Teacher supply and the quality of schooling in South Africa

Patterns over space and time

MARTIN GUSTAFSSON

Stellenbosch Economic Working Papers: 03/16

KEYWORDS: SOUTH AFRICA, TEACHER SUPPLY, EDUCATION PLANNING, SPATIAL ANALYSIS
JEL: C21, D73, I28

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

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

Teacher supply and the quality of schooling in South Africa
Patterns over space and time

MARTIN GUSTAFSSON[1]

ABSTRACT

The paper addresses policy questions in South Africa's education system using a newly merged 1999 to 2013 panel of data that includes school enrolments by grade, staff details from the payroll system, examination and test results and the geo-coordinates of schools. This combination of data, which is seldom used, at least in developing countries, permits new and important knowledge about a schooling system to be uncovered. Whilst policy conclusions are South Africa-specific, the methods would be largely transferable to other contexts. It is shown that school data can complement official population data with respect to the monitoring of within-country migration and in determining the rate of urbanisation. An approach for calculating the viability of small schools in a context of migration out of rural areas is presented, using assumptions around maximum distance to be travelled by pupils and the degree to which multi-grade teaching by teachers should be permitted. Cost reductions associated with a reduced presence of small schools, and greater economies of scale associated with larger schools are found to be smaller than what is generally assumed. Correlations between pupil under-performance and the under-staffing of schools are found to be higher at the primary than the secondary level, apparently confirming the greater importance of personal interaction with a teacher for younger pupils. Between-school movements of pupils other than those associated with urbanisation are found to be high, and highly variable across districts. This further complicates the allocation of publicly paid teachers. An approach to gauging whether teachers avoid moving to schools on the other side of provincial boundaries is presented. It is confirmed that movement across provinces, which are the employers of teachers, is restricted, creating further obstacles to efficient teacher allocation. It is confirmed that teachers tend to move to better performing schools, but that the performance signals that influence this movement are often inaccurate and a few years old.

Keywords: South Africa, teacher supply, education planning, spatial analysis
JEL codes: C21, D73, I28

---

[1] Martin Gustafsson is a researcher based at the Department of Basic Education (DBE) and a member of Research on Socio-Economic Policy (ReSEP), a group within in the Department of Economics at the University of Stellenbosch. Special thanks go to Nic Spaull for inputs and advice regarding the paper. This document has been produced with the financial assistance of the Programme to Support Pro-Poor Policy Development (PSPPD), a joint initiative between the Department of Planning, Monitoring and Evaluation (DPME) and the European Union

Note on summaries

Given the fact that the sections of the report deal with relatively separate matters, summaries appear for each section, and there is no summary for the report as a whole, other than the short abstract. Readers can find a summary box after each level one heading, starting from section 2.

Contents

Tables and figures

1 Introduction

This report forms part of a set of reports produced by various researchers and focussing on binding constraints within South Africa's schooling system. The present report contributes towards understanding constraints relating to teachers. Here the focus is mostly on the more quantitative side of the constraint, in the sense of numbers of teachers. Separate reports pay attention the quality of teachers. Yet the words 'quality of schooling' appear in the title of the report, because educational outcomes are a concern which informs much of the discussion in the report. At times, this is explicit, at other times implicit. Moreover, some of the analysis presented involved linking teacher supply data to learner performance data in novel ways.

It goes without saying that for schooling to occur, teachers need to be employed at schools. As is discussed within the report, to some extent the perception exists that many schools in the country lack even a basic supply of employed teachers, because the teacher workforce is ageing, not enough young people are entering the profession and schools in certain areas are seen as such undesirable places to work, so that teachers do whatever they can to avoid working there. A key purpose of the report is to provide fresh analysis, using previously unexplored or under-explored data analysis techniques, to examine the size and shape of the perceived teacher under-supply problem. There is a strong emphasis in the analysis on the dynamic nature of many teacher supply problems. They change over time and a part of the challenge is attempting to forecast what recent trends suggest will be the situation five, ten or twenty years from now. In this regard, appreciating the implications of South Africa's ongoing urbanisation process for teacher supply is important. The 'small school problem', which influences the sufficiency of school-level teacher supply in a number of ways, must be understood within the context of urbanisation.

Methodologically, this report attempts to innovate through the application of spatial analysis, both mapping for descriptive purposes and the use of between-school distances for analytical purposes. There is little spatial analysis in education policy work, and the report attempts to demonstrate that this represents a lost opportunity. In carrying out the spatial analysis, the report draws from recent innovations in the Stata statistical software package which now permit a much richer variety of spatial analysis techniques than were possible just a few years ago within Stata[2].

The analysis presented in the report is by no means exhaustive. A lot more could be done with the data and suggestions are provided in this regard. In many ways, a purpose of the current report is to encourage a wider range of analysts to use the data, and for the data to become more publicly available.

A short summary of the various sections follow. The topics covered in the report are various and the sections are relatively independent of each other, even if they draw from each other to some extent. This explains why executive summaries are provided for each section, rather than for the report as a whole.

Section 2 below describes some of the literature that informs this report and which the report attempts to build on. The 'binding constraints' approach to policy analysis is discussed here.

Section 3 describes the key datasets used in the report's analysis, including the work undertaken to optimise the linkability of the various datasets and to create a normalised panel of data.

Section 4 presents an analysis of changes in the distribution of enrolments between 1999 and 2013 in ordinary schools, both public and independent, partly to create some sense of the

[2] Some of this innovation is South African. See Brophy, Daniels and Musundwa (2014).

dimensions of the urbanisation process occurring in the country, a process that should inform the education policy debates to a larger degree. Section 4 also makes an argument for stronger use of school enrolment data in order to verify trends seen in the population data, but also to fill the considerable gaps that characterise the latter data.

Section 5 offers an analysis of the evolving relationship between school size and school remoteness. The focus is partly on exploring what the opportunities and risks are with respect to the further closure of exceptionally small schools in sparsely populated areas, and how differences in the inherited spatial distribution of schools, partly a result of the fragmented pre-1994 apartheid administration, play a role.

Section 6 examines the extent to which changes in the distribution of teachers in the 1999 to 2013 period promote resourcing equity and specifically a sufficient availability of teachers in historically disadvantaged schools. In this section a model is presented for simulating the national policy implemented by provinces, somewhat differentially, which determines what the official equity standard is for teacher distribution. This policy is the post provisioning norms, which are widely seen as requiring better monitoring and implementation, and possibly fairly fundamental redesign. Section 6 puts forward a number of recommendations in relation to this policy. What is also presented is an analysis of the economies of scale permitted by the shift towards larger and more urban schools.

Section 7 presents findings from an analysis that has never been performed on South African data before, and is hardly ever performed elsewhere, namely an analysis of the movements over time of individual teachers between schools, between schools and the education administration, and into and out of the education system. The focus here is partly on identifying useful indicators of staffing mobility, or immobility, in individual schools and how the values for these indicators are distributed across the country.

2 Pointers from the literature

This section discusses what the literature says about two topics. Firstly, what has been written about what can be called the 'binding constraints' approach to prioritisation in government planning is critically discussed. This approach has been used to frame a set of reports of which the current report is one. The approach, which has not been applied in an education planning context previously, is relatively straightforward and emphasises planning advice which readers would probably be familiar with. The priorities relating to teacher supply put forward in South Africa's National Development Plan are described and the plan's success in advancing good planning principles is briefly discussed.

Secondly, literature relating to South Africa's urbanisation process is discussed.

Subsequent sections in the current report include further discussion of other literature which is of relevance to the analysis.

2.1 Data analysis that truly assists the policymaking and change process

How should analysis of data and the translation of this analysis into policy advice proceed? This is a difficult and important question that is too seldom addressed directly. There appear to be no 'magic bullets', or straightforward and infallible solutions. Yet some of the answers do appear to have become clearer over time. This report draws from the 'growth diagnostics' approach of Ricardo Hausmann and others. This approach is not unlike many other approaches to using evidence to inform policymaking. Importantly, the growth diagnostics approach as originally conceived is focussed on advising policymakers on interventions that maximise economic growth, though as the authors of the approach themselves say, there is no reason why the approach cannot be adapted to specific sectors, such as basic education[3]. But if this is to be done, it may be better to refer to the approach as the 'binding constraints' approach, as binding constraints are a key element of growth diagnostics. Importantly, no evidence could be found of the use of the binding constraints approach in any specific sector of a country, such as education. The suggestions made below are thus exploratory[4].

The approach of Hausmann and his team emphasises two necessary shifts. Firstly, there needs to be a shift towards a greater use of analysis and evidence from the country in question. This is of course not a controversial call and is partly a reaction to bad experiences with more formulaic and one-size-fits-all approaches that, in the case of advice on economic growth, have often relied on simplistic Washington Consensus-type arguments. Hausmann does not put forward a detailed model on how country-specific data should be analysed, but instead advances certain analytical principles. One is thus dealing with 'disciplined art', rather than a specific model[5].

Secondly, and somewhat more controversially, the approach calls for a very strong prioritisation of just one or two interventions that tackle one or two key binding constraints. Binding constraints are constraints, such as the availability of schoolbooks, attendance of learners at school, or teacher attendance, which are considered binding in the sense that if we do not resolve these constraints, resolving non-binding or less binding constraints will not make a difference to our desired outcomes, better learning outcomes in the case of education. For instance, non-attendance of teachers at schools would be a typical binding constraint. If this problem is pervasive, and if schools have books, a relatively good curriculum is in place,

[3] Hausmann, Klinger and Wagner, 2008: 20.
[4] A good and relatively independent critical overview of the growth diagnostics approach can be found in Habermann and Padrutt (2011).
[5] Habermann and Padrutt, 2011: 7.

teachers are more or less capable of teaching, then clearly what one should prioritise is getting teachers into schools. Why is prioritisation so important? One answer to this question is that constraints and hence solutions have a hierarchy. There are certain things that must be tackled first, for instance getting teachers into schools. A further justification for prioritising would be that it is often impossible for governments to focus well on many interventions at once. It is often better to do a few things right than tackle a wide range of things and obtain mediocre results. One could moreover argue that getting one or two things right carries with it a positive externality. If one succeeds in getting teachers to attend school, a general sense of despair about the education system may simultaneously be tackled, and this could lead to across-the-board improvements in many areas, from textbook provisioning to school management, even if those other areas were not explicitly mentioned in the government's change strategy. Success in one area breeds success in other areas, in other words. This positive externality argument could be used to counter someone like Jeffrey Sachs, who is sceptical about what he sees as an over-emphasis on prioritisation within the growth diagnostics approach. For Sachs, it is valid to pursue a 'big push' approach where many things receive focus at the same time. Yet the evidence seems to suggest that at least in the policy area of economic growth, change comes about when a government tackles different small bundles of issues sequentially, at different points in time[6].

Clearly the basic education sector is not like the economy as a whole. What sector-specific features should one bear in mind if one wants to apply the binding constraints approach to education? One can probably safely say that the education sector, as a sub-section of the entire economy, is less complex than the economy as a whole. Whilst one should not under-estimate the complexities around determining what might bring about better educational quality in South Africa, for instance, this question seems considerably less complex than the question of how to raise South Africa's economic growth. There is probably a lower risk within the education arena that tackling two binding constraints at once will result in harmful contradictions, in the way that tackling high taxes and the overall deficit might. Tackling more than one constraint at a time in education may thus not be that risky. At the same time, however, the positive externality effects described above are likely to be strong in one circumscribed sector, such as education, because the people one is dealing with in one intervention tend to be the same people one would be dealing with in another intervention. This would strengthen the argument for focussing on just one or two binding constraints.

How well has South Africa done when it comes to, firstly, basing education policy advice on evidence and, secondly, prioritising education interventions? The 2012 national development plan, *Our future: Make it work*, provides some answers[7]. The use of evidence on what brings about educational quality improvements of what size, something one would expect in evidence-based policymaking, does not feature strongly in the plan. But evidence-based policymaking is more than this. It is also about using reliable descriptive data on enrolments, education outcomes and the various strands of service delivery to inform the discussion. Here the national development plan fares a bit better. One could argue, however, that there is not a strong sense of prioritisation. The emphasis on one or a very few interventions advocated within the growth diagnostics approach is not reflected in the plan. It is instructive to list key high-priority interventions mentioned in the national development plan focussed on improving learning outcomes:

- Bursaries for the in-service training of teachers
- Stronger requirements for teacher to show evidence of ongoing professional development
- More support initiatives run from district offices and targeting specific schools
- High-speed broadband for all schools

[6] Habermann and Padrutt, 2011: 9.
[7] National Planning Commission, 2012.

- Reductions in class sizes, possibly by relaxing entrance requirements into the teaching profession
- Greater management leeway for good school principals
- Improvements to the Annual National Assessments programme to make results more comparable
- Better information packages directed at parents on learner performance
- Monetary rewards for schools linked to improvements in standardised tests.

The above list provides a sense of where the political will is focussed and should inform any analyst wishing to identify constraints, prioritise them, and find ways of tackling them.

2.2 Urbanisation and educational progress

Urbanisation, despite problems such as over-crowding created by it, is generally considered a progressive trend, with positive educational associations. The 2014 *World urbanisation prospects* of the United Nations says the following[8]:

> Cities are important drivers of development and poverty reduction in both urban and rural areas, as they concentrate much of the national economic activity, government, commerce and transportation, and provide crucial links with rural areas, between cities, and across international borders. Urban living is often associated with higher levels of literacy and education, better health, greater access to social services, and enhanced opportunities for cultural and political participation.

Comparing levels and rates of urbanisation across countries is common, but these comparisons need to be treated with caution. The widely quoted figures of the United Nations Population Division accept national definitions of what constitutes ‘urban’ and efforts are instead devoted to ensuring that measures for each country use the same approach across years. This is perhaps the best one can expect, given how difficult it would be to standardise definitions of ‘urban’ for the world.

South Africa is said to be 64% urbanised in 2014, against for instance 61% for Southern African, 40% for Africa and 51% for middle income countries. South Africa thus emerges are rather urbanised. As countries become more urbanised, the rate of urbanisation tends to slow down. Thus it is not surprising that South Africa is becoming more urbanised at a rate of 0.8 percentage points a year, so at a slower rate than in middle income countries as a whole, wehre the rate is 1.3 percentage points a year[9].

It is not easy to find what South African sources the UN Population Division uses, but what has been published for South Africa specifically agrees broadly with the UNPD’s figures. For instance, Kok and Collinson (2006: 19) concluded that in 2001 56% of the South African population was urbanised, against a 2001 figure of 58% used by the UNPD. Kok and Collinson defined areas as urban largely on the basis of people’s sources of income, specifically the degree to which this was non-agricultural, population density and physical features such as the types of roads in an area. These authors make the point that the apartheid system created a rather unusual situation of densely populated rural areas with only limited reliance on agricultural income, due to migrate labour and remittances. Laldaparsad (2013) arrives at an urbanisation figure of 65% for 2011. Laldaparsad, and presumably analysts, classify densely populated areas in the former homelands, even outside of the ‘homeland capitals’, as urban[10]. Whilst such an approach may follow standard methodologies, it would

---

[8] United Nations: Population Division, 2014: 3.
[9] United Nations: Population Division, 2014.
[10] This is clear if one examines Laldaparsad’s map of urban areas in Limpopo in a slideshow titled “Statistical Approaches for Classifying and Defining Areas in South Africa as ‘urban’or ‘rural’” and available at

not be in keeping with popular South African definitions of what it means to live in an ‘urban’ area.

Turning to the benefits of urbanisation for education, the following extract from a report by Centre for Development and Enterprise[11] explains the diversity of ways through which urban centres facilitate education.

> ... urbanisation and densification can help improve education by reducing the cost of delivering education and raising its quality through increased competition and greater specialisation in both public and private provision of education and training. In addition, the density of urban areas makes the exchange of information, knowledge and skills easier, increasing the productivity gains from education. This also increases the individual returns on education which encourages people to stay in school and to seek to upgrade their skills. Finally, the presence of large and medium size firms in cities means that there is far more opportunity for (and supply of) in-firm training to workers, raising the quality of the labour force.

To the above could be added the fact that teachers tend to prefer working in urban areas. Attracting teachers to remote areas is a typical policy challenge in many education systems across the world.

---

https://www.statssa.gov.za/ycs/SpeakerPresentations/Acropolis4/Day2/SessionVIC/Session%20VIC_Ms.%20Cecilia%20Makupe_Acrop_4/Sharthi%20Laldaparsad%20ISI2008.pdf.

[11] Centre for Development and Enterprise, 2014: 60.

3 Data opportunities and constraints

The data used for the current report are described. Specifically, four different data sources were linked to each other: enrolment by school and grade; payroll data per educator and school; Grade 12 mathematics results; geo-coordinates of schools. It is explained that combining datasets of this kind is largely a question of overcoming hurdles relating to how schools are identified differently in different sources. Data spanned the period 1999 to 2013.

The rationale for using Grade 2 enrolments as a relatively comparable indicator of trends in the child population is explained.

Education systems around the world tend to be characterised by a considerable presence of data, but little capacity to analyse these data in a manner that assists policy processes. The relative absence of informative data analyses can often create the impression that there are no data, or that data are not usable, which in turn can lead to a situation where policymakers focus all their attention on improving the supply of data[12]. This report demonstrates that certainly in South Africa, much can be done with the available data. Clearly there is a need to improve the availability of good data, but if the data which are already available are not used and understood, chances are that any new data will be limited by the same problems as in the past.

The present report makes use of a recently compiled panel dataset that combines four different datasets: Snap Survey learner and staff counts; Persal teacher payroll data; Grade 12 mathematics examination results; and geographical coordinates from the master list of schools. The four data sources can be described as follows:

- **Snap Survey learner and staff counts.** The Snap Survey data collection occurs through two methods, depending on the kind of information system employed by different provinces and schools. The definitions for enrolment and staff are however national. In certain schools, school principals complete, at the start of each school year, enrolment aggregates by grade and gender, and staff by gender and staff category. In other schools, schools-based databases with details on each learner are used to compile a computer file with the Snap Survey learner aggregates, which is then transferred to the provincial education department, mostly through a physical medium such as a compact disc. The Snap Survey data for 1997 to 2013 were recently made public through the DataFirst portal based at the University of Cape Town, as part of an initiative to make government data on the basic education sector increasingly available for researchers[13].

- **Persal teacher payroll data.** The Persal payroll system covers public servants across a range of sectors. The Persal records used were those of the nine provincial education departments, covering educators, non-educator staff, and including staff in schools and other institutions falling under the provincial departments. Privately paid educators employed in public schools would be excluded from these data. Persal numbers identifying individuals are managed rigorously. One can assume that every person has the same unique Persal number forever, no matter where he or she works in the public sector.

- **Grade 12 mathematics examination results.** These data come from the public national examinations system. Indicators of mathematics performance developed for an earlier analysis by Gustafsson and Taylor (2013) were included in the dataset.

[12] Heeks (2003) offers an interesting discussion on a closely related topic, namely why the e-education projects of developing country governments so often fail.
[13] See https://www.datafirst.uct.ac.za.

- **Geographical coordinates from the master list of schools.** Starting with the 1996 School Register of Needs, geo-coordinates of schools have been collected and periodically updated. These data are available in the master lists of schools that the Department of Basic Education makes available on its website[14].

For the type of analysis undertaken in this report, much time must be invested in ensuring that different datasets can be linked. Links across the four abovementioned data sources had to be established through school identifiers. This involved considerable work setting up 'translation tables' indicating how to link datasets where different norms and practices had been used to identify schools. Even within one data source, it was necessary to ensure that the same school was linked across different years. The details of the work can be seen in the Stata do-files produced[15]. The national EMIS numbers of schools as they appeared in the 2011 master list of schools was used as an anchor to which schools from various datasets and periods were joined. Obviously the closure of schools and opening of new schools would not allow for this joining in all instances. The final panel of data (actually a relational database with linking across time) is far from perfect. Some links across data sources and over time could be not be established within the time that seemed reasonable to dedicate to this task (roughly ten days' work).

The following table summarises the degree of linkability across the data sources. It should be noted that the staff information from the Snap Survey for 2000 clearly had many large gaps, but for other years coverage was much better (the next worst year after 2000 was 2002 for which data for 20,797 schools were available). With respect to the Persal data, the relatively low 2005 count of schools is not a reflection of missing records (in fact there are no missing employee records for any Persal year), but rather a reflection of problems encountered in linking school identifiers across years. This latter factor is not really a problem if one considers that employee identifier numbers (Persal numbers), as opposed to school identifiers, can be used to link schools over years, using certain assumptions around the constancy of staff over years (this approach is employed, as described below in this report).

