 that, in general, higher language diversity has a positive influence on test scores for those that are not LOLT home language speakers (positive peer language effects); comparing across schools with similar socioeconomic diversity, the general trend is for the gap to narrow with increased language diversity. There is also some indication that socioeconomic diversity may play a small positive role on the test gap between LoLT and non-LoLT speakers, at least within the contexts of higher language diversity (***lfi*** = 2 and ***lfi*** = 3).

Figure 6: Distribution of SFI scores by school



It needs to be kept in mind that these gaps are computed after accounting for differences in school quality and other unobservable resources. Table 3 summarises the average school fixed

effect for each ***lfi*** and ***sd*** combination. The average school performance (as measured by the school fixed effect) is positively, and non-linearly, related to the socioeconomic diversity of the school, irrespective of the level of language fractionalisation. The link between average school performance and the language fractionalisation of a school, however, depends on the socioeconomic diversity of the school; only in the case of socioeconomically homogenous schools, language diversity and average school performance are positively related.

Table 3: School fixed effects by socioeconomic and langauge fractionalisation of the school

		Classification of learner SES fractionalisation (***sd***)			
		0	1	2	All
Classificaiton of learner language fractionalisation (***lfi***)	0	0.196	0.027	-0.134	-0.018
		(0.092)	(0.043)	(0.036)	(0.028)
	1	0.406	-0.063	-0.304	-0.126
		(0.125)	(0.048)	(0.037)	(0.031)
	2	0.562	0.002	-0.175	0.004
		(0.128)	(0.048)	(0.042)	(0.034)
	3	0.937	0.317	-0.063	0.251
		(0.108)	(0.057)	(0.054)	(0.041)
	All	0.477	0.048	-0.182	0.004
		(0.059)	(0.024)	(0.021)	(0.017)

Notes: own calculations

4. Conclusion

This paper employs empirical analysis to identify the effect of various language and socio-economic factors on the numeracy performance of Grade 3 learners in South Africa, with a focus on the effects of different levels of language and socio-economic fractionalisation on performance. Various language and socio-economic variables were created from learner, principal and parent responses, and fractionalisation indices were constructed in order to examine the aforementioned effects. OLS regression results show a large home language differential effect, with non-LOLT home language speakers performing significantly worse than otherwise identical peers whose home language corresponded to the school's LOLT. This finding provides an argument for home language instruction, and could motivate extending the number of years in which learners learn in their home language, in order to improve South Africa's numeracy outcomes.

By introducing fractionalisation indices, along with the other language and socio-economic variables and a number of interaction effects into an OLS regression, support is found for the existence of peer effects regarding language fractionalisation and socio-economic fractionalisation. Whilst a performance deficit remains for non-LoLT speakers when accounting for school fixed effects, the interaction between language and socio-economic fractionalisation is partly able to eliminate this "non-LOLT home language penalty". This indicates that it may be possible, through

higher levels of language and socio-economic fractionalisation, to minimise the negative effects of non-home language instruction without incurring a cost for those already learning in their home language.

Although the results found in this paper are encouraging, the empirics are not infallible. The socio-economic variables were created using multiple component analysis of asset ownership and living conditions, and are therefore not an entirely accurate reflection of learner and school wealth, resulting in possible biases in the socio-economic fractionalisation variables. As is the case with many school survey data, missing responses and discrepancies between parent, school, and learner responses limited the sample size, thereby introducing potential sample bias, compounded by omitted variable and selection biases through the omission of important confounding variables not available in the dataset, as well as non-random selection of observations into the sample. To control for these issues, school fixed effects were accounted for by the model; however, this may not have solved the problem of omitted variable bias completely.

There are a number of possibilities for further research in order to verify the peer effects found in this paper. Further studies may want to experiment with a number of differently constructed measures of socio-economic fractionalisation in order to determine if the interaction of socio-economic status with language is dependent on how socio-economic fractionalisation is defined. Additionally, it would be useful for more accurate measures of socio-economic status to be used, if such data can be obtained or estimated accurately. There is also an opportunity for investigation into different measures of fractionalisation and minority status, in terms of both language and socio-economic status, and the definition of fractionalisation can be expanded to include other fractionalisation measures such as gender and racial fractionalisation (if so desired). Language and socio-economic fractionalisation measures that are invariant across learners could also be explored.

As no other South African studies have focussed on home language peer effects, it is impossible to compare this paper's findings to existing South African literature. The finding of significant peer effects warrants further investigation into the effect of home language and socio-economic fractionalisation in South Africa. Through the aforementioned suggestions for extension of this research, the possibility exists for greater insight to be gained of the working of peer effects in South Africa, which could aid educational economists in their understanding of South Africa's educational outcomes and social mechanisms.

REFERENCE LIST

Boonen, T., Speybroeck, S., de Bilde, J., Lamote, C., Van Damme, J. & Onghena, P. 2014. Does it matter who your schoolmates are? An investigation of the association between school composition, school processes and mathematics achievement in the early years of primary education. *British Educational Research Journal,* 40(3): 441-466.