Table 1: Summary of panel size

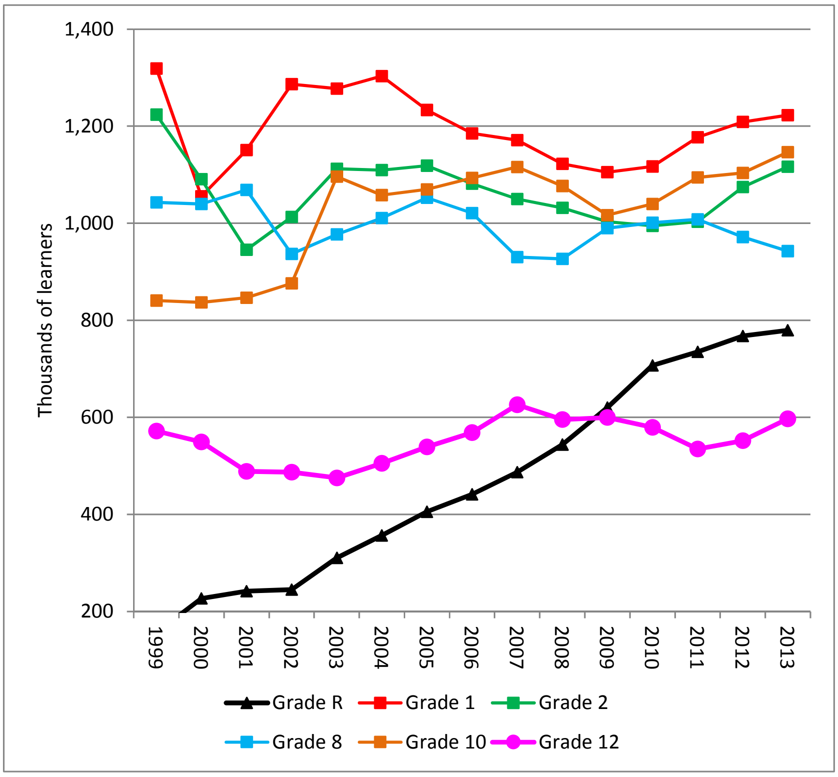
	Years in panel	*Max schools*	*Min schools*	*Max links to 2011 master*	*Min links to 2011 master*
Snap Survey (enrolments)	1999-2013	26,967 (2001)	25,709 (2013)	25,694 (2011)	22,794 (1999)
Snap Survey (staff)	1999-2013	26,296 (2003)	13,132 (2000)	25,656 (2011)	11,932 (2000)
Persal	2005, 2010-2012	24,527 (2012)	18,510 (2005)	25,588 (2012)	19,313 (2005)
Grade 12 exams	2005-2013	5,843 (2007)	5,456 (2013)	5,843 (2007)	5,456 (2013)
Master 2011	2011	25,748 (2011)	25,748 (2011)	25,748 (2011)	25,748 (2011)
Master with geo-coordinates	2005, 2011	25,373 (2005)	25,134 (2011)	25,134 (2011)	23,298 (2005)

Note: The first maximum and minimum values for Persal reflect schools which could be identified as public ordinary schools within Persal's own data. The low value for 2005 is indicative of the fact that the 2005 data did not include component type information (a component could be a school or an administrative office, for example) and so linking to 2012 Persal data was necessary, yet that link was reduced by the fact that component identifiers changed considerably between 2005 and the 2010 to 2012 period.

[14] http://www.education.gov.za.
[15] The best starting point is a document titled 'Details on Martin's school- and employee-level panel dataset' available from the author of the present report.

The following graph illustrating ordinary school enrolments (both public and independent schools) by key grades over a 15 year period provides important background information against which the analysis in the remainder of the report should be seen. Grades 1 to 12 enrolment figures have remained relatively stable over the period, with the exception of one or two noteworthy shifts. The abrupt drop in Grade 1 enrolments in 2000 was the result of the imposition of stricter age-grade norms, which shifted the average age of Grade 1 learners up. The ripple effect of the 2000 drop in Grade 1 enrolments can be seen in the Grade 2 dip one year later, in 2001. The sudden increase in Grade 10 enrolments in 2003 is also striking, but was partly pre-empted by similar trends in earlier grades in earlier years. The steady increase in Grade R enrolments is partly what explains the relatively strong decline in Grade 1 enrolments between 2004 and 2009. As Grade R became increasingly available, there was less pressure for Grade 1 classes to 'baby-sit' younger children. Lastly, the increases in primary level enrolments seen after 2010, for instance the Grade 1 increase in 2011, followed by the Grade 2 increase in 2012, is noteworthy. Analysis of enrolment data by age, plus analysis of Home Affairs birth registrations data, suggest that these post-2010 increases reflect peaks in the number of children born in the 2003 to 2005 period.

Figure 1: Enrolment by grade 1999-2013



Source: Department of Basic Education, 2014.

To some extent, this report uses Grade 2 enrolments as an indicator of trends in the number of children in general. The above graph illustrates why Grade 2 enrolments would serve this purpose better than Grade 1 enrolments. Grade 2 enrolments are lower than Grade 1 enrolments partly because Grade 1 has continually been the grade at the primary level with the greatest degree of grade repetition.

4 The context of urbanisation

This section explains an important part of the education planning dynamic which is too often over-looked, namely the ongoing process of urbanisation. The analysis presented serves as a reminder of how concentrated South Africa's population, and hence her schools, are. This is a concentration which is steadily increasing. Whilst understanding education policy solutions which cater for the specific conditions of rural areas is important, there is probably not enough appreciation of the need to ensure that policies cater for a rapidly urbanising population too. A few indicators of urbanisation, using school data, are explored. For instance, the distance for the average Grade 2 learner to the closest school also offering Grade 2 declined from around 1.4 kilometres to 1.2 kilometres during the 1999 to 2013 period. The trend slowed fairly markedly in 2012, which is indicative of the fact that the increase in births in the years 2003 to 2005 was strong in rural areas. Enrolment decline at the Grade 2 level, at least up to 2011, has occurred for the country as a whole, but this has been particularly concentrated in the rural parts of Eastern Cape and KwaZulu-Natal. The enrolment trends by each of the eight metro municipalities match very closely the population trends seen in Stats SA census data, which affirm the quality of both sources. Both sources confirm that eThekwini (Durban), the country's third-largest metro, has seen relatively slow growth in the 2001 to 2011 period.

It is argued that enrolment data, which is collected more frequently than population census data, ought to be used to a greater extent to monitor urbanisation.

School data allow us to gauge the speed and shape of South Africa's urbanisation process in ways that population data do not. Obviously one should keep in mind the limitations inherent in using enrolment data as a proxy for the population, even the child population. As alluded to in the previous section, grade repetition creates 'noise' in the data. Yet enrolment data are collected universally every year and are readily linkable to fairly dense geo-coordinates, features one does not find in the population census data. In this section four key questions are tackled. Firstly, how large have the school population increases or decreases been in the various parts of the country over the last decade or so? Secondly, to what extent could these changes be characterised as urbanisation, as opposed to some other kind of redistribution? Thirdly, how can we measure the speed of the changes occurring, and has the speed been changing over time? Fourthly, to what extent do the patterns seen in the enrolment data confirm existing internationally comparable statistics on urbanisation? These questions have important implications for education planning and policymaking, in particular as far as school size and access to urban amenities that support the educational process are concerned.

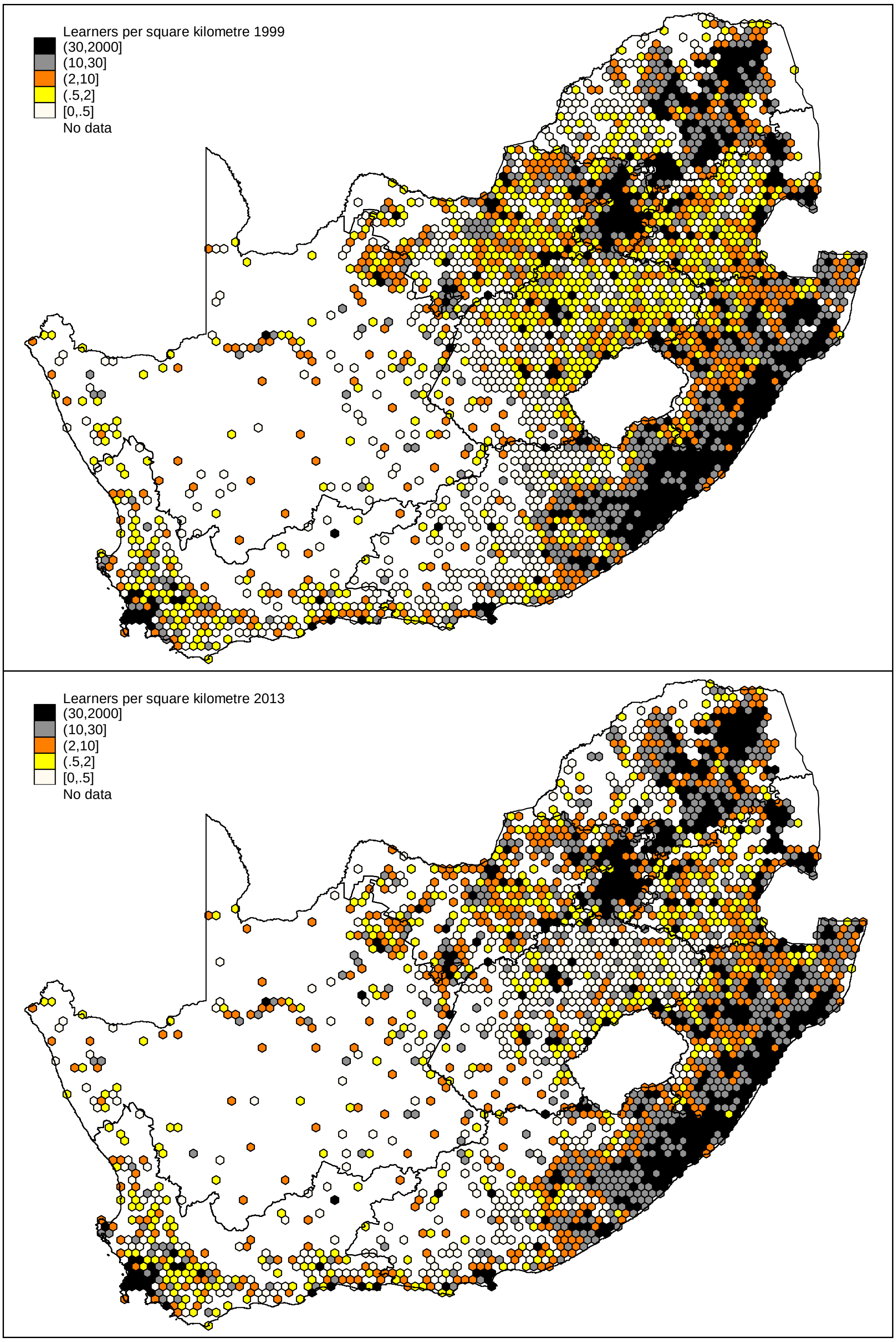
The maps that follow have 'honeycomb' grids which divide the country into relatively small hexagons. One advantage with this approach, over using the 88 education districts or 234 magisterial districts to divide up the country, is that the honeycomb design allows for a more fine-grained picture of distributions and movements. The sizes of the hexagons selected for the maps presented below result in 5,809 cells covering the entire country. However, just under half of these cells appear in the maps below as cells covering areas where there are no schools, in other words very sparsely populated areas, are omitted[16]. On average, the cells are 18 kilometres across and 210 square kilometres[17].

[16] To illustrate, in a 2013 map covering all school grades, 2,647 cells would appear in the map. If one limits this map to grade 2, the number of cells appearing drops slightly to 2,608.

[17] Because one is not dealing with a flat surface, the hexagon size varies with latitude, from 198 square kilometres in far south of South Africa, to 222 square kilometres in the far north of the country. This variation is taken into account when, for instance, population density is calculated.

The two Figure 2 maps illustrate the number of learners per square kilometre within the honeycomb cells, counting all learners in ordinary schools, both public and independent, of any school grade, 1999 and 2013. Between these two years total enrolments increased somewhat, by 5.6%. The general distribution across the country would be familiar to those who have worked with population density maps. Two areas of the country where considerable declines in learner density occurred stand out: rural areas in the Free State, and the far eastern parts of Eastern Cape, in other words the former Transkei.

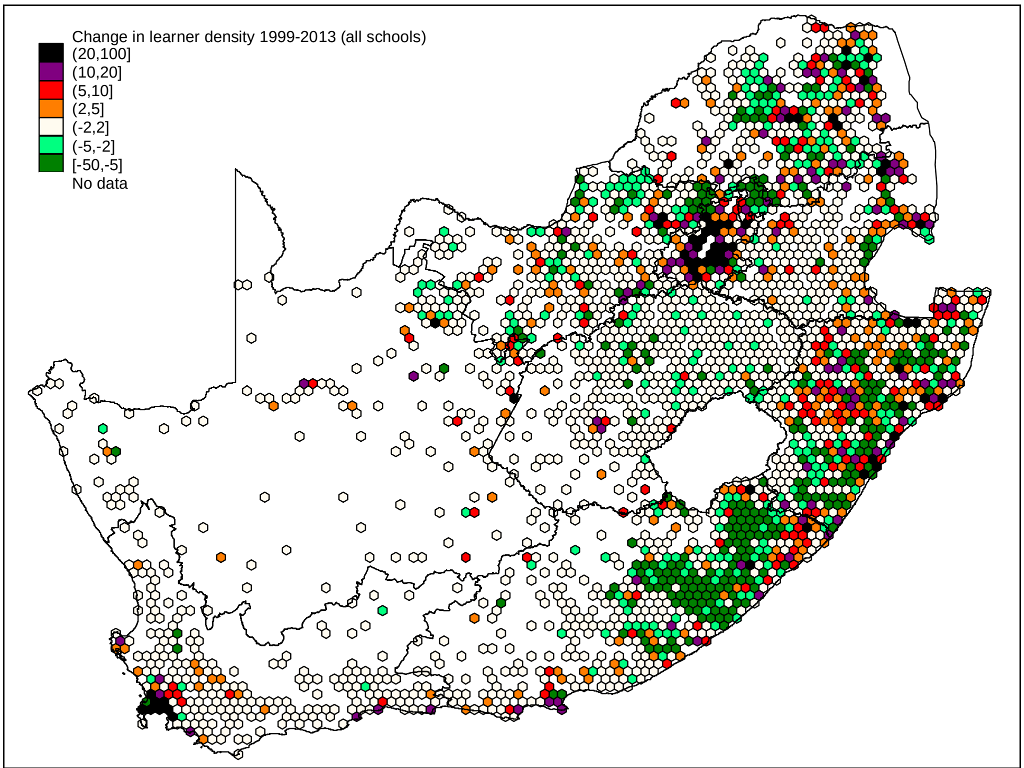
Figure 2: Density of school enrolment in 1999 and 2013

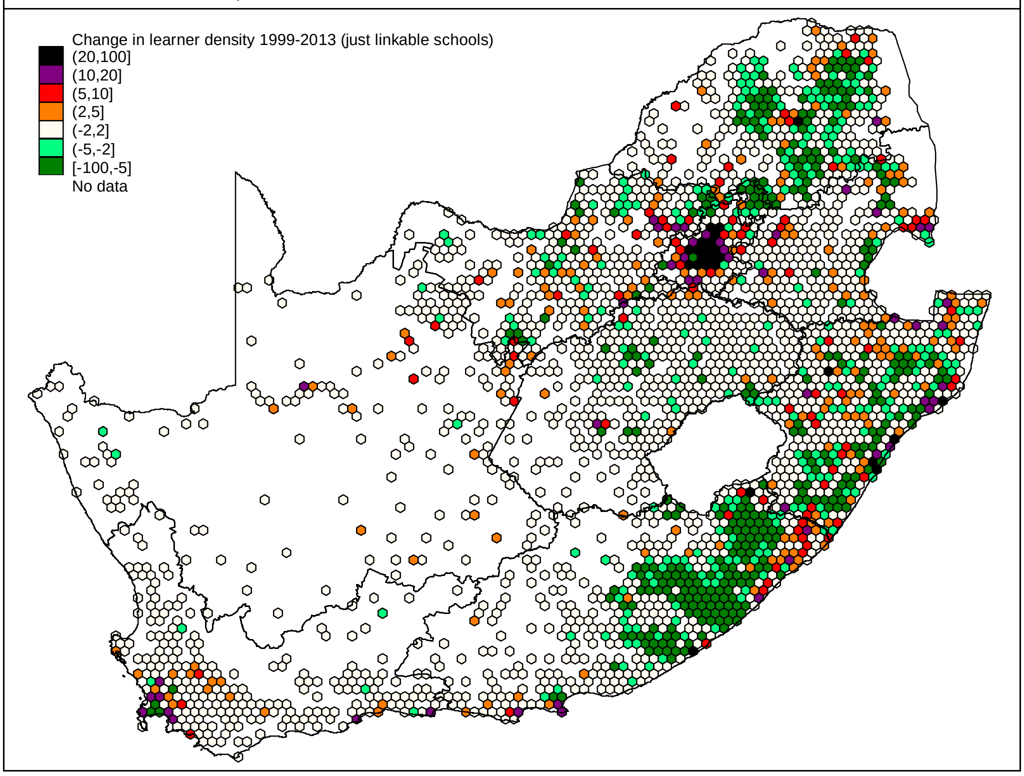


Source: Department of Basic Education, 2014.
Note: The legend indicates the range represented by each colour. For instance, orange represents 2 to 10 learners per square kilometre.

Figure 3 provides two maps, the first of which simply illustrates the differences between the previous two maps within one map. The large pockets of 'learner loss' in the former Transkei and in parts of KwaZulu-Natal are clearly visible. The loss in rural Free State is less striking in the first Figure 3 map as in absolute terms, this decline is not that large. Fairly large pockets of loss immediately to the north of Gauteng stand out. The first Figure 3 map covers all schools available in the 1999 and 2013 data. However, only 21,880 schools could be linked across both years. There were 4,110 schools present in 1999 which were not found in the 2013 data, and there were 2,949 schools present in 2013 which were not found in the 1999 data. A part of this discrepancy would be real, in other words a reflection of closing and opening schools. But a part would be data problems, both missing schools and schools whose unique identifiers changed and could also not be linked after the normalisation process.

Figure 3: Change in density of school enrolment 1999-2013

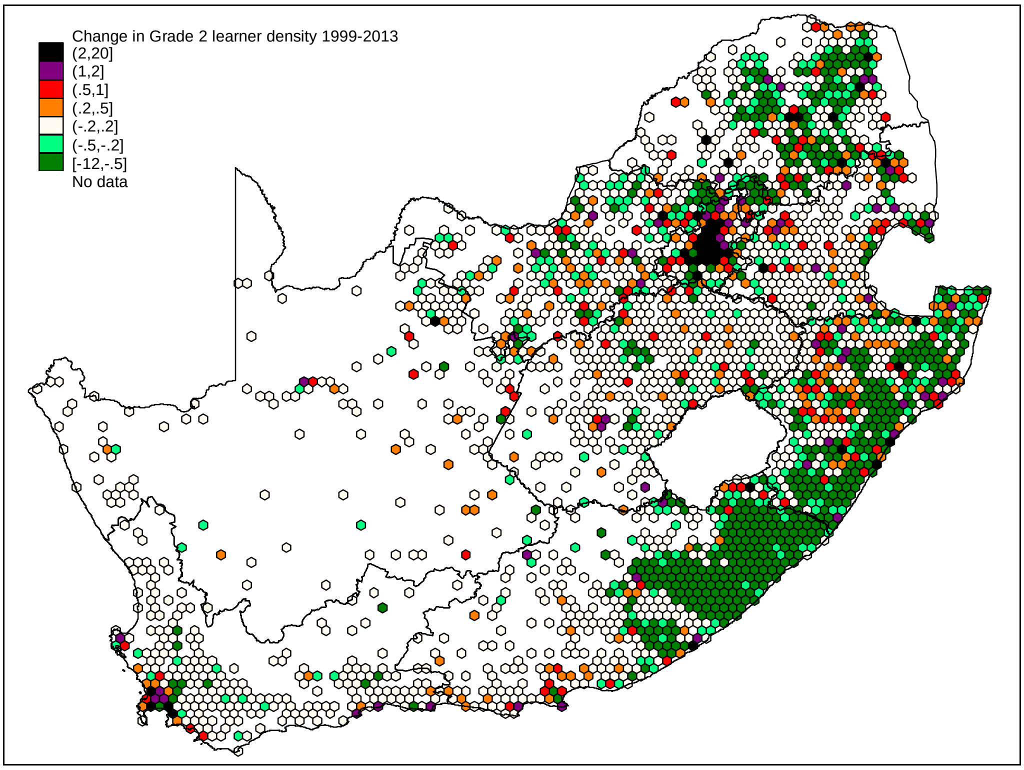




The second Figure 3 map uses data from just the 21,880 schools found in both years. Enrolment in these schools actually declined between 1999 and 2013, by 3.2%. Using only linkable schools does not change the picture to a large extent, but there are some noteworthy differences. In particular, learner gains in northern KwaZulu-Natal are not as pronounced in the second map as the first one. This could point to the fact that schools in this area were missing in the 1999 data, creating a false picture of learner gains in the first map.

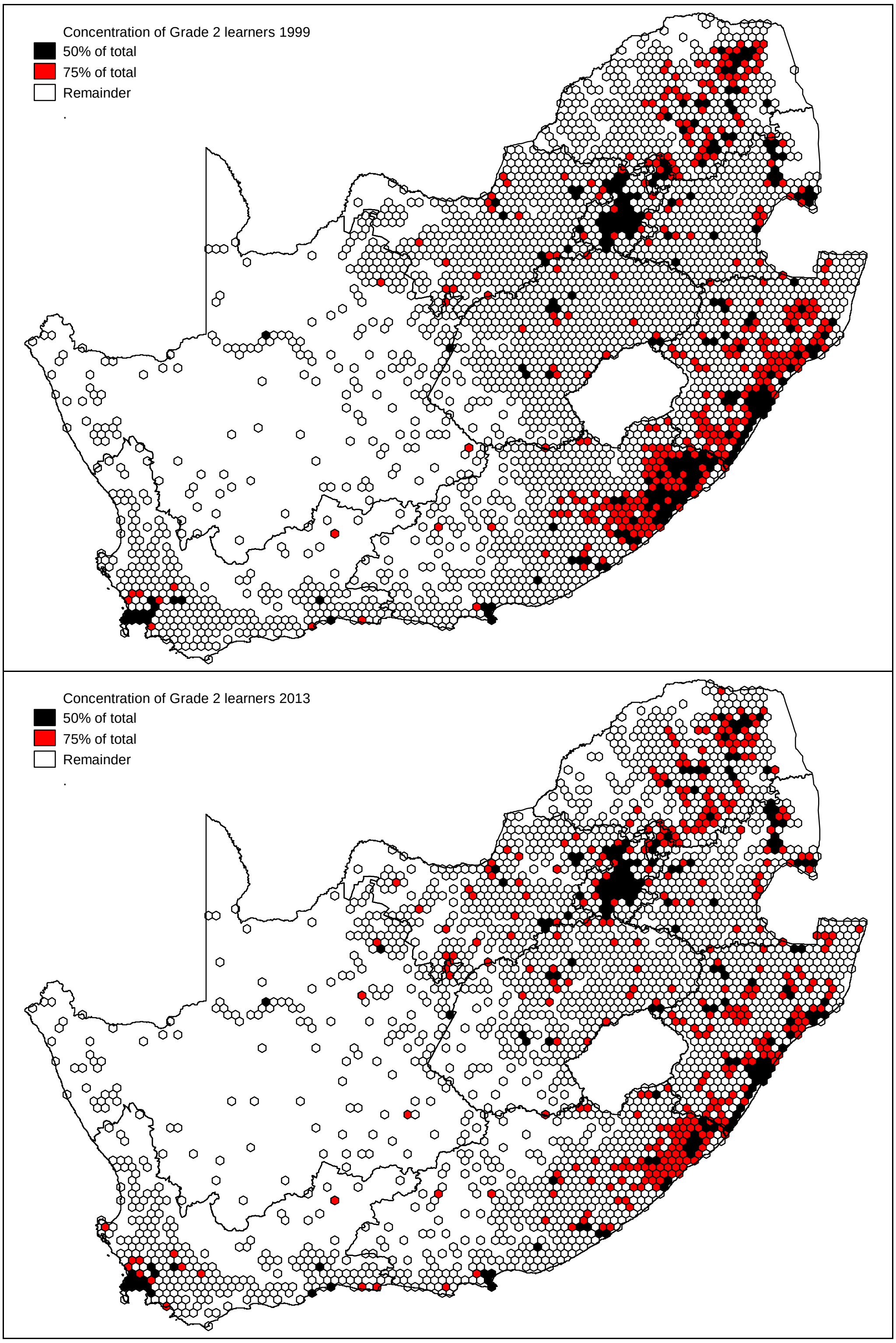
Focussing on just Grade 2 is likely to produce a picture that is closer to that for the population during the period in question given that the percentage of children receiving at least Grade 2 would have been close to 100% in both 1999 and 2013. Higher grades, in particular grades beyond Grade 9, would have seen large enrolment ratio improvements between 1999 and 2013. The problem with Grade 1, on the other hand, is that this grade has been used for 'baby-sitting' over the years, and that grade repetition in this grade has been particularly high, though with large differences across, for instance, provinces. These are all factors that support the use of Grade 2 as a grade that gets relatively close to reflecting one young age cohort in the population, approximately age 8. In Figure 4, all schools from both years have been used. The picture that emerges is of far more widespread learner losses in the Eastern Cape and KwaZulu-Natal. The explanation for this is almost certainly that increases in enrolments at the secondary level mask, within the Figure 3 maps, some of the reductions occurring at the primary level.

Figure 4: Change in density of Grade 2 enrolment 1999-2013



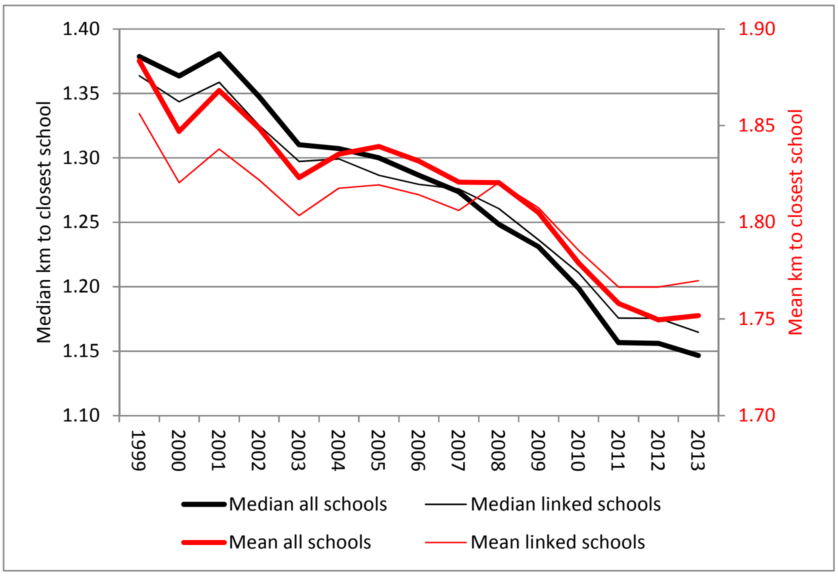
The next two maps attempt to illustrate the changes in the distribution of learners during the 1999 to 2013 period in a different manner. Here honeycomb cells coloured black reflect cells that cover 50% of all Grade 2 learners. Cells were considered starting from the cells with the highest learners per square kilometre values. Red cells expand the coverage to 75% of learners. Black cells in the first map cover 47,504 square kilometres, whilst the figure for the second map is 37,740. The high density area within which half of South Africa's learners are found thus shrunk by around 20%. The 2013 area of 37,740 square metres mentioned here accounts for only 3% of South Africa's overall area. Of course the statistics mentioned here are sensitive to the size of the honeycomb cells selected for the maps.

Figure 5: Geographical concentration of Grade 2 learners in 1999 and 2013



One useful indicator of geographical concentration is the average distance experienced by learners to the closest other school offering the same grade. Schools are probably closer to each other than what many would assume. If in the 2013 ordinary schools data one takes every school offering Grade 2, and measures the distance between that school and the closest other school offering Grade 2, then the median distance obtained across all schools is 1.7 km. At the 90th percentile (near the maximum), the distance is 4.7 km. The next graph illustrates the median learner-weighted value for all schools in the data for each year in the range 1999 to 2013 (see the curve 'Median all schools'). There has been a general downward movement, reflecting the fact that learners have moved closer to each other, as for instance schools in urban schools became larger, rural schools closed, and new urban schools opened. If one uses the mean instead of the median, and if one limits the analysis to the 14,744 schools with Grade 2 for which data existed in every year, the same general pattern emerges. The slowing down of the trend in 2012 is a manifestation of the large 'wave' of children who entered Grade 1 in 2011, and Grade 2 in 2012, and who tended to have a relatively high presence in less densely populated parts of country.

Figure 6: Measure of increasing geographical concentration

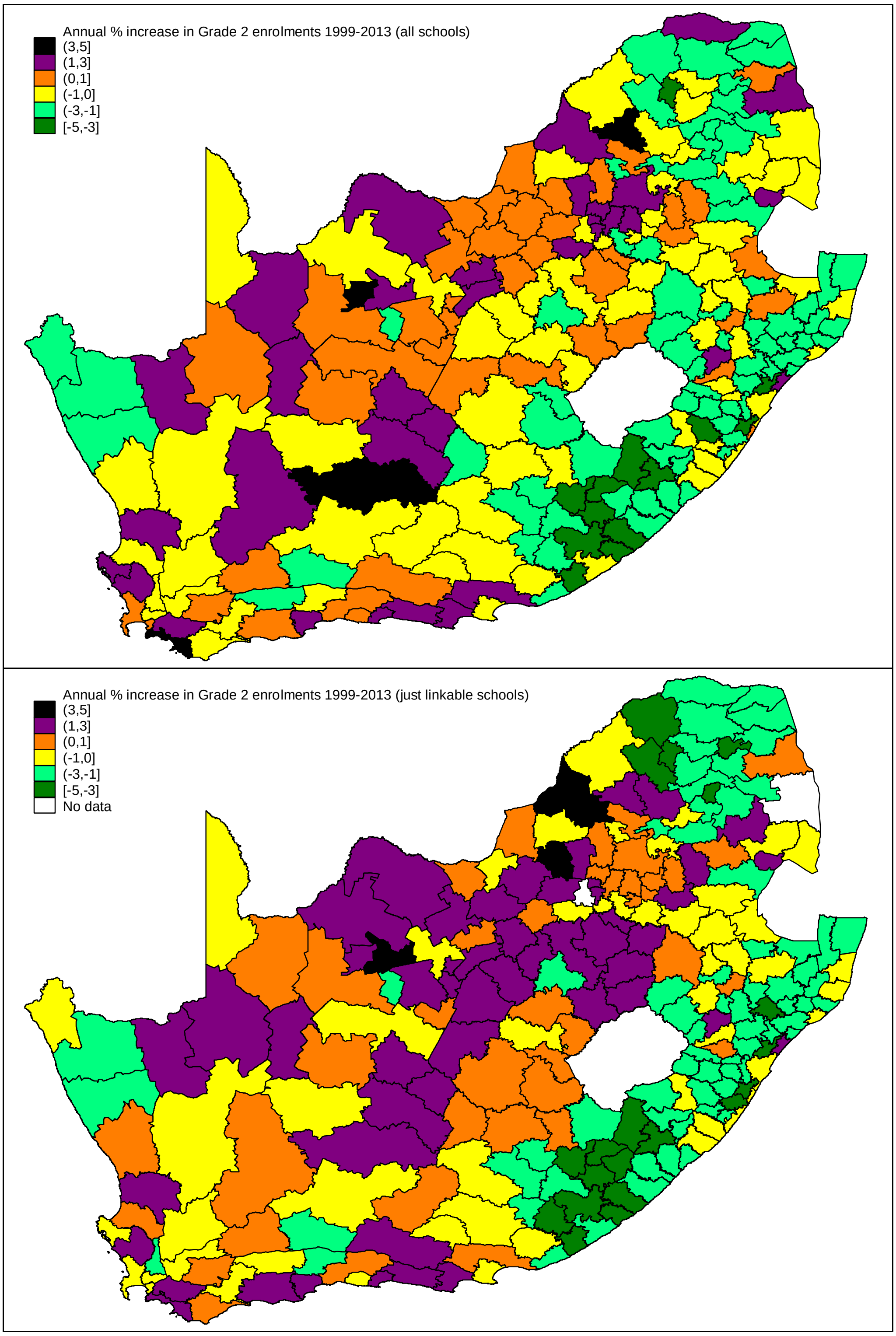


Source: Department of Basic Education, 2014.
Note: The analysis underlying the graph focuses on Grade 2 only. The schools covered in the 'all schools' curves ranged from 19,842 in 2001 to 18,489 in 2013.

At the top end, measures of distance to the nearest school also declined, for instance from 3.6 km to 3.4 km at the 90th percentile between 1999 and 2013 using all schools (learner-weighted) with Grade 2 available in the data.

In order to understand the extent to which changes in the distribution of learners in the country represent a process of urbanisation, it is necessary to identify more and less urban parts of the country. The two Figure 7 maps below represent a first step in this direction. Here Grade 2 data from all years in the range 1999 to 2013 were used. Divisions on the map are those of the country's 234 municipalities, including the eight metropolitan municipalities. For each municipality, the linear trend in Grade 2 enrolments was found and the slope divided by the mean enrolment across all fifteen years was considered the average annual percentage increase, the statistic illustrated in the first map. For the country as a whole, the statistic was minus 0.5%.

Figure 7: Change in municipality-level Grade 2 enrolment 1999-2013

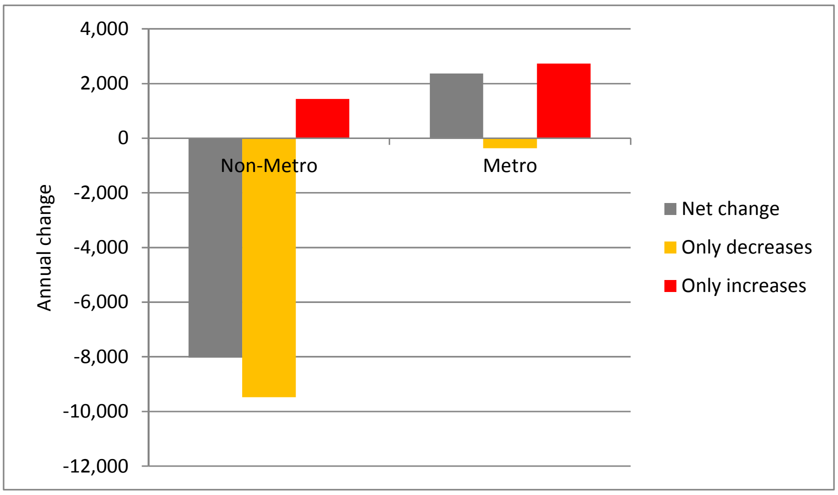


Note: The user-created Stata command 'gpsbound' by Brophy, Daniels and Musundwa (2014) was used to place schools within municipalities.

The second Figure 7 map differs from the first insofar as it used data from only the 14,744 schools for which Grade 2 enrolment values existed for all 15 years. Here the overall national trend would be a 1.0% annual decline in Grade 2 enrolments.

If the annual changes, using a linear trend, are calculated for four groups of municipalities defined by whether they are one of the eight metropolitan municipalities and whether they saw a positive or negative change, the picture seen in the next graph emerges. Clearly non-metro municipalities have on the whole seen large enrolment declines, whilst metro municipalities have mostly seen increases. Specifically, in non-metro municipalities the annual loss in Grade 2 learners has been around 8,000 learners a year, whilst in metro municipalities there here has been a gain in Grade 2 learners of on average 2,368. There were three metros with a decline, namely eThekwini, Nelson Mandela Bay and Buffalo City. But certain municipalities outside of the eight metros have also experienced substantial growth. Figure 8 is based on an analysis where all schools are used. Using just schools that can be linked across all years produces a picture that is not substantially different. Which non-metro municipalities account for most of the growth reflected in the first red bar below? The concentration in this regard is not that large: the five strongest growing non-metros account for 25% of the overall growth amongst non-metro municipalities. The five municipalities in question are Rustenburg, Mogale City, KwaDukuza, Modimolle and Matlosana.

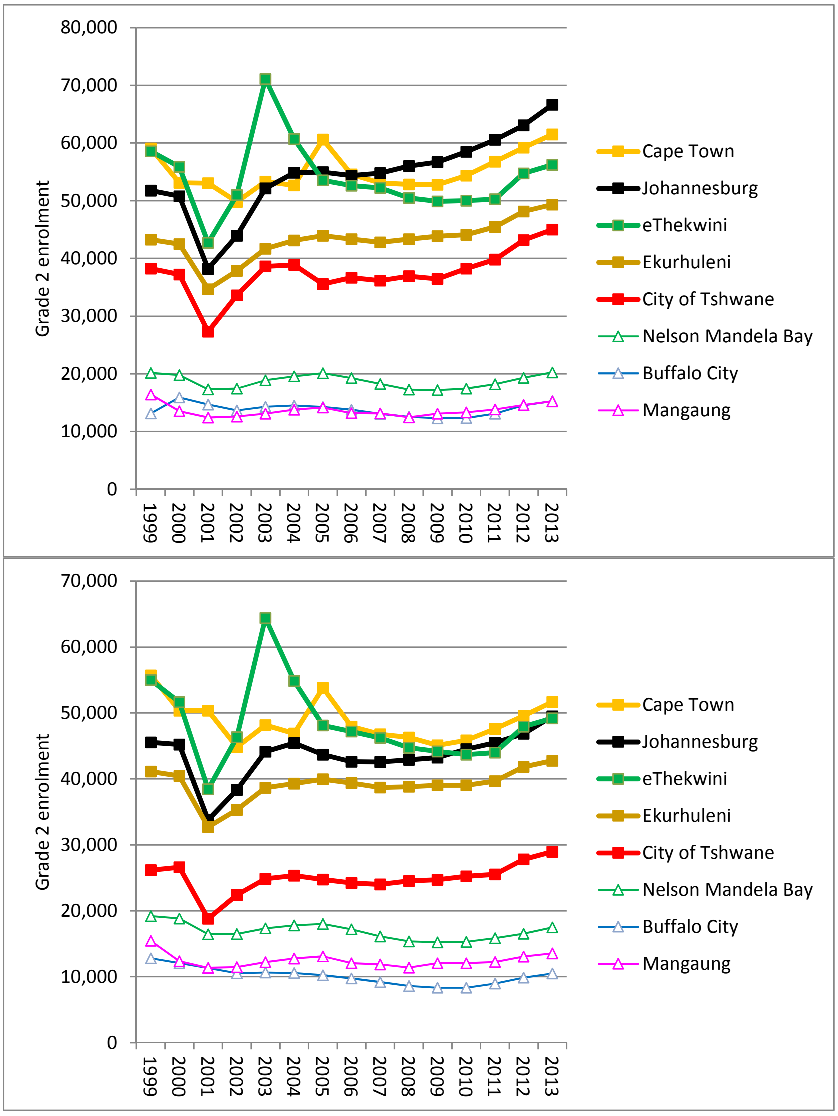
Figure 8: Metro versus non-metro Grade 2 change 1999-2013



The proportion of Grade 2 learners found in the eight metro municipalities grew from 25% to 30% over the 1999 to 2013 period. The annual change in this statistic was a positive 0.4 percentage points. These figures can be compared to the widely quoted figures on South Africa’s urbanisation process discussed in section 2.2. Urbanisation is said to have reached 64% by 2014, confirming that urban areas are typically understood to encompass far more than the country’s metropolitan municipalities. Official figures indicate that by 2011, 39% of the population were in the eight metros (in 2001 the figure stood at 36%). Assuming all figures are correct, this suggests that adults were more likely than children to live in the metros, a hypothesis which would be consistent with theories around how urbanisation occurs. Specifically, young adults will often make a special effort to improve their lives by seeking employment in cities.

The Grade 2 trends for the eight metropolitan municipalities are provided in the next two graphs. The curve for eThekwini (Durban) is particularly striking, for two reasons. Firstly, the 2001 trough in enrolments, resulting from the change in the age of admissions criteria (see above discussion), was followed by an exceptionally high peak (specifically, a 2003 peak for Durban). Closer analysis of the data reveals that this peak for Durban was widely distributed across most schools. It was not the product of a few outlier schools. Secondly, from 2003 onwards Durban saw a decline in enrolments all the way up to the 2012 rise, a rise prompted mostly by increases in the sizes of birth cohorts (see discussion around Figure 1). Durban’s unusual patterns emerge whether one uses all the available data, or data only from schools present in each of the fifteen years. The data suggest that Durban has not followed the general pattern of child population increases seen in other large cities. But one should guard against hard conclusions on the basis of just these trends. A possible hypothesis would be that there were exceptionally strong reductions in grade repetition in Durban, which mask actual population increases within these two graphs.