Contini, D. 2013. Immigrant background peer effects in Italian schools. *Social Science Research,* 42(2013): 1122-1142.

DBE. 2008. *2007 Grad 3 SE Leaflet* [Online]. Available: http://www.education.gov.za/Portals/0/DoE%20Branches/GET/System%20Evaluation/2007%20Grade%203%20SE%20Leaflet,%2006%20Nov%202008.pdf?ver=2009-09-10-094455-567 [2016, May 18].

Draper, K. & Spaull, N. 2015. *Examining oral reading fluency among grade 5 rural English Second Language (ESL) learners in South Africa: An analysis of NEEDU 2013*. Stellenbosch Economic Working Papers, WP09/15.

Friesen, J. & Krauth, B. 2008. *Enclaves, peer effects and student learning outcomes in British Columbia*. Metropolis British Columbia: Centre of Excellence for Research on Immigration and Diversity.

Gabriel, U., Lilla, N., Zander, L. & Hannover, B. 2014. How immigrant students' self-views at school relate to different patterns of first and second language use. *Social Psychology of Education,* 17(2014): 617-636.

Guldemond, H. & Bosker, R.J. 2005. Ontwikkeling in de groei van achterstanden? Op zoek naar die rol van de schools (concept). Groningen, the Netherlands: GION.

Hall, J.C. & Leeson, P.T. 2010. Racial Fractionalisation and School Performance. *American Journal of Economics and Sociology,* 69(2)

Hattie, J.A.C. 2002. Classroom composition and peer effects. *International Journal of Educational Research*, 37(2002): 449-481.

Lee, V.E. & Croninger, R.G. 1994. The relative importance of home and school in the development of literacy skills for middle-grade students. *American Journal of Education,* 102: 286-329.

Luyten, H., Schildkamp, K. & Folmer, E. 2009. Cognitive development in Dutch primary education, the impact of individual background and classroom composition. *Educational Research and Evaluation,* 15(3): 265-283.

Marks, H.M. 2002. School composition and peer effects in distinctive organizational settings. *International Journal of Educational Research,* 37(2002): 505-519.

Masland, L.C. & Lease, A.M. 2013. Effects of achievement motivation, social identity, and peer group norms on academic conformity. *Social Psychology of Education,* 16(2013): 661-681.

McEwan, P.J. 2008. Can Schools Reduce the Indigenous Test Score Gap? Evidence from Chile. *Journal of Development Studies*, 44(10): 1506-1530.

Mda, T. 2004. Multilingualism and education, in Chisholm, C. (ed.). *Changing Class*. Cape Town: HSRC Press. Chapter 6.

Pijl, S.J., Frostad, P. & Mjaatvatn, P. 2011. Segregation in the classroom: What does it take to be accepted as a friend? *Social Psychology of Education,* 14(2011): 41-55.

Resh, N. & Dar, Y. 2012. The rise and fall of school integration in Israel: research and policy analysis. *British Educational Research Journal,* 38(6): 929-951.

Shepherd, D. 2011. *Constraints to school effectiveness: what prevents poor schools from delivering results?* Stellenbosch Economic Working Papers, WP05/11.

Sommerfeld, A.K. 2016. Education as a collective accomplishment: how personal, peer, and parent expectations interact to promote degree attainment. *Social Psychology of Education,* 19(2016): 345-365.

Soudien. C. 1998. We Know Why We're Here': the experience of African children in a 'coloured' school in Cape Town, South Africa. *Race Ethnicity and Education*, 1(1): 7-30.

Soudien, C. 2004. 'Constituting the class': an analysis of the process of 'integration' in South African Schools, in Chisholm, C. (ed.). *Changing Class*. Cape Town: HSRC Press. Chapter 3.

Spaull, N. 2012. *Poverty & Privilege: Primary School Inequality in South Africa*. Stellenbosch Economic Working Papers, WP13/12.

Taylor, S. & Coetzee, M. 2013. *Estimating the impact of language of instruction in South African primary schools: A fixed effects approach*. Stellenbosch Economic Working Papers, WP21/13.

Thrupp, M., Lauder, H., & Robinson, T. 2002. School composition and peer effects. *International Journal of Educational Research,* 37(2002): 483-504.

Vandenberghe, V. 2002. Evaluating the magnitude and the stakes of peer effects analysing science and math achievement across OECD. *Applied Economics,* 34(2002): 1283-1290.

Vardardottir, A. 2015. The impact of classroom peers in a streaming system. *Economics of Education Review,* 49(2015): 110-128.

Von Suchodoletz, A., Larsen, R.A.A., Gunzenhauser, C. & Fäsche, A. 2015. Reading and spelling skills in German third graders: Examining the role of student and context characteristics. *British Journal of Educational Psychology,* 85(2015): 533-550.

Wilkinson, I.A.G., Parr, J.M., Fung, I.Y.Y., Hattie, J.A.C., & Townsend, M.A.R. 2002. Discussion: modeling and maximizing peer effects in school. *International Journal of Education Research,* 37(2002): 521-535.