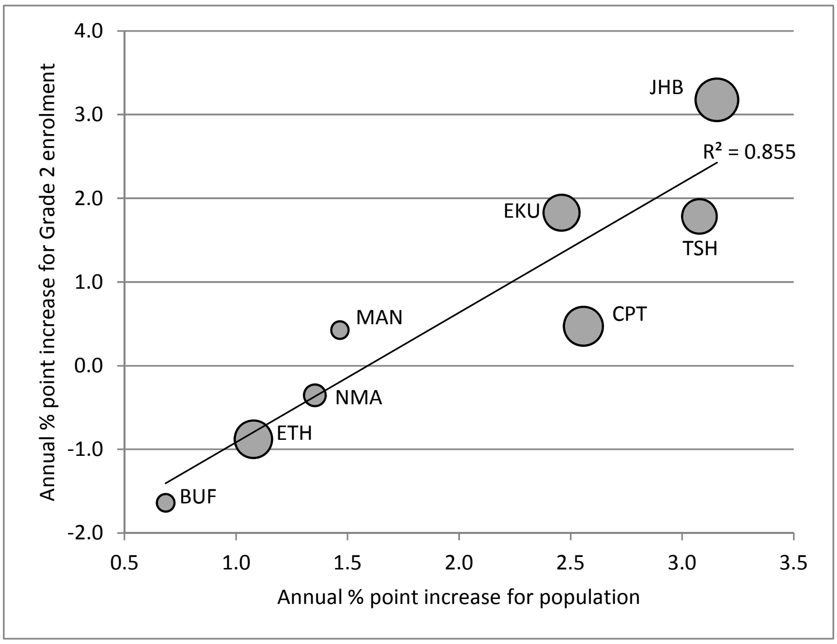
Figure 9: Enrolment trends for individual metros



Note: The first of the two graphs reflects all the available data, whilst the second graph reflects data for only those schools which had data in all the 15 years.

The following graph, Figure 10, indicates that metros with fast enrolment growth, also tended to have fast population growth in the 2001 to 2011 period.

Figure 10: Comparison of metro population and enrolment growth 2001-2011



Sources: Department of Basic Education (2014); Statistics South Africa (2013).
Note: Areas of circles are proportional to the 2011 national census totals.

The above analysis is partly aimed at demonstrating that school enrolment data can complement population data in the quest for a clearer picture of where South Africans live and how these patterns have changed over time. The utility of enrolment data here needs to be noted in South Africa and beyond. There is surprisingly little in the demographics literature, including the manuals of the United Nations[18], on how school data can assist in establishing population patterns, especially in countries with less than ideal population data (one can assume that most developing countries would fall into this category[19]). It is worth noting that half a century ago school enrolment data was clearly considered an important part of the demographer's raw materials. A 1951 handbook issued by the United States census office includes a section on how school data were regularly used in the calculation of population and migration statistics within the United States[20].

[18] http://www.un.org/en/development/desa/population/publications/manual/index.shtml.
[19] Jerven (2013) offers a thorough analysis the problems that characterise national statistics in African countries, including population statistics.
[20] Jaffe, 1951: 223.

5 The inherited spatial distribution of schools

The analysis presented in this section quantifies and maps the prevalence of grade configurations in schools which do not follow the predominant pattern of grades 1 to 7 in primary schools and grades 8 to 12 in secondary schools. Around 29% of learners were in schools with a different grade configuration in 2013. This important fact seems to be under-appreciated in the policy discourse, which seldom takes into account the implications of unusual grade combinations. Eastern Cape is a particularly non-standard province in the sense that over half of learners follow a system of grades 1 to 9 in one school, and grades 10 to 12 in another. A mapping of this phenomenon confirms that it is geographically concentrated in what was the Transkei 'homeland'.

A method is developed for identifying schools which are too small to avoid multi-grade teaching, given prevailing staffing patterns. Enrolment thresholds from this calculation are used together with a distance threshold as inputs into a further model to identify schools which could be closed without leaving learners without sufficiently nearby schools. It is found that over 5,000 schools were of a small enough size to require multi-grade teaching in 2013, but that only around 100 of these schools could be closed. The latter figure is obviously sensitive to assumptions and the method used, yet the conclusion seems supported that multi-grade teaching must be accepted as a reality in many thousand schools over the longer term. In fact, it argued that in a context of urbanisation, where populations in rural areas decline, but never reach zero, because there are always some people left behind, one should an expect an *increase* in the number of schools so small that multi-grade teaching is required (though learners in these schools should in theory decline as a percentage of all enrolments).

Trends with regard to small schools in recent years have been complex, and partly counter-intuitive. The number of schools requiring multi-grade teaching in theory, because their enrolments fall below critical thresholds, has in fact *decreased*. This is probably because closing of small schools has occurred over the 1999 to 2013 period studied. At the same, enrolments in these schools as a percentage of all learners in the country has *increased.* The percentage of all learners (grades 1 to 7) in schools so small that multi-grade teaching was required increased from 7.2% in 2007 to 8.9% in 2013.

The analysis in this section provides a basis for further necessary work involving the modelling of small school trends in future years, which should be informed by assumptions around the speed of the urbanisation process.

This section presents an analysis of school size and school distribution. Understanding, for instance, the shape and size of the 'small school problem', a problem referred to in the policy debates from time to time[21], seems very important. There are a number of questions. How does one determine when a school is too small, given factors such as remoteness and the demand for schooling? Do different parts of the country display different versions of the 'small school problem'? How has the presence of very small schools changed in South Africa over time?

To begin, an examination of the distribution of school size, and the related matter of grade configuration, is presented. The examination first occurs in relation to all schools, public and independent, but the subsequent more detailed modelling of when schools can be considered too small considers public schools only.

Table 2 below confirms that despite the predominance of a configuration where primary schools offer grades 1 to 7 and secondary schools offer grades 8 to 12, there are a number of

[21] Department of Education, 2005.

alternative configurations accounting for 29% of learners. For instance, 9% of grades 1 to 12 learners find themselves in schools offering grades 1 to 9. (In this analysis Grade R has been excluded to simplify matters. Of all learners in Grade 1, 93% were in schools which also offered Grade R in 2013.)

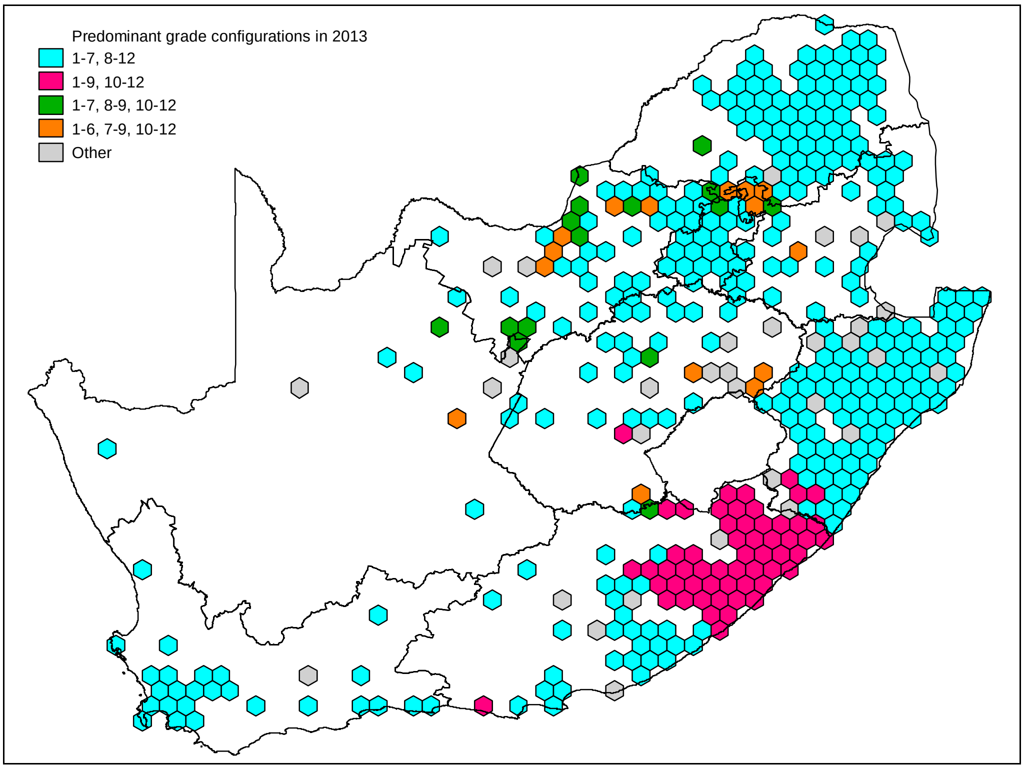
Table 2: Learners and their grade configurations in 2013

Grades	EC	FS	GP	KN	LP	MP	NC	NW	WC	SA
1 to 7	21	36	51	42	48	39	36	46	52	42
8 to 12	13	23	31	36	42	28	20	21	32	29
1 to 9	42	9	1	5	1	6	6	3	4	9
1 to 12	3	6	7	5	3	5	5	4	4	5
1 to 6	3	9	2	1	1	8	10	10	1	3
10 to 12	11	5	1	1	0	3	3	5	0	3
1 to 4	2	0	1	4	2	2	0	1	0	2
1 to 8	2	3	1	1	0	1	8	1	3	1
7 to 9	0	4	1	0	0	3	3	5	0	1
5 to 7	1	0	0	2	2	1	0	0	0	1
Other	2	5	3	4	1	4	9	5	2	3
Total I	100	100	100	100	100	100	100	100	100	100
Total II	15	5	17	23	14	9	2	6	8	100

Note: Figures refer to percentages of learners within a province in the range grades 1 to 12, and cover both public and independent schools. 'Total II' indicates the overall grades 1 to 12 enrolment found in a particular province.

Figure 11 below illustrates where different grade configurations for schools predominate. The most noteworthy deviation from the '1-7, 8-12' configuration of primary then secondary school, is the pattern of grades 1 to 9 followed by grades 10 to 12 in a different school. This pattern is clearly a phenomenon of areas which fell under the Transkei 'homeland' existing before 1994. Patterns following a three-school format, for instance the '1-7, 8-9, 10-12' pattern, are still relatively common in what used to be Bophuthatswana (largely sections of North West Province).

Figure 11: Distribution of grade configurations in 2013

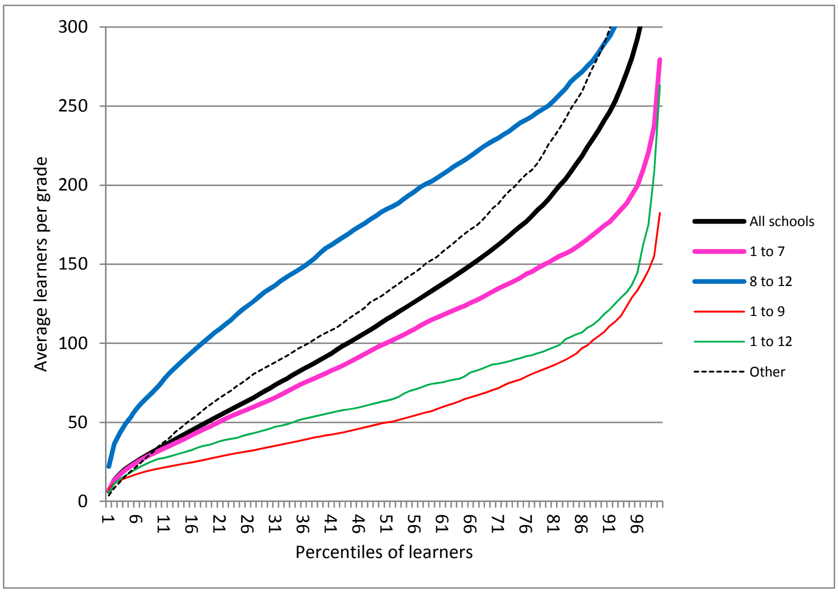


Note: The cells in this map are roughly twice the width of cells in previous graphs (for instance Figure 5). In this map, only cells with at least two schools with Grade 1 and at least two schools with Grade 12 were included. The predominant grade configuration was found by establishing the modal configuration for Grade 1 learners and then the modal configuration for Grade 12 learners.

A better measure of a small school than the total enrolment of the school, is what we can refer to as the 'average grade group size', meaning the total enrolment of the school divided by the number of grades offered in the school. Put differently, it makes more sense to think of a school with 90 learners offering grades 1 to 12 as a small school with potential economy of scale problems, than a school with 90 learners offering just grades 1 to 3. The next graph, Figure 12, illustrates the distribution of the average grade group sizes of schools, with schools weighted by learners. The weighting by learners plays a large role. For instance, the curve for all schools indicates that 10% of learners are in schools where the average grade group size was less than 35 in 2013. The value 35 is often considered a threshold for an economically sized class, for instance for grades 1 to 3 in the policy that distributes educator posts to schools[22] (the 'post provisioning norms'). If we do not weight by learners, we find that 35% of schools experience an average grade group size of less than 35 learners. The difference between the 10% and 35% is large and can lead to confusion. The risk is that the extent of the small school phenomenon is exaggerated, if the 35% of schools is mistakenly taken to mean that 35% of *learners* are in small schools. Using the 35-learner threshold here, in fact only 10% of learners are in small schools.

[22] Government Notice 1451 of 2002.

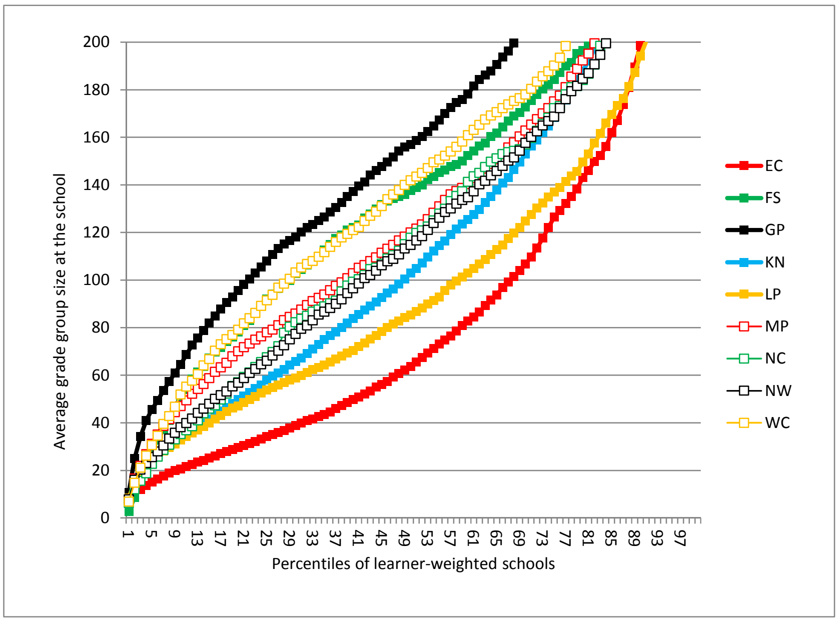
Figure 12: Distribution of grade group size in 2013



Note: To illustrate, '1 to 7' in the legend refers to schools offering grades 1 to 7.

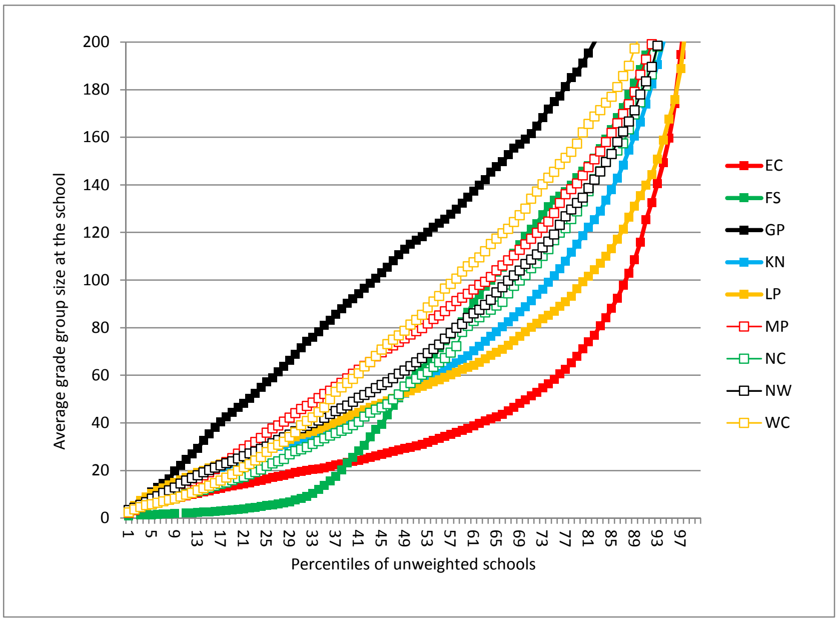
The importance of differentiating between unweighted schools and learner-weighted schools is especially important when examining provinces, as the approach taken makes a large difference to the picture one obtains. Eastern Cape and Free State are often referred to as provinces with large 'small school problems'. As can be seen in Figure 13, Eastern Cape does indeed stand out as a special case. In this province, 25% of learners were in schools where the average grade group size was less than 35 learners in 2013. Free State, however, appears not to experience a similar situation, according to Figure 13. According to Figure 14, however, which focuses simply on the number of schools, without any weighting by learners, Free State does emerge very much as a special case. Here 43% of *schools* in 2013 were schools where an average grade group size of less than 35 was seen. These were exceptionally small schools. The smallest 20% of schools in Free State had an average grade group size of 2.3, against a figure of 8.8 in Eastern Cape. One can think of the Free State 'small school problem' as being one of many extremely tiny schools, where the Eastern Cape's small schools are slightly larger.

Figure 13: Grade group size by province in 2013 (weighted schools)



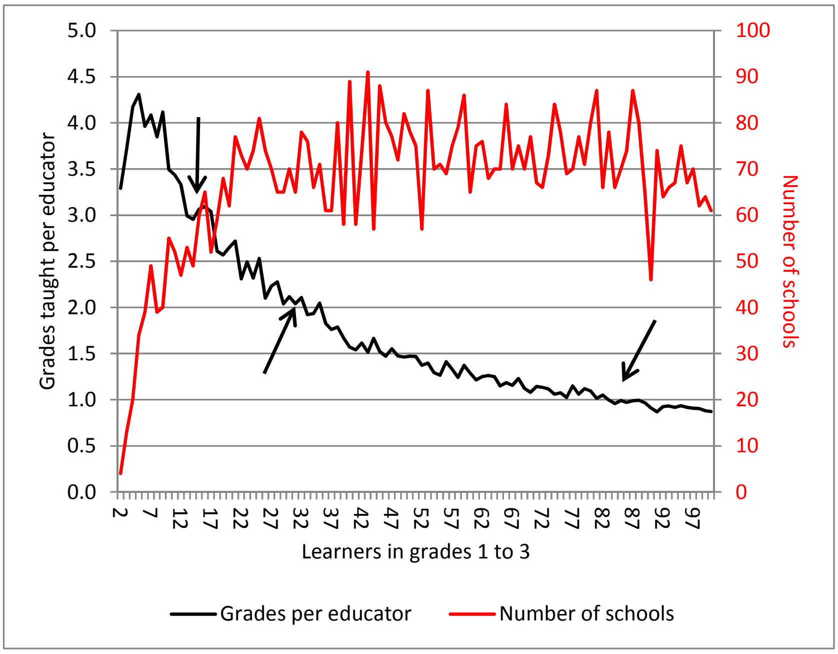
Note: In this graph and the next one, all grades 1 to 12 learners in public and independent schools are considered in the analysis.

Figure 14: Grade group size by province in 2013 (unweighted schools)



How should one determine when a school is too small, given factors such as remoteness and the demand for schooling? Figure 15 below presents a picture of the relationship between the number of grades offered in a school, available teachers and numbers of learners. This picture is necessary if one is to obtain an idea of which schools can be considered too small. To arrive at the graph, the 17,289 public schools with grades 1, 2 or 3 in 2013 were identified. Then 180 schools which did not offer all three grades in question were dropped. The average grades taught per teacher in each school was calculated by taking the number grades offered in the range of grades 1 to 12 and then dividing this by the number of educators employed in the school (publicly paid, and those paid by the school governing body). Whether educators included Grade R teachers was not possible to determine for all schools. To check whether this could be influencing the results substantially Figure 15 was reproduced using only schools without Grade R. The resultant patterns were in fact not substantially different. According to Figure 15 below, in general schools must have at least around 83 learners enrolled in grades 1 to 3 (the 'foundation phase' according to the curriculum) if they are to have enough teachers for one teacher for each grade, in other words a level of teacher resourcing which would not necessitate multi-grade teaching. If one's criterion was that ideally there should not be multi-grade teaching, then one would attempt to avoid having schools with fewer than around 83 learners in grades 1 to 3 (see the right-hand arrow in the graph). The number of schools with fewer than 83 learners in the graph is 5,398. On the other hand, if one's criterion was that there should be at least one teacher per phase (three grades for the purposes of the current analysis), then the enrolment threshold for grades 1 to 3 would be around 13 learners. There are 390 schools represented in the graph with fewer than 13 learners (of the three arrows, the one furthest to the left). An intermediate criterion would be to say the ratio of grades to teachers should not exceed 2.0, which would result in an enrolment threshold of around 30. The thresholds of 83 learners will be used in the scenario discussed below dealing with the number of small schools one could possibly merge.

Figure 15: Multi-grade teaching thresholds in 2013



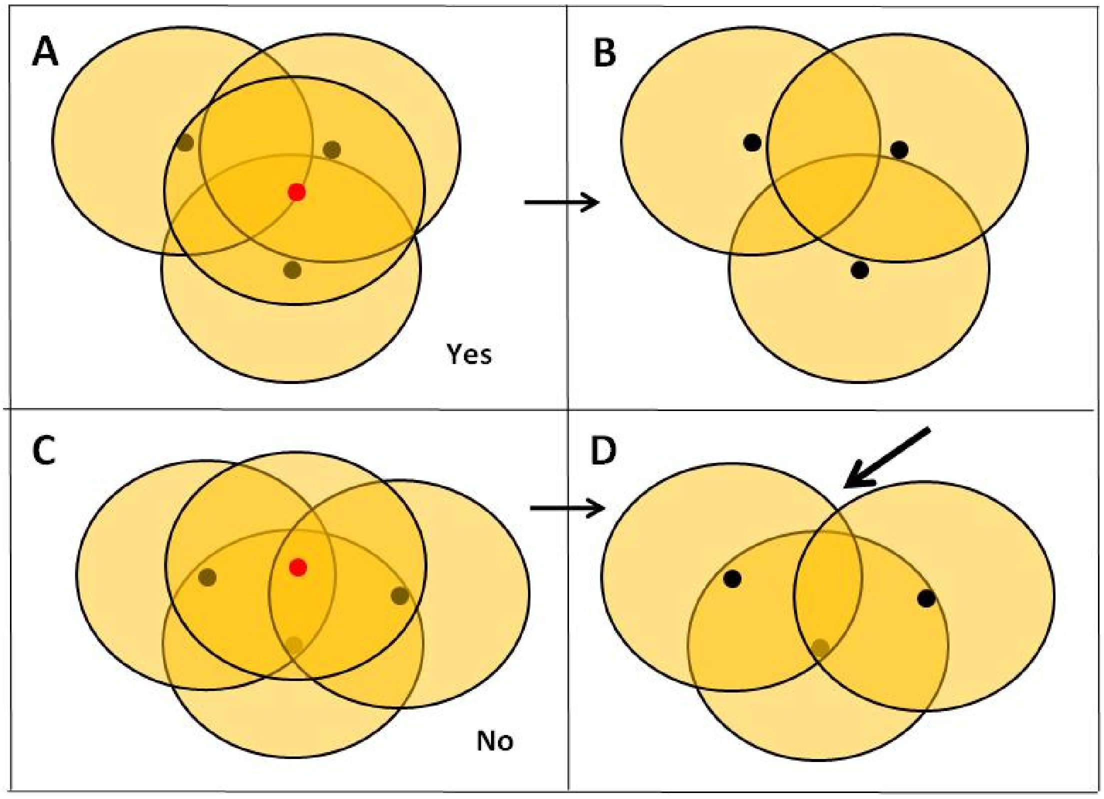
Source: Department of Basic Education, 2014.
Note: The graph uses data from 6,634 public schools. Had the horizontal axis extended beyond 100, many more schools would have been included. Grades taught per educator uses all grades in the school in the range grades 1 to 12, divided by the number of educators working in the school, including school principals, but excluding 'Grade R practitioners'.

Apart from an enrolment threshold, what is also needed is a distance threshold. What is the maximum distance a child should be expected to walk to school? An answer to this question would inform whether merging schools would result in unreasonable increases in the distances walked by learners. The Department of Transport (2009: 22) considers 3 km travelled by learners in one direction, from home to school, a threshold beyond which state-subsidised scholar transport should be provided. The analysis that follows uses not distance travelled, but straight-line distances. The rough assumption was used that the former would be 50% greater than the latter on average, meaning a straight-line distance of 2 km seemed appropriate as a distance threshold in the analysis that follows. It is worth bearing in mind that it would be impossible, given the data that are currently available, to simulate actual distances between people's homes and schools, using pathways that pedestrians are likely to choose. Straight-line distances in analyses of the kind presented here are likely to be the best possible solution for the foreseeable future.

It appears there is no published guidance on how to proceed with the analysis. The most relevant text found was an interesting one dealing with the optimisation of the location of public facilities, such as schools, so that the average distance travelled per user (or learner) was minimised[23]. However, what was needed here was a method that could be used to adapt an existing system, not to establish an ideal system from a blank slate. A set of procedures was thus devised that would deal with the specific situation in South Africa (though presumably many other countries would experience similar situations).

Figure 16 below explains the basic approach. In Panel A the red dot represents a small school, the black dots nearby schools. The question is whether the red dot school could be closed, and the learners could be moved to one of the neighbouring schools. The shaded circles represent areas around each school up to the threshold distance, say 2 km. Panel B illustrates the result if the red dot school is removed. We could say that one could safely close the red dot school because the overall coverage, shaded in yellow, has not changed. Specifically, no learner from the closed school would need to travel more than 2 km (in a straight line) to get to her new school. Panel C represents a different situation. Here, if the red dot school were closed, there would be a reduction in the shaded yellow area (compare panels C and D). Assuming there were learners living in the areas which ceased being yellow, certain learners would find themselves more than 2 km from a school if Panel D. We would thus not want to close the red dot school in Panel C.

Figure 16: When a small school can be closed



[23] Gastner and Newman, 2006.

Appendix A at the end of this report presents the Stata code which was used to discover how many small schools existed which could be merged, using the 83 enrolment threshold and the 2 km distance threshold. Appendix A also explains a few situations where merging may have been possible, but it turned out to be computationally very challenging to confirm this through the methods used. What this means is that the estimates presented for the number of schools which could be merged would be a slight under-estimate, though the under-estimate seems unlikely to be larger than around 10%.

The results of the analysis are presented below.

Table 3: Outputs of school merging simulation

Number of schools analysed	16,866
Total enrolment in these schools (only grades 1 to 3)	3,162,183
Distance threshold used (km)	2.0
Enrolment threshold used	83
Schools with enrolment below the threshold	5,343
Number of schools closed	117
Number of learners in closed schools	6,152
Number of schools receiving learners from closed schools	110

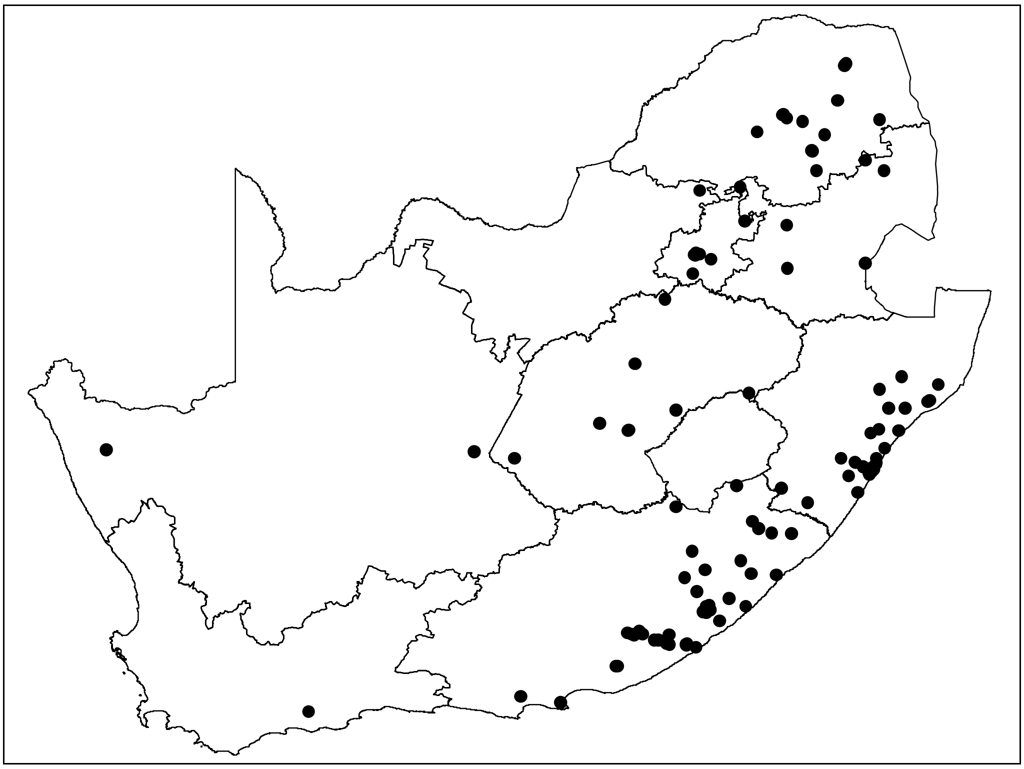
The number of schools which could be closed and merged, 117, is relatively low, representing just 0.2% of overall enrolment. This number is also very low in comparison to the total number of schools which could be considered small, namely 5,343. To compare, if the distance threshold had been raised from 2 km to 4 km, 261 schools would have been merged, in other words still a rather low number. What the analysis suggests is that despite the large numbers of small schools, this is perhaps a rational setup. The small schools that exist are mostly not close enough to other schools to make their closure desirable. Of course the analysis presented here is crude. In particular, the distance between learners' homes and the school has been greatly simplified. But the simplification occurs in two opposite directions. In the analysis presented here, obstacles posed by natural barriers such as rivers have not been considered, meaning that some schools which were merged in the simulation should in fact not be merged. On the other hand, the analysis has essentially assumed that learners live everywhere around a school, within a 2 km radius. Of course there would be many schools where the population would be concentrated on just one side, for instance the eastern side, of a school. Not taking this into account means the simulation has failed to close certain schools which probably could be closed. The net effect of all this, whether the 117 closed schools referred to Table 3 can be can considered too high or too low, seems impossible to tell without a detailed analysis of individual schools. Yet the general conclusion that closing and merging schools is not something that can be widely implemented in South Africa, probably stands.

Table 4 breaks the 117 closed schools down by province and the map in Figure 17 provides more details about their location. Clearly it is over-simplification to say that only one or two provinces ought to be concerned, at least to some degree, about the closure of unsustainably small schools.

Table 4: Provincial breakdown of closed schools

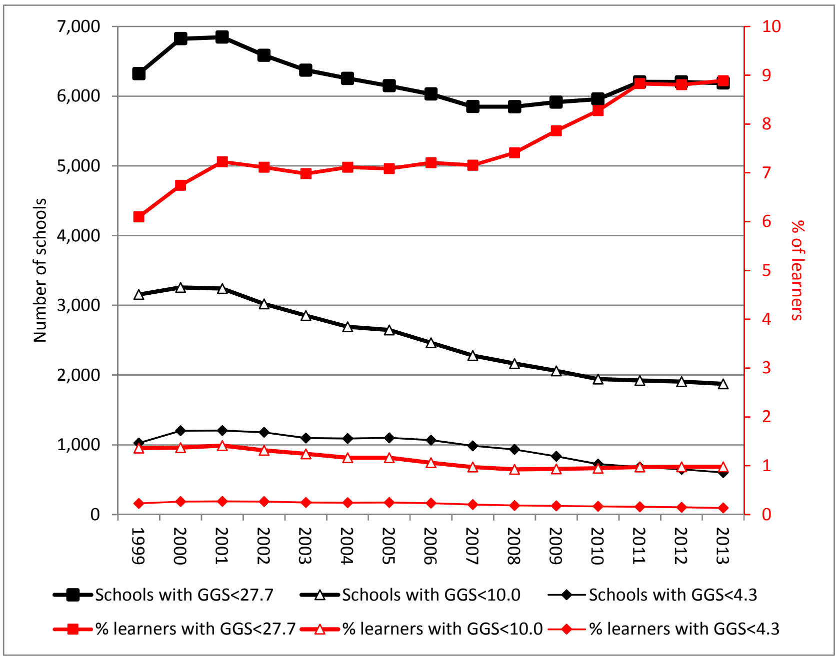
	Number of schools
EC	42
FS	8
GP	9
KN	31
LP	17
MP	6
NC	2
NW	1
WC	1
SA	117

Figure 17: Location of schools which could be closed



What has the trend been with respect to small schools over the period 1999 to 2013 at the primary level? The number of schools and percentage of learners falling below the thresholds discussed above, for each year in the 1999 to 2013 period, were calculated and are represented in Figure 18 below. If we start with the highest threshold, 83 learners across three grades, and convert this to the average grade group size (GGS), we obtain 27.7 (after division by 3). The uppermost curve in the graph indicates that the number of schools with fewer than 27.7 learners per grade (counting all grades in the range 1 to 7) declined over time, from just under 7,000 to just above 6,000. The increase in the years 1999 to 2000 could be the result of problems with the earlier data, specifically missing small schools. There seems no reason to believe that the number of small school *increased* suddenly between 1999 and 2000. The second curve in the graph, from the top, indicates that the percentage of learners at the primary level in small schools (defined as schools with an average grade group size less than 27.7) in fact increased. We can think of this as a reflection of the fact that small schools were not quite as small in, say, 2013, compared to 2007. Analysis of the data does in fact point to an absolute increase in the average grade group size of the small schools in question over the period 2007 to 2013. This explains why the percentage of learners in very small schools with a grade group size below 10.0 decreased (see the relevant curve in the graph). Given that the 27.7 threshold being considered here is the threshold below which multi-grade teaching became necessary (using the methods and assumptions explained earlier), one can conclude that the demand for multi-grade teaching would have *increased* between 2007 and 2013, despite the decline in the number of small schools. This is a pattern one may easily miss if the data are not examined carefully. In fact, an increase in the demand for multi-grade teaching as urbanisation occurs in South Africa should not completely surprise us. As population density in rural areas declines, an increasing number of schools will find that their enrolment levels fall below the thresholds at which it is no longer feasible for there to be one educator per grade. As indicated above, there is not much room available to close remaining small schools, without depriving learners who are in rural areas of the education they require. The number of schools where multi-grade teaching becomes necessary will almost certainly increase in future years, in particular if one considers that the teacher supply bottlenecks currently experienced cannot be resolved quickly.

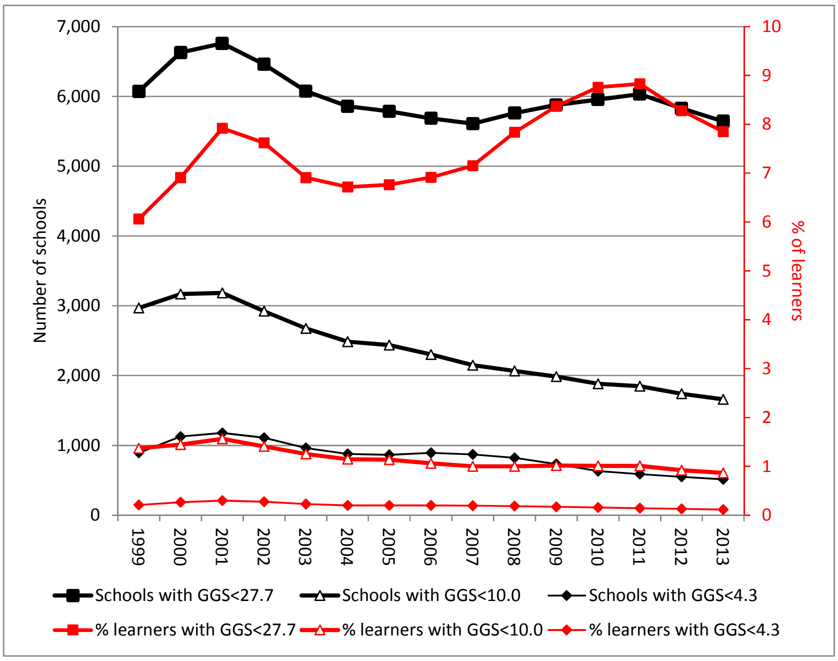
Figure 18: Small school trends grades 1 to 7 (1999-2013)



Source: Department of Basic Education, 2014.

The previous graph reflected small school trends considering learners in grades 1 to 7, though the thresholds for the determination of 'small' were drawn from an analysis covering just grades 1 to 3. If Figure 18 is recalculated using just grades 1 to 3 enrolment figures, the result is Figure 19 below. The two pictures are roughly similar, with one noteworthy exception. In the grades 1 to 3 analysis, the proportion of learners in schools requiring multi-grade teaching appears to drop after 2011. This is largely due to the enrolment increases in the lower grades brought about by demographic changes (see discussion in sections 3 and 4). Enrolments increased substantially in rural areas, meaning that if additional teachers were allocated to schools to retain historical learner/educator ratios, then the need for multi-grade teaching would indeed decline. Of course if historical learner/educator ratios were not maintained, then multi-grade teaching would continue as before.

Figure 19: Small school trends grades 1 to 3 (1999-2013)



Source: Department of Basic Education, 2014.

6 Inequities over time in the availability of teachers

This section focusses strongly on problems around the implementation and design of the post provisioning norms, given the prominence of this policy in the current debates. The focus is thus on how well provinces implement the norms, if current implementation patterns suggest there are problems with the norms themselves, and what poor implementation, or inequities in the distribution of publicly paid educators, means for education outcomes.

The section begins by explaining the parameters of post provisioning norms. It then explains a few Snap Survey data problems which complicate the analysis, but still make it possible. These problems are largely related to how Grade R teachers (who are mostly not distributed using the post provisioning norms) are counted in the Snap Survey.

Whilst the emphasis is on the distribution of publicly paid educators, how the picture changes when privately paid educators in public schools are counted is considered. For instance, Western Cape has exceptionally high learner/educator (L/E) ratios when only publicly paid educators are counted, but especially low (better) ratios when privately paid staff are also counted.

The post provisioning norms are simulated, and what each school *should* have in terms of the number of educators in 2013 is then compared to what it *actually* has. The larger the discrepancy between these two, the greater the problems relating to the post provisioning norms (and, for instance, the moving of excess teachers). It is found, using Snap Survey data, that one province exceeds by far other provinces in its ability to implement the norms. This province is Western Cape. Second in line is Gauteng, whilst three further provinces do a relatively good job (Northern Cape, Mpumalanga and Free State), whilst one province, Eastern Cape, stands out as particularly problematic. To illustrate, in Western Cape 2.3% of educators can be considered 'misplaced', against 6.3% in Eastern Cape.

The norms attach relatively low weights to grades 5 to 7 learners. Most provinces subvert this by effectively increasing the weights for these learners. It appears that provinces are responding to a real need, and that the problem lies with the policy. It may be preferable to allow provinces to legitimately adjust these weights, within reasonable limits, on condition that they justify any adjustments.

A further departure from the official norms is the use of generous thresholds to provide a second educator to schools which would otherwise be given only one educator. Such thresholds have clearly been implemented in KwaZulu-Natal, but not in Free State or Eastern Cape. Guidance on this matter (though perhaps not hard prescription) is probably needed in the norms, and this should take into account province-specific budget pressures.

Only three provinces were found to be distributing educators in a pro-poor manner, in accordance with the norms. These provinces are Free State, Northern Cape and Western Cape. In two provinces, Mpumalanga and KwaZulu-Natal, it is clear that *fewer* educators are employed in poorer schools, which clearly contravenes the post provisioning policy.

If one compares the primary and secondary levels, important province-specific patterns emerge. Gauteng stands out as the province with the highest L/E ratio difference between primary and secondary schools (37 for the former, 29 for the latter). KwaZulu-Natal, on the other hand is, relative to the standards set by the norms, more generous to primary schools than to secondary schools.

An analysis of the 'post provisioning problem' by district and by the primary-secondary split indicates that getting educators to 'follow' learners is a far greater challenge at the secondary than the primary level. What is particularly striking is the number of districts in Eastern Cape with a high proportion of secondary schools with *more* educators than they should have.

The relationship between under-staffing (in terms of a school's actual and prescribed L/E ratios) and learner performance was examined. The link between under-staffing and learner under-performance is more visible at the primary than the secondary level, which is possibly a reflection of the fact that younger learners are more dependent on face-to-face teaching, and thus more sensitive to having fewer teachers. Interestingly, the relationship between under-staffing and under-performance only comes through clearly in provinces which are relatively good at implementing the norms. This is possibly because provinces which are good at implementing the norms, are probably also good at implementing other policies affecting schools, so the absence of teachers is felt more strongly.

Contrary to what is often believed, the relationship between under-staffing and the remoteness of schools is rather weak. It is least weak at the secondary level, where four provinces display clearly the expected relationship of greater remoteness associated with under-staffing. These four provinces are Eastern Cape, KwaZulu-Natal, Limpopo and North West.

An historical analysis, using Snap Survey data from 2008 to 2013 was undertaken to establish the trends which led to the 2013 situation. Three provinces, Eastern Cape, Northern Cape and North West, have seen their correlation between educators and enrolments (adjusted using weights from the norms) decline, or deteriorate.

One key measure of a province's ability to manage the post provisioning processes is the lag between enrolment distribution patterns and educator distribution patterns. Ideally, there should be a lag of just one year, because the best a province can do is to distribute educators in one year based on the previous year's enrolments. The ideal lag of just one year is clearly present in Gauteng, Western Cape and Northern Cape (although, as mentioned previously, there is evidence that Northern Cape's situation has been deteriorating). The lag in Eastern Cape is clearly bad, at four years.

Eastern Cape's difficulty in implementing the post provisioning norms is partly related to the fact that the distribution of learners across schools is remarkably volatile in this province. This is surprising, as it is often assumed that more urban areas, where the close proximity of schools should make it easier for parents to switch schools, display the highest degree of enrolment instability. It was found that this assumption is in many respects ways false. Eastern Cape's enrolment volatility is partly linked to the prevalence of small schools in the province, but there are clearly other factors specific to Eastern Cape which also play a role.

An analysis of L/E ratios over the period 2003 to 2013 indicates that over the longer term the provisioning of educators to schools has improved substantially, at least insofar as L/E ratios have dropped, at the both the primary and secondary levels. At both levels the drop has been around 4.0 learners. This has been driven largely by higher numbers of educators in absolute terms. Clearly a declining overall L/E ratio would tend to make the allocation of educators easier as the removal of educators from certain schools is less likely to result in larger classes. However, despite the overall decline in the L/E ratio, from around 2010 the ratio has been rising slowly, at both the primary and secondary levels.

To conclude, the rather effective implementation of the wider post provisioning process in a few provinces suggests that in other provinces the key solutions would have to be, firstly, better technical capacity to implement the norms and, secondly, better ability to deal with the politics of teachers not wishing to move from schools with shrinking enrolments. The problem is not mainly one of policy design, though there are some aspects of the norms which require revisiting. Eastern Cape stands out as having particularly large capacity problems. Not only is capacity clearly weak, but Eastern Cape is one of three provinces where there has been a deterioration in capacity between 2009 and 2013.

In some ways, a 'low-hanging fruit' is to ensure that all primary schools are adequately staffed, given that the link between under-staffing and learner under-performance emerges as stronger at the primary level (something one should expect), and given that primary-level teachers seem to be easier to move around, so that they 'follow' learners. The secondary-level L/E ratio inequality problem appears more complex and more extensive (though the strong link to learner performance does not come through). Here more complex responses, including better use of remoteness incentives, appear to be necessary.

A further post provisioning 'low-hanging fruit', dealt with in section 7.3, is to eliminate barriers which currently make it difficult for educators to apply for and fill posts in different provinces. These barriers imply that matching the best educator to an available post is made a bit more difficult than it should be.

What is not a major 'low-hanging fruit', according to the analysis presented below, is the freeing up of budgets, possibly for the hiring of more educators, through a reduction in the dependence on small schools. Relying less on small schools, which have exceptionally low L/E ratios, does bring about some saving, but this saving is tiny. Specifically, a model was run with two scenarios, both with the same total enrolment and the same L/E ratios per school size. The only difference was that in one scenario small schools played a larger role. The school size distributions followed the actual pattern in 1999, and the actual pattern in 2013, when the country was more urbanised. The 2013-like scenario was found to cost just 0.6% less than the 1999-like scenario, with respect to educators.

This section deals with teacher supply in basic terms, largely the total number of educators available per school relative to learners. Whilst such an approach ignores the quality of teachers, and their specialisations, understanding the basics of teacher availability is important and, as will be shown, not as straightforward as may initially be believed. In particular, it is appropriate and equitable for the ratio of learners to educators to vary according to school size and the socio-economic advantage of the school community. After a description of the post provisioning norms in section 6.1, section 6.2 establishes an approach for assessing inequities in the availability of teachers in 2013. An approach for simulating the post provisioning norms to arrive at the number of educators each schools *should* have, is explained. Some kind of equity standard is obviously needed if inequities are to be measured. The point of departure in the analysis of actual equity is thus an assumption the official policy represents a good yardstick for equity. Yet the analysis includes some critiquing of the existing post provisioning norms. Section 6.3 turns to trends over time. Here the focus is partly on the kinds of lags seen in getting teachers to follow learners in the system. Section 6.4 describes long-range learner/educator trends, from 2003 to 2013. Section 6.5 addresses important questions around possible savings one might expect from a schooling system with a greater proportion of learners in larger and urban schools. In other words, economies of scale relating to the need to staff small schools differently are examined. Finally, section 6.6 analyses the degree to which the distribution of enrolments across schools is stable over time, a matter which influences the difficulty of allocating educators to schools fairly.

Throughout section 6, the emphasis on the provincial level is strong, given that it largely provincial practices, which may or may not follow national policies to the letter, which determine how equitably teachers are distributed.

The policy questions surrounding the post provisioning norms and related policies can be summarised as follows:

- To what extent do provinces comply with the norms, and how has this changed over time? To what extent do better complying provinces set a yardstick for what can be achieved in other provinces, given different provincial contexts?

- To what extent do provinces comply with the norms in a delayed manner, in other words in a manner whereby educators do follow learners, in line with the rules, but with a lag?

- To what extent might a lack of compliance, for instance with respect to specific school grades, represent sensible choices made by provinces, and hence an indication of where the existing policy must change?

- What do geographical patterns of under-staffing (and even over-staffing) suggest regarding policy solutions to ensure a more equitable distribution of educators?

- What do correlations between learner performance and under-staffing, conditioned on factors such as socio-economic status, suggest regarding policy solutions?

- To what extent will the urbanisation process outlined earlier in this report allow for economies of scale?

6.1 The post provisioning norms

What is commonly referred to as the 'post provisioning norms' are the national policy determining how many educator posts each school is entitled to for the purposes of grades 1 to 12 schooling. Though there is an option within the post provisioning norms to calculate Grade R educators, this is an option which only two provinces use[24]. Grade R funding and staffing is moreover dealt with comprehensively through a separate policy· The separation between Grade R and the other grades is by no means always maintained in practice. The grades 1 to 12 norms do not break down entitlements by rank (for instance how many deputy principals a school should have) or by subject (for instance how many mathematics teacher a school should have). Essentially the policy stipulates that a province should on an annual basis determine how many educator posts it can afford, and that it should then calculate school-level post entitlements using weightings which are based largely on enrolments by grade and subjects taken by grades 10 to 12 learners.

The 2002 policy is titled 'Amendment of regulations for the distribution of educator posts to schools in a provincial department of education' and is published as Government Notice 1451 of 2002. This policy amends an earlier similar policy, published in Government Notice 1676 of 1998. The 2002 amendments dealt mainly with the introduction of a pro-poor element to the distribution formula. In 2005, a further revision was distributed to provincial departments, though this revision was not published as a government notice, and hence its status has been the matter of debate. The 2005 policy, titled 'Post distribution model for the allocation of educator posts to schools', is available on the Department of Basic Education website. The 2005 policy differs from the 2002 policy mainly insofar as it aligns weightings producing post entitlements at the grades 10 to 12 level to the revised curriculum, the National Curriculum

[24] North West is reported to have used the Grade R option for many years, whilst Limpopo began using it in 2013.

Statement. The 2002 weighting rules, with 2005 amendments, where applicable, are summarised in the following table.

Table 5: Parameters of the post provisioning norms

	Unit	*Weight from the policy*	*Leeway for provincial adjustment?*
1	Grade 1 to 4 learner	1.166 **1.190**	No
2	Grade 5 to 6 learner	1.020 **1.042**	No
3	Grade 7 learner	1.103 **1.126**	No
4	Grade 8 to 9 learner	1.229 **1.2426**	No
5	Subject taken by Grade 10 to 12 learner, seven per learner	Varies from 1.134 to 0.179. At the learner level, varies from 3.94 to 1.25 (possible subject combinations taken into account). **Various subject weightings which translate to a range of 1.30 to 3.05 at the learner level.**	Yes (reweighting done at subject level, no restrictions)
6	School	10 to 20	Yes, but within 10 to 20 range.
7	School, if it offers at least one primary grade (1 to 7) and at least one grade in range 10 to 12.	10 to 20	Yes, but within 10 to 20 range.
8	Grade in range 1 to 7	2	Yes, 2 can be increased up to limit of 4.
9	Grade in range 8 to 12	3	Yes, 3 can be increased up to limit of 5.
10	Grade in range 1 to 7 with a second language	4	Yes, 4 can be increased up to limit of 6.
11	Grade in range 8 to 12 with a second language	6	Yes, 6 can be increased up to limit of 8.
12	Poverty quintile of school (1 to 5)	Overall inflation of weights ranging from 9.2% (poorest) to 1.3% (least poor). **Slightly more equal spread, creating a narrower range of 7.9% (for poorest) to 1.3% (least poor).**	No.
13	**Disabled learner (if counted here, cannot be counted for any other weighting purposes)**	**3 to 6**	**Yes, inflation up to 20% allowed.**

Note: Text in bold refers to 2005 amendments to the 2002 policy. Numbers in the first column are specific to this table and are not found in the policies.

The DBE has committed itself to a major revision of the post provisioning system, but most of these changes relate not so much to the formula which is the central concern of the actual post provisioning norms, but to related matters such as the overall supply of appropriately qualified teachers, the ease with which teachers are moved from one school to another, addressing excessive class sizes, and incentives to teach in hard-to-teach schools[25].

6.2 Teacher supply inequities in 2013

As a prelude to the main analysis, two sets of assumptions to be used in relation to 'practitioners' in the Snap Survey data, had to be resolved. Practitioners are in general understood to be Grade R teachers who are not fully qualified educators, though the term is somewhat loosely applied, so it might occasionally be used to refer to a well-qualified teacher teaching Grade R. Practitioners need to be taken into account when learner-to-educator ratios

[25] Department of Basic Education (2015), African National Congress (2013).

are calculated as practitioners essentially assume the role of educators. They teach learners in a class. However, for the analysis presented below there was also a need to differentiate between Grade R teachers (generally 'practitioners') and grades 1 to 12 teachers (generally 'educators') insofar as the analysis of the post provisioning norms focusses on how these norms operate at the grades 1 to 12 level only[26].

Before 2010, the Snap Survey did not cater for the category 'practitioner' and a look at the questionnaire suggests schools would have counted Grade R practitioners as 'educators' prior to 2010. This precedence supports the argument that the way the term 'practitioner' is currently used in the Snap Survey process could be unpredictable. To provide a sense of magnitudes, in half of public schools in 2013 which offered Grade R, practitioners were reported to be employed. By implication, in the other half of these schools Grade R was being taught by people reported to be 'educators'.

The first set of assumptions that had to be resolved has to do with a problem whereby practitioners were clearly under-reported by schools in certain tables in the Snap Survey questionnaire[27]. If the data were used without adjustments, the number of practitioners would be under-stated. Detailed analysis of the data led to adjustments which yielded an overall count of practitioners for public schools in 2013 of 18,353 (without any adjustment, the number would be 11,538). Similar adjustments were applied to the 2010 to 2012 data. The adjustments are obviously driven by particular assumptions around the kinds of errors which occurred when the Snap Survey questionnaires were completed[28].

The second set of assumptions related to how to define a *publicly* paid Grade R practitioner. In the survey questionnaire, a practitioner is either paid by the state, or by the school governing body (SGB) or paid by both of these sources. The first and third categories were automatically considered to be publicly paid. The question was what to do with the SGB-paid category considering that policy states that the state can pay the SGB in order to pay practitioners. Of the 18,353 practitioners referred to above, 3,393 were paid by the SGB in 2013. Of these, 1,867 practitioners were working in schools where there were no SGB-paid 'educators', meaning relatively disadvantaged schools. It was assumed that these 1,867 practitioners could be added to the number of practitioners considered to be funded by the state, as the policy focuses on public funding of Grade R practitioners in less advantaged schools[29].

Key values relating to educators and practitioners in 2013, by province, are given in the next table. The practitioner totals (columns D and E) are those obtained after all adjustments mentioned above have occurred. The 'effective L/E' ratio, which counts all staff members who teach, is perhaps the most significant of the three ratios presented, in terms of understanding the overall human resourcing of schools. If only publicly paid employees are counted (the second ratio), overall the L/E ratio rises by two learners at the national level. The ratios in the last column are difficult to interpret, as they exclude practitioners, and who is counted as a practitioner and who is counted as an educator when it comes to Grade R is a

---

[26] As pointed out in section 6.1, two provinces do use the post provisioning norms to calculate the number of Grade R teachers. To reduce the complexity of the analysis done for the current report it was decided to focus on post provisioning for just grades 1 to 12 across all provinces (as opposed to focussing on grades R to 12 in the case of two provinces, and grades 1 to 12 in the case of seven provinces).

[27] Specifically, since 2010, and at least up to 2013, practitioners employed by the state, and practitioners employed by the school governing body (SGB) had to be entered *twice*, in different tables. However, practitioners paid by *both* the state and the SGB had to be entered just *once.* The questions clearly resulted in some confusion. Analysis of the data reveals that in many schools, practitioners who should have been entered twice were entered just once.

[28] The author of this report can be contacted for a Stata .do file containing details of the adjustments.

[29] Government Notice 26 of 2008.

somewhat arbitrary matter. Yet knowing even roughly the number of fully qualified educators, relative to learners, is relevant (the ratio would be rough not just because of problems around the description of Grade R teachers, but also because 'educators' in other grades may also be under-qualified).

Table 6: Provincial breakdown of 2013 educator and learner numbers

	Learners all grades (A)	State-paid educators (B)	SGB-paid educators (C)	Publicly paid practitioners (D)	Privately-paid practitioners (E)	Number of SGB-employed practitioners considered publicly paid	Effective L/E ratio (A / (B + C + D + E))	Public-only L/E ratio (A / (B + D)	Public-only educator-only L/E (A / B)
EC	1,881,605	59,265	3,992	4,014	151	305	27.9	29.7	31.7
FS	649,806	22,465	1,468	203	119	6	26.8	28.7	28.9
GP	1,899,542	53,997	5,971	4,440	391	266	29.3	32.5	35.2
KN	2,798,975	87,800	4,569	3,937	188	471	29.0	30.5	31.9
LP	1,662,106	54,249	906	492	7	77	29.9	30.4	30.6
MP	1,025,859	32,259	1,268	1,556	128	102	29.1	30.3	31.8
NC	279,445	8,264	608	680	75	113	29.0	31.2	33.8
NW	773,040	23,716	1,564	298	94	43	30.1	32.2	32.6
WC	1,005,466	27,285	5,387	1,207	373	484	29.4	35.3	36.9
SA	11,975,844	369,300	25,733	16,827	1,526	1,867	29.0	31.0	32.4

A problem for the analysis which is somewhat separate from the matter of the two assumptions discussed above, is the fact that around half of schools with Grade R in 2013 reported having no practitioners. This problem resulted in a number of checks being run on the modelling presented below, to establish whether the data problem was driving the policy conclusions. Details on schools with Grade R but with no practitioners are presented in the next table. One pattern that is noteworthy is that in Gauteng and Western Cape it seems to be mainly better off quintiles 4 to 5 schools which classify in the Snap Survey their Grade R teachers as 'educators', as opposed to 'practitioners'. This is to be expected, given that Grade R teachers in these two provinces would tend to be better educated.

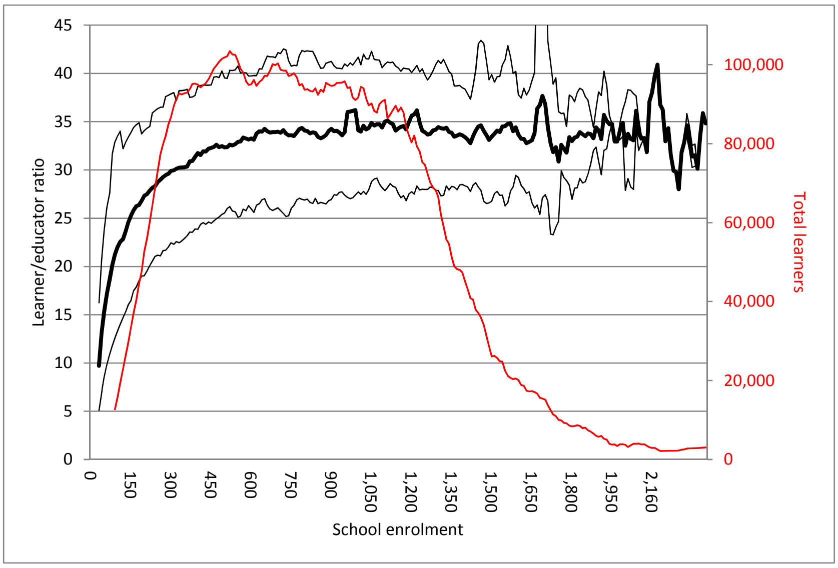
Table 7: Details on Grade R but no practitioners 2013

	All schools with Grade R (A)	Schools in A with no 'practitioners' (B)	% of learners in column B schools over learners in column A schools	% of Grade R learners in column B over Grade R learners in column A	% of learners in column B who are in quintiles 4 or 5
EC	4,487	1,237	30	28	12
FS	584	564	97	96	17
GP	1,286	171	13	13	64
KN	3,963	2,500	60	58	22
LP	2,328	1,949	83	83	9
MP	995	444	38	38	12
NC	361	65	16	15	42
NW	915	876	96	96	16
WC	931	497	52	51	66
SA	15,850	8,303	51	51	22

The following graph illustrates the distribution of what was referred to in the previous table as the 'public-only L/E ratio', according to school size. The fact that smaller schools, specifically schools with a total enrolment of less than about 700, experience a more favourable L/E ratio is clear. Above this threshold, the mean ratio does not change much.

Deviations from the mean can be large, however, as indicated by the 10th and 90th percentile curves. For instance, for schools in the size range 1,000 to 1,100 learners, the L/E ratio range, for the 'middle' 80% of schools, is 28 to 41. How these variations work relative to, for instance, province and poverty quintile, are key questions explored below.

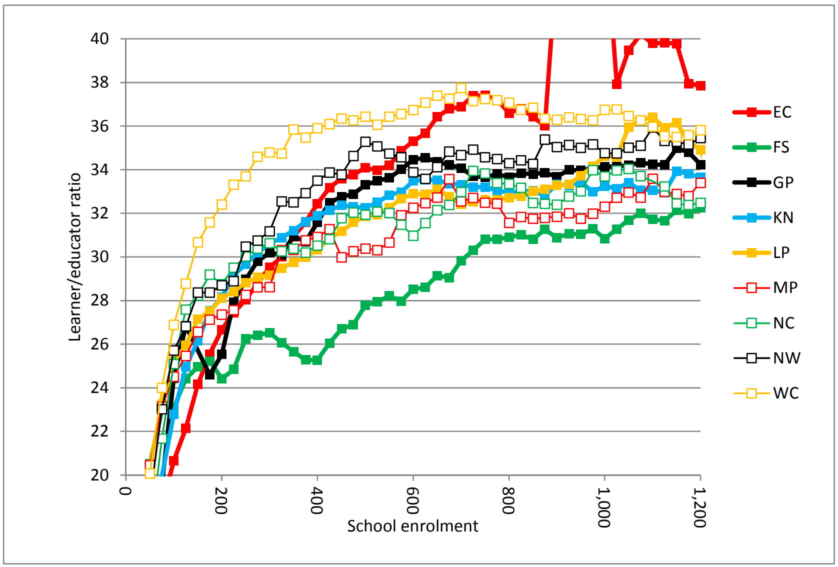
Figure 20: Learner-to-educator ratio distribution 2013



Source: Department of Basic Education, 2014.
Note: This graph counts only staff (educators and practitioners) paid by the state, and only in public ordinary schools. The red curve should be read against the right-hand axis. The thick black curve represents the mean ratio for each group of schools, where schools are grouped according to total learners rounded to the closest multiple of 10 ('Total learners' in the case of the red curve is also calculated using these groups). The thin black curves represent the 10th and 90th percentiles within each group.

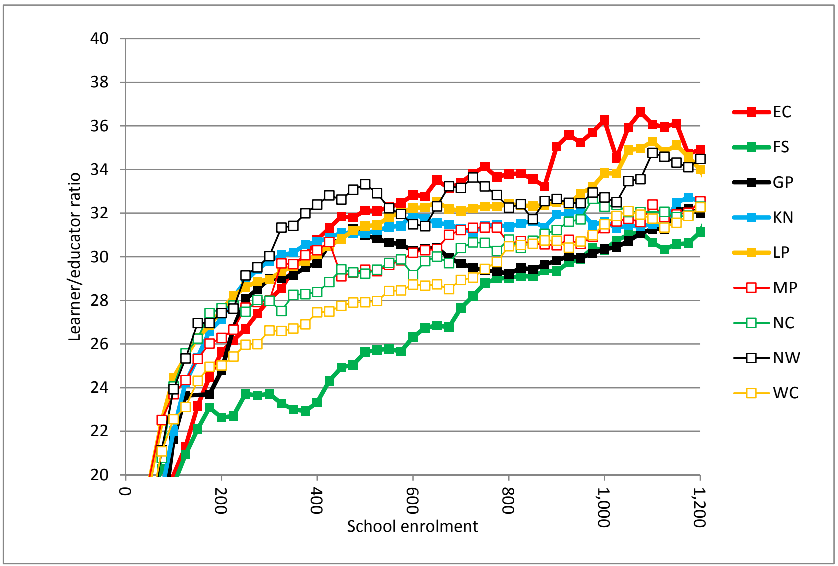
The next graph illustrates the shape of the mean 'public-only' L/E ratio distribution, relative to total enrolment, for each province. What is clear is that Free State schools enjoy exceptionally low L/E ratios, and that counting only publicly paid staff results in high L/E ratios for Eastern Cape and Western Cape. However, as can be seen in the following graph (Figure 22), once privately paid educators in public schools are taken into account, Western Cape emerges as a relatively well-off province in terms of the L/E ratio. The reason for the unusual shape of the Gauteng curve in Figure 22 is that larger schools tend to be more advantaged schools with a greater ability to raise funds to pay for additional teachers.

Figure 21: Provincial L/E ratios counting only publicly paid staff



Source: Department of Basic Education, 2014.
Note: Grade R practitioners have been counted as educators. Only public ordinary schools are considered.

Figure 22: Provincial L/E ratios including privately paid staff



Source and notes same as for previous graph.

The problem with the type of analysis seen in the previous two graphs is that it does not take into account school factors other than overall school size which, according to policy, should influence the L/E ratio, such as the grades offered by the school. To obtain a better sense of where schools in the country are advantaged or disadvantaged, the educators each school *should* have were calculated, using a simulation of the post provisioning norms. The difference between the actual number of educators in a school and the number of educators a school should have would then provide a more meaningful picture of the L/E ratio inequalities. Moreover, using a simulation which limited the available educators to the total actual educators *per province* would provide a picture of which provinces were most successful at maintaining within-province equality, but also at complying with the national policy.

In Table 8 below two different measures of the degree to which provinces complied with the norms in 2013 are presented. The adjusted R^2 measure is derived from a regression analysis, where the actual number of educators (excluding practitioners) was regressed on twelve school-level weights defined by the norms. The details of this analysis appear in Appendix B. An R^2 value of 1.000 would mean that the twelve weights predicted completely the actual number of educators per school. The best R^2 value is that of Western Cape, the worst that of Eastern Cape. This points to greater policy compliance in the case of the former province, and greater within-province equity in terms of the L/E ratio. Compliance here should be understood broadly, as encompassing both the formal declaration of posts, and the filling of these posts, and ensuring that educators in excess of declared posts in a school were moved. The same R^2 values would be obtained, whether one used the 2002 or 2005 versions of the post provisioning norms (though slope coefficients, discussed in Appendix B, would change).

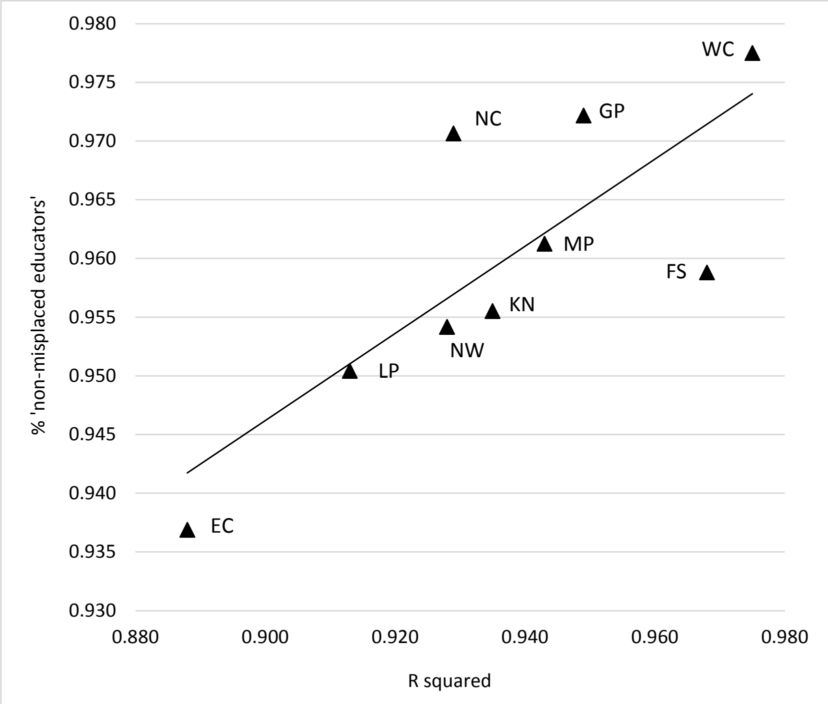
The second column in Table 8 presents a more intuitive indicator. Here the actual and predicted educators per school were compared. Where the actual figure exceeded the predicted figure by more than 0.5 (keeping in mind that the predicted figure would mostly be a non-integer), the difference between the two was considered a measure of 'misplaced educators', or educators who should have been working at other schools. The values in Table 8 refer to *non*-misplaced educators as a percentage of all educators. For instance, in Western Cape only 2.3% of educators were misplaced, against 6.3% in Eastern Cape.

Table 8: Measures of compliance with the post provisioning norms

	Adjusted R^2	% 'non-misplaced educators'
EC	0.888	93.7
FS	0.968	95.9
GP	0.949	97.2
KN	0.935	95.6
LP	0.913	95.0
MP	0.943	96.1
NC	0.929	97.1
NW	0.928	95.4
WC	0.975	97.7

As one might expect, the two indicators of policy compliance reflected in the previous table provide very similar rankings of provinces. Below, in Figure 23, the two indicators are graphed. It is clear that faring somewhat better than Eastern Cape are Limpopo, North West and KwaZulu-Natal (in that order), with Gauteng being the closest to achieving the Western Cape level of compliance.

Figure 23: Measures of compliance with the post provisioning norms



A key conclusion one can draw is that the 'post provisioning problem' is more an implementation problem than a policy design problem. Whilst there are undoubtedly problems with the design of the norms, Western Cape and to some extent Gauteng suggest they are implementable, with the resultant misallocation of educators being acceptably low. Mpumalanga is perhaps especially interesting because, like most provinces, it includes large rural areas. Yet this province is able to achieve a situation where just 3.9% of educators are misplaced. This is not excessive, in particular if one keeps in mind that a province, at best, allocates educators in line with the *previous* year's enrolment figures (the analysis presented here compares actual educators in 2013 to weights calculated using enrolment data also from 2013, in effect making it virtually impossible for a province to achieve 100% compliance). The data suggest a key challenge is to get the four provinces clearly performing worse than Mpumalanga to move to the level of this province, through better use of data and better human resources management.

The coefficients for the various school-level weights point to possible problems in the existing norms (Table 26 in Appendix B). In particular, the general pattern is for provinces to be more generous to grades 5 to 7, relative to other grades, than what is implied by the policy. The policy in fact weights grades 5 to 7 less than the other grades, so what provinces appear to be doing is to 'correct' this so that grades are treated more equally. Provinces are not meant to be doing this, as the weights in question are not allowed to be adjusted. Given the extent of the 'correction', it appears provinces are getting around a design flaw in the policy. The best may be to give provinces the official leeway to deviate, and then to be open about the deviations they made, and the reasons for this. The same finding regarding grades 5 to 7 emerges if one excludes the 8,300 or so schools where Grade R teachers appear to have been counted as 'educators'.

The post provisioning norms should probably also explicitly allow provinces to set their own enrolment thresholds permitting a one-educator school to become a two-educator school. As seen in Table 28 of Appendix B, currently provinces with large numbers of small schools approach this matter rather differently, presumably in response to province-specific budget pressures.

The analysis in Appendix B indicates that for four provinces, patterns are most consistent with the 2002 norms, as opposed to the 2005 norms, at least as far as the weighting of enrolment by grade is concerned (the provinces are Gauteng, Mpumalanga, Northern Cape and North West). The matter of whether the 2002 or the 2005 norms should be used by provinces, and which of the two provinces actually use, has been a point of concern[30]. The remaining five provinces appear to have used the weights in the 2005 policies (see Table 9 below). Only three provinces displayed patterns in the data which were clearly consistent with the implementation of the pro-poor weights, and in all three provinces the distribution was more consistent with the 2005 weights than the 2002 weights. The three provinces are Free State, Northern Cape and Western Cape. It seems possible the remaining six provinces are unable or unwilling to implement this important element of the policy.

Table 9: Version of policy provinces comply most with

	Grades 1 to 9 weights	Poverty weights
EC	2005	
FS	2005	2005
GP	2002	
KN	2005	
LP	2005	
MP	2002	
NC	2002	2005
NW	2002	
WC	2005	2005

Note: The year values come from Table 27 in Appendix B.

The next table aims to represent whether learners are better or worse off, relative to national averages, in terms of the L/E ratio they experience, where 'E' includes only publicly paid educators. In producing Table 10 below, a predicted L/E ratio per school was calculated. This involved a few steps. Firstly, predicted educators per school, or the number of educators each school *should* have, were calculated using one *national* regression with the post provisioning weights. In effect this meant using a national standard. The resultant number of predicted educators per school was a non-integer value, such as 10.4 educators in a school. Integer values were obtained by rounding non-integers upwards. However, this resulted in a total number of predicted educators which was higher than the actual number of educators. To deal with this problem, non-integer values were adjusted downwards, by a common factor, until a factor was found whose end result was the correct total number of predicted educators (in other words a total equal to the actual number of educators overall). Two L/E ratios were calculated per school, one using the actual number of educators in each school, and one the predicted, or ideal, number of educators per school. The first L/E ratio minus the second one was calculated for each school. Table 10 displays the average difference per quintile and province cell, where schools were weighted by learners.

[30] See for instance Deloitte (2013). It should be noted that what policy provinces claim to be using may be at odds with the patterns emerging from the data. For instance, according to the Deloitte (2013: 29) report, Western Cape claimed to be implementing the 2002 norms, yet according to Table 9 the patterns in the data are closer to the 2005 weights. In the final analysis this matter is probably of less importance than whether provinces are able to achieve an equitable distribution of educators, regardless of the precise equity standard used (the differences between the 2002 and 2005 weights are actually fairly small).

Table 10: L/E ratio advantage by province and quintile 2013 (public E)

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Overall	Indicator of variance
EC	2.1	1.0	0.9	1.7	1.7	1.4	4.5
FS	-3.4	-4.1	-3.6	-3.7	-2.3	-3.5	3.9
GP	1.1	0.5	0.5	0.7	1.1	0.7	3.1
KN	0.0	-0.4	-0.3	-1.2	-0.6	-0.4	3.3
LP	-1.2	-1.1	-1.5	0.5	0.4	-1.1	3.8
MP	-1.2	-1.3	-1.5	-1.6	-1.8	-1.3	2.8
NC	0.7	-0.6	0.5	0.2	1.8	0.4	2.4
NW	0.0	-0.1	0.3	-0.2	2.4	0.1	3.4
WC	2.1	1.4	1.5	2.3	3.8	2.5	3.6
SA	0.0	-0.5	-0.1	0.1	1.1	0.0	3.5

Note: This table compares the L/E ratio schools actually experience against the L/E ratio schools would experience if the same approach for distributing educators, based on the official post provisioning norms, were applied across all schools in the country. The colour-coding produces green for better (low) L/E ratios, relative to what a school should have, and red for worse (high) ratios, relative to what one would expect the norms to produce. Values in all cells were reduced by 0.6 in order to produce an overall national statistic of 0.0. Why would the original national statistic be 0.6 despite the total number of learners and educators being identical for both L/E ratios? This is because even with weighting of schools by total enrolment, the national statistic will always exceed 0.0, unless every school experiences exactly the same L/E ratio. In order to avoid distortion of the statistics by what were clearly outliers, schools where the L/E ratio difference was less than -10 or greater than 20 were excluded. Overall this meant roughly 2% of school observations were dropped. This trimming also occurred for the next three tables.

To illustrate, Free State as a whole, and schools within each of the quintiles in Free State, enjoy exceptionally low L/E ratios, relative to what the post provisioning norms would predict, and assuming (unrealistically) that post provisioning occurred *nationally*, so the norms were used to distribute a central 'pot' of educators across the country. This national approach is followed here to allow for comparison across provinces. In a subsequent analysis, the more realistic approach of applying the norms separately to provincial 'pots' is followed. In contrast to Free State, Western Cape has relatively high (disadvantageous) L/E ratios. In many ways, Table 10 reflects what was seen in Figure 21 above. One important check made possible by Table 10 is whether provinces are treating poorer schools more favourably when it comes to educator staffing. It is clear that at least at the national level quintile 5 schools are indeed receiving lower numbers of publicly paid educators than the other quintiles, which is broadly correct and in line with the norms (the quintile 5 difference at the national level in Table 10 is 1.1, a higher value than for the other quintiles, meaning a higher, or 'worse', L/E ratio). There are noteworthy patterns in the across-quintile distribution of the statistic within specific provinces. In KwaZulu-Natal and Mpumalanga the *lowest* (best) L/E ratios are those of the better off quintiles 4 and 5 schools. This should be a matter of concern. However, it should also be noted that despite these irregularities, poorer schools in Mpumalanga are still well staffed relative to national patterns, simply because overall Mpumalanga's L/E ratio is rather low.

How is it possible for Eastern Cape to enjoy a relatively good (low) L/E ratio in the second-last column of Table 6, yet experience relatively high L/E ratios according to Table 10? This is mainly because Eastern Cape experiences a particularly large burden of many small schools (see Figure 13) and should thus ideally have a much lower overall L/E ratio than the rest of the country if schools are to be treated like similar schools in other provinces, and in line with the post provisioning norms. This is a burden that the equitable share formula, which distributes public revenue across provinces, does not take into account.

The final column of Table 10 provides the mean across the absolute differences between the actual and ideal L/E ratios at each school. The higher this value, the more individual schools deviate from the predicted, or ideal, L/E ratio. Clearly, Eastern Cape performs worst against this indicator. This would largely be due to the fact that within each of the cells (or quintiles), there is considerable inequality with respect to the distribution of educators.

What does change the picture substantially is the inclusion of privately employed educators. This can be seen in the next table. Quintile 5 schools are clearly in a much more favourable situation than other schools, across all provinces. On average, these schools enjoy an L/E ratio that is 7.3 lower than what they would have if there were only publicly paid educators and these were distributed nationally according to the post provisioning norms. But what is also interesting is that even in the poorer quintiles more favourable L/E ratios are found when privately paid educators are counted. In fact, of the approximately 25,000 privately paid (or school governing body paid) educators in public schools in 2013, 20% worked in quintiles 1 to 3 schools. But whereas the percentage of all educators who are privately paid was 28% and 8% respectively for quintiles 5 and 4, in quintiles 1 to 3 this figure was between 1% and 2%.

Table 11: L/E ratio advantage by province and quintile 2013 (public and private E)

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Overall	Indicator of variance
EC	1.1	0.1	-0.2	-3.9	-10.3	-0.3	4.3
FS	-4.2	-4.4	-3.9	-6.6	-9.3	-4.9	5.0
GP	1.0	0.4	-0.1	-1.5	-8.0	-2.0	4.2
KN	-0.2	-0.7	-0.6	-2.7	-6.5	-1.6	3.9
LP	-1.3	-1.2	-1.8	-0.2	-7.8	-1.5	3.9
MP	-1.5	-1.5	-1.8	-3.5	-9.5	-2.2	3.4
NC	0.2	-1.0	-0.2	-0.7	-6.2	-1.1	2.9
NW	-0.5	-0.5	-0.6	-4.5	-9.4	-1.6	3.9
WC	0.1	-0.1	0.0	-0.8	-5.7	-2.2	4.3
SA	-0.5	-0.9	-0.8	-2.2	-7.3	-1.7	4.0

Note: Original values were reduced by 0.6, as for Table 10, so that statistics would be comparable across the two tables.

L/E ratio problems should be understood in terms of the level of the school. In Table 12 below one thing that stands out is how very poorly schools at the primary level are staffed in Gauteng, relative to the secondary level. In KwaZulu-Natal the opposite problem exists, namely an under-staffing of the secondary level co-existing with a relative over-staffing of the primary level. These patterns are also seen if one focusses on the raw actual L/E ratios. For instance, Gauteng experiences by far the largest difference between the primary and secondary L/E ratios, this difference being 8.0 (36.6 for primary against 28.7 for secondary). The raw L/E ratios can be seen in Table 29 in Appendix B.

Table 12: L/E ratio advantage by school level (public E)

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Overall	Indicator of variance
Schools with learners anywhere in grades 10 to 12 and nowhere in grades 1 to 4 (5,834 schools)							
EC	4.7	2.2	0.8	0.7	3.2	2.0	4.5
FS	-5.6	-4.6	-4.7	-5.4	-2.7	-4.7	5.1
GP	0.4	-2.5	-2.2	-2.0	-2.2	-1.9	4.1
KN	2.4	1.4	1.1	-0.7	0.1	1.0	3.7
LP	0.3	0.1	-0.7	2.0	-0.1	0.1	4.1
MP	-1.7	-1.4	-1.3	-2.5	-2.3	-1.6	3.2
NC	-0.3	-1.8	-0.2	0.4	1.6	-0.1	2.6
NW	0.0	-0.1	-0.9	-1.4	-0.2	-0.7	4.0
WC	3.4	1.9	2.2	3.4	3.9	3.2	4.2
SA	0.6	0.0	-0.2	-0.6	0.3	0.0	4.0
All other schools, generally primary and combined primary-secondary schools (18,186)							
EC	1.8	0.7	0.9	2.2	0.7	1.2	4.3
FS	-2.6	-3.9	-3.0	-3.0	-2.1	-3.0	4.9
GP	1.5	2.0	1.9	2.5	3.1	2.3	4.2
KN	-1.1	-1.4	-1.3	-1.5	-1.2	-1.3	4.0
LP	-2.2	-2.1	-2.2	-1.1	0.7	-2.0	3.8
MP	-1.0	-1.2	-1.5	-1.2	-1.6	-1.2	3.4
NC	1.1	-0.2	0.7	0.2	1.9	0.6	3.1
NW	0.0	-0.2	0.7	0.7	4.4	0.4	3.9
WC	1.8	1.2	1.0	1.8	3.8	2.2	4.3
SA	-0.2	-0.7	-0.1	0.6	1.6	0.0	4.1

Note: In this table the values of each sub-table are reduced to produce overall values of 0.0.

Does the picture seen in the bottom half of Table 12 change substantially if one excludes schools with Grade R but no practitioners? It appears as if the changes are small enough not to invalidate or change the conclusions drawn above. The bottom half of Table 12 is reproduced in Table 13 below using data from only the 9,907 non-secondary schools which did not display the Grade R practitioner anomaly. Differences between the two sets of figures would not only be due do data problems. It is also possible, for instance, that schools in Gauteng which are better staffed are more inclined to report their Grade R teachers as being 'educators' (and not 'practitioners'). This could explain some of the improvement for Gauteng seen in Table 13 relative to Table 12. To what extent it would cannot be ascertained, at least not using the Snap Survey data.

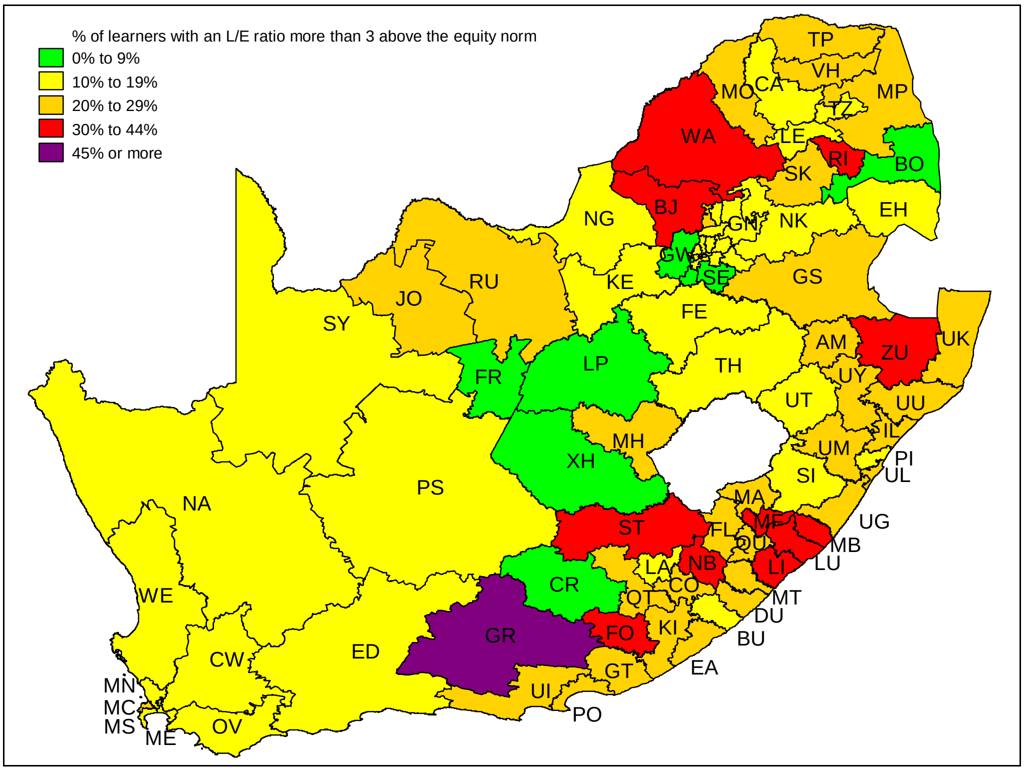
Table 13: Primary L/E ratio advantage with Grade R problem removed (public E)

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Overall	Indicator of variance
EC	1.5	0.0	0.7	2.5	0.0	0.9	4.7
FS	-1.8	-3.1	-3.5	-4.2	-2.1	-2.7	4.8
GP	0.4	1.3	0.9	1.5	2.2	1.3	4.0
KN	-0.8	-0.6	-1.7	-2.5	-1.8	-1.5	4.2
LP	-2.0	-2.5	-2.6	-1.9	1.8	-2.2	3.9
MP	-1.1	-1.0	-2.0	-1.8	-2.9	-1.3	3.1
NC	-0.1	-1.3	0.6	-0.1	0.9	-0.2	3.2
NW	-0.7	-0.7	-0.8	0.6	5.9	-0.3	4.2
WC	0.7	0.1	0.1	1.1	2.5	1.1	3.7
SA	0.1	-0.4	-0.2	0.1	0.9	0.0	4.0

The next step in the analysis was to obtain whole (integer) numbers of educators per schools for 2013, using a province by province simulation of the post provisioning norms. Essentially the national approach described above was repeated for each province. One difference in the method was that the apparent total enrolment threshold applicable within a province to add a second educator to one-educator schools was implemented, even if this is officially not a part of the post provisioning norms. Thus, for instance, any school in Limpopo which according to the norms should have just one educator, but which also exceeded the provincial threshold, which is 12 learners in the case of Limpopo, was given a second educator (see Table 28). The simulation still ensured that, within a province, the total actual number of educators and the sum of all school-level educator entitlements produced the same grand total. The correlation between the educators a school should have and the educators it did have was, as one would expect, high and the implied provincial compliance rankings were almost identical to what one sees in Table 8.

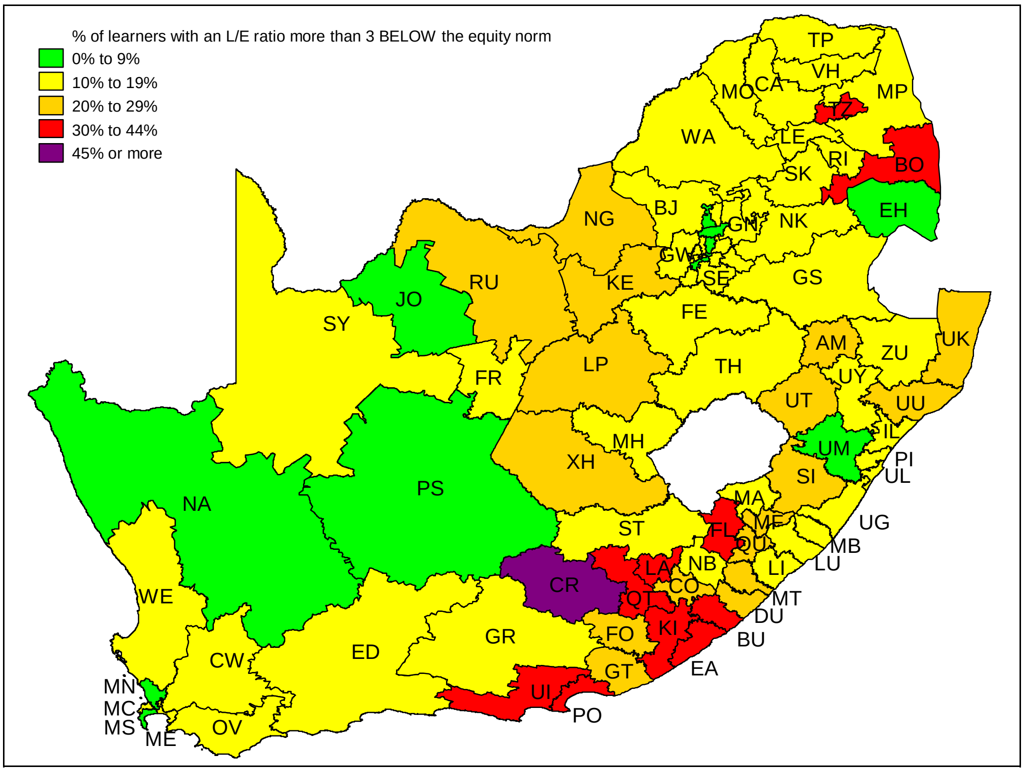
The first two maps below illustrate the compliance situation by district at the primary level. A threshold of 3 was used in deciding which schools were under- or over-staffed. Thus a school whose actual L/E ratio was 35 when the province-level simulation of the norms suggested it should be 31, was considered to be problematically under-staffed. In Figure 24 one sees there is a particularly high number of districts in Eastern Cape where 30% or more of primary-level learners are in problematically under-staffed schools. Figure 25 illustrates the reverse side of the problem, districts with many *over*-staffed schools, in other words schools which should be releasing educators to other parts of the province. Districts with a high proportion of under-staffed schools in Figure 24 tend to also be districts with a *low* proportion of *over*-staffed schools, but there are exceptions. For instance Port Elizabeth (PO) has over 30% of its learners in clearly *over*-staffed schools whilst over 20% of learners are in clearly *under*-staffed schools. This suggests that a part of the challenge is to reallocate educators *within* this district.

Figure 24: Percentage of primary learners with high L/E relative to provincial norms



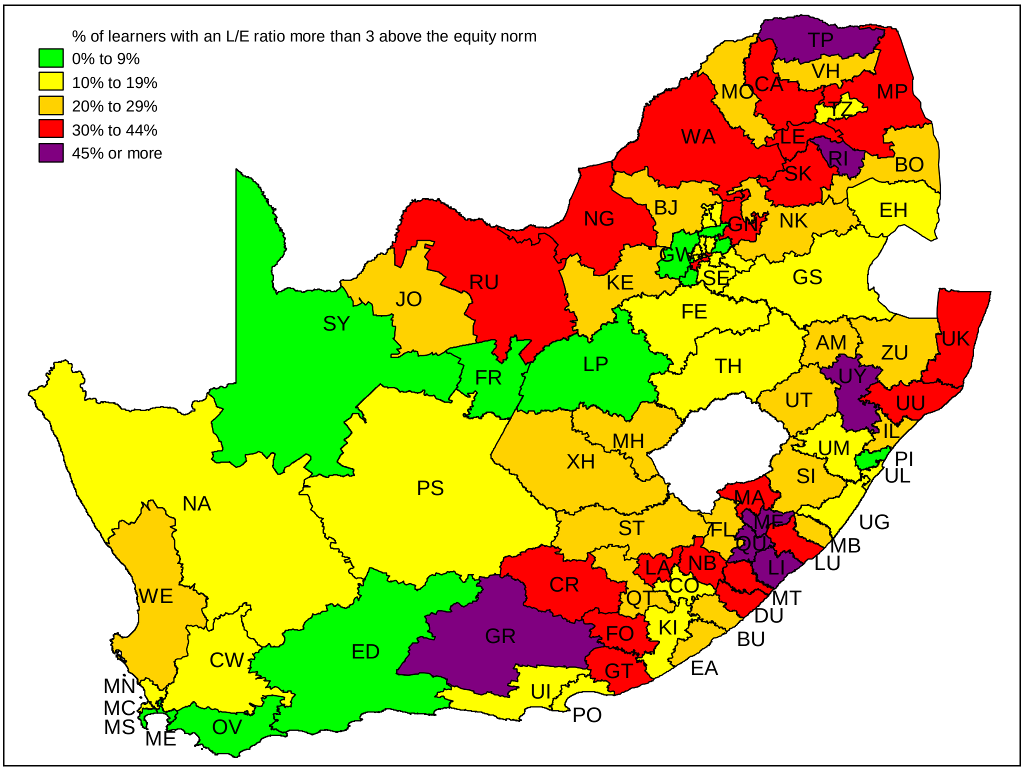
Note: Considered within this graph (and the next one) are all schools which have no learners in the range grades 10 to 12, or have some learners in the range grades 1 to 4. In other words primary and combined primary-secondary schools would be covered here. District codes are explained in Appendix C.

Figure 25: Percentage of primary learners with LOW L/E relative to provincial norms



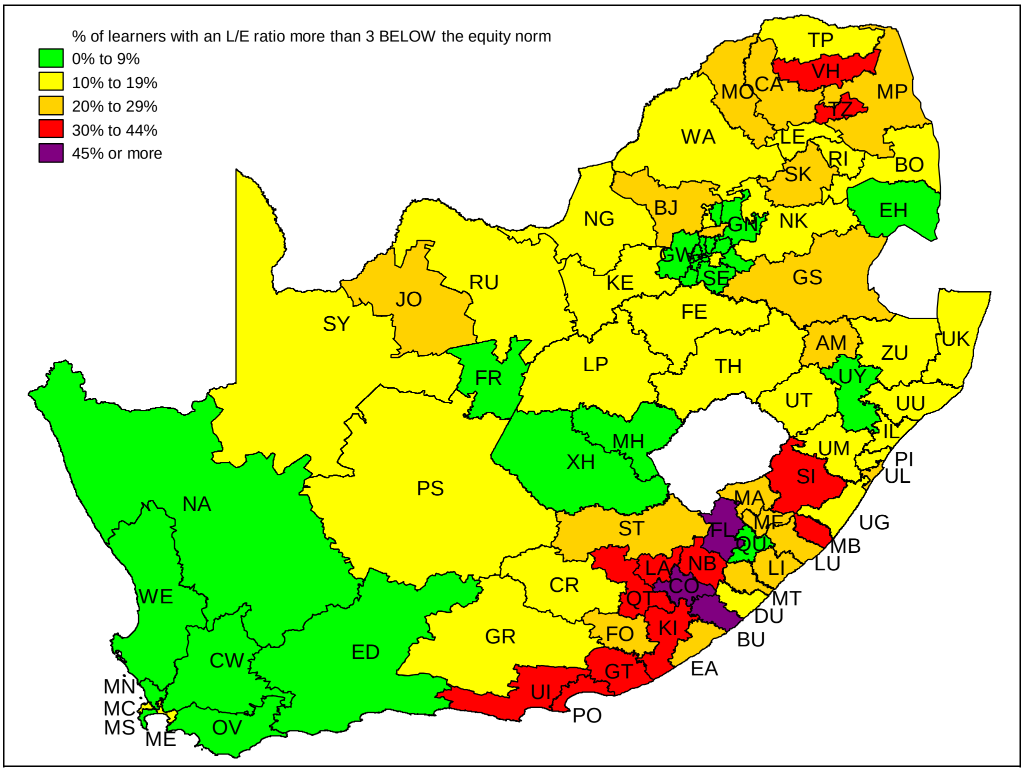
The next two maps present a similar analysis for the secondary level. It is clear that the under-staffing problem is more serious at the secondary level than at the primary level. Part of the reason for this is almost certainly that secondary-level teaching posts are more difficult to fill because of their specific requirements and the fact that it is more difficult to hire under-qualified teachers at this level as a 'gap-fill' measure. But a further explanation could be that moving secondary-level teachers is more difficult than at the primary level. This last explanation seems particularly important for Eastern Cape considering the number of districts in this province with a relative *over*-supply of teachers at the secondary level.

Figure 26: Percentage of secondary learners with high L/E relative to provincial norms



Note: Considered within this graph (and the next one) are all schools which have learners in the range grades 10 to 12, and have no learners in the range grades 1 to 4.

Figure 27: Percentage of secondary learners with LOW L/E relative to provincial norms



The following table illustrates an important relationship, namely that between the L/E ratio advantages and disadvantages used for the previous tables and maps, and learner performance. Appendix B provides details on the analysis behind the Table 14 summary. One can conclude that only in four provinces was greater *under*-staffing (measured by the degree to which the L/E ratio disadvantage statistic exceeded zero) associated, to a statistically significant degree, with *lower* learner performance. What is very noteworthy is that it is the provinces which are *better* at implementing the post provisioning norms where the expected relationship with learner performance emerges (see Table 8). A likely explanation is that the provinces in question (Free State, Gauteng, Mpumalanga and Western Cape, according to Table 14) are provinces which are good at implementing education policies in general, not just the post provisioning norms, and that this results in particularly effective teachers, meaning that a shortage of teachers is likely to be clearly felt in the results of a school. To put it crudely, in a province which does *not* implement policies well, under-staffing may matter less because there is a wider problem of, for instance, teacher absenteeism, weak support to schools, poor teacher and school accountability, and so on. What Table 14 and the Appendix B analysis moreover point to is a stronger link between staffing problems and performance at the primary level, compared to the secondary level.

Table 14: Relationship between post provisioning compliance and performance

	Was the L/E ratio disadvantage associated with lower learner performance at the primary (P) and secondary (S) levels in 2013?
EC	
FS	P
GP	PS
KN	
LP	
MP	P
NC	
NW	
WC	S
SA	P

Note: The symbols in this table are derived from the details seen in Table 31.

It is often argued that schools are under-staffed because they are in remote areas where it is difficult to attract educators. This matter is explored in Table 15 below. Specifically, this table explores the relationship between the following two variables: the L/E ratio disadvantage (counting both positive, disadvantage, and negative, advantage, values) and the distance of a school to the closest school with at least one grade in common. Clearly, the relationship between the two is not strong. For instance, at the secondary level the second variable explains only 0.5% of the variation in the first variable (this is at the national level). At the primary level, the level of explanation, or prediction, is also low, at 1.1%. The correlation is mostly positive at the secondary level (worse under-staffing situation, greater remoteness), but mostly *negative* at the primary level (more remote schools tend to be *better* staffed). This could be because rural schools tap into a reserve of locally available under-qualified, or unqualified, teachers, something which secondary schools in remote areas cannot do. However, it should be emphasized that the pattern is a weak one. At the primary level, only in one province, North West, is there a statistically significant slope coefficient (-0.4 here means that on average for every kilometre of remoteness, the L/E ratio disadvantage value *diminishes*, so gets better, by 0.4 learners). At the secondary level, where the expected relationship is more common, the slope coefficients support the assumption that remote schools have difficulties attracting staff, and that incentives to teach in remote areas are justified. It is noteworthy that this problem is most clearly visible, in the form of positive and statistically significant slope coefficients, in four provinces (three of which are the 'large and poor provinces'): Eastern Cape, KwaZulu-Natal, Limpopo, North West.

Table 15: School remoteness and L/E ratio disadvantage

	Secondary level (with grades 10 to 12 enrolment, no enrolment grades 1 to 4)			Primary and combined (all other grade combinations)		
	Correlation	R^2 (dependent is L/E advantage)	Slope coefficients	Correlation	R^2 (dependent is L/E advantage)	Slope coefficients
EC	0.213	0.045	0.3	-0.144	0.021	
FS	0.025	0.001		-0.177	0.031	
GP	0.059	0.004		0.074	0.005	
KN	0.054	0.003	0.1	-0.094	0.009	
LP	0.064	0.004	0.2	0.045	0.002	
MP	0.072	0.005		-0.014	0.000	
NC	0.001	0.000		-0.080	0.006	
NW	0.175	0.031	0.2	-0.245	0.060	-0.4
WC	-0.006	0.000		0.001	0.000	
SA	0.074	0.005	0.1	-0.103	0.011	-0.2

Note: Slope coefficients appear when these are statistically significant at the 10% level.

6.3 Compliance with the provisioning norms over time

The province-level regression analyses performed for Table 26 in Appendix B were repeated across different years of the Snap Survey data to assess the degree to which the distribution of educators followed the enrolment patterns of the previous year. One would not expect the distribution of educators to follow the enrolment patterns of the *same* year as historical data must by necessity be used when post entitlements are calculated. The system is currently not able to establish precisely school-level enrolments in a *future* year. Altogether 189 regression analyses were run, 21 for each province. For instance, actual educators per school in a province in 2013 would be regressed on enrolment patterns in 2012, with these enrolment patterns converted to the post provisioning weights seen in Table 26 (so the 2005 policy parameters were used). The adjusted R^2 value would then be taken (this was the value used in Table 8 above to assess a province's degree of compliance with the norms). Then 2013 educators would be regressed on 2011 enrolment patterns. Then 2013 educators would be regressed on 2010 enrolment patterns. And so on. Educators from years before 2013, going back as far as 2008, were used as the dependent variable. Explanatory variables would be enrolment variables from the *same* or from a *previous* year. The result was 21 R^2 values per province. To illustrate the methodology, the 21 values for KwaZulu-Natal are shown in Table 32 in Appendix B. Those 21 values were then used to produce the row for this province in Table 16 below. In the case of KwaZulu-Natal, 2013 educator distributions across schools were best described using 2011 enrolment data, meaning there was a two-year lag. Similarly, 2012 educator distributions were best described using 2010 enrolment data, also resulting in a two-year lag. And so on. There is no value in the 2009 column in Table 16 for KwaZulu-Natal as the post-2009 patterns suggest that one would not be able to obtain a reliable lag statistic for 2009 (because 2007 enrolment data were not used).

Table 16: Lags in years in the implementation of post provisioning

	2009	2010	2011	2012	2013	Average		Slope	
EC				4	4	4.0	High lag	0.0	No change
FS		2	2	1	2	1.8	Medium lag	-0.1	Getting better
GP	0	1	1	1	1	0.8	Low lag	0.0	No change
KN		2	2	2	2	2.0	High lag	0.0	No change
LP			3	2	3	2.7	High lag	0.0	No change
MP		1	1	1	2	1.3	Medium lag	0.3	Getting worse
NC	1	1	1	0	1	0.8	Low lag	0.0	No change
NW		1	1	2	2	1.5	Medium lag	0.4	Getting worse
WC	1	1	1	1	1	1.0	Low lag	0.0	No change

Table 16 brings to the fore vital information regarding staffing practices in provinces. The fact that such clear and consistent patterns within each province emerge is a good sign insofar as this means that the Snap Survey data are fairly reliable. There are only a few anomalies. For instance, it is unlikely that Northern Cape was able to use 2012 enrolment data to provision teachers in 2012 (this is what zero means). Clearly some provinces are succeeding in maintaining the best possible situation, namely a one-year lag. This applies in Gauteng, Northern Cape and Western Cape. These three provinces are not only able to use the previous year's enrolment data to create accurate school post establishments for the current year, they are also able to move educators rather quickly in accordance with the new establishments. At the other extreme, Eastern Cape displays a lag of four years, so for instance the 2013 spread of educators is best explained by 2009 enrolment data. This confirms the point made in a previous section, namely that much of the 'post provisioning problem' is an implementation problem, as opposed to a policy design problem. The final columns in Table 16 point to a *deterioration* in the implementation of the norms for North West and Mpumalanga. In these provinces the ideal one-year lag used to exist, but it has shifted to a two-year lag.

The lags picture seen in Table 16 does not change substantially if the 8,300 or so schools with the Grade R data problem (Grade R exists, but no practitioners) are excluded from the analysis.

In the following table, the R^2 values obtained by comparing one year's actual educator counts with the enrolment-based weights from just one year back, are presented. What this analysis permits (unlike the previous table) is to gauge whether provinces with consistent lags, such as Western Cape, are getting better or worse at ensuring that educators follow learners. It is noteworthy that provinces such as Western Cape and Gauteng are indeed getting better at getting educators to follow learners, insofar as their R^2 values have shown a general upward trend. But it is also noteworthy that in three provinces, Eastern Cape, Northern Cape and North West, the correlation one would want to see between educators and enrolments has in fact been declining, or getting worse.

Table 17: Lags in years in the implementation of post provisioning

	2009	2010	2011	2012	2013	Slope	
EC	0.921	0.904	0.892	0.902	0.901	-0.004	Getting worse
FS	0.950	0.974	0.966	0.975	0.974	0.005	Getting better
GP	0.934	0.962	0.942	0.961	0.957	0.004	Getting better
KN	0.907	0.944	0.850	0.964	0.952	0.011	Getting better
LP	0.924	0.922	0.903	0.919	0.936	0.002	Getting better
MP	0.536	0.962	0.938	0.976	0.960	0.086	Getting better
NC	0.980	0.987	0.982	0.964	0.951	-0.008	Getting worse
NW	0.956	0.953	0.947	0.948	0.936	-0.004	Getting worse
WC	0.973	0.982	0.983	0.986	0.980	0.002	Getting better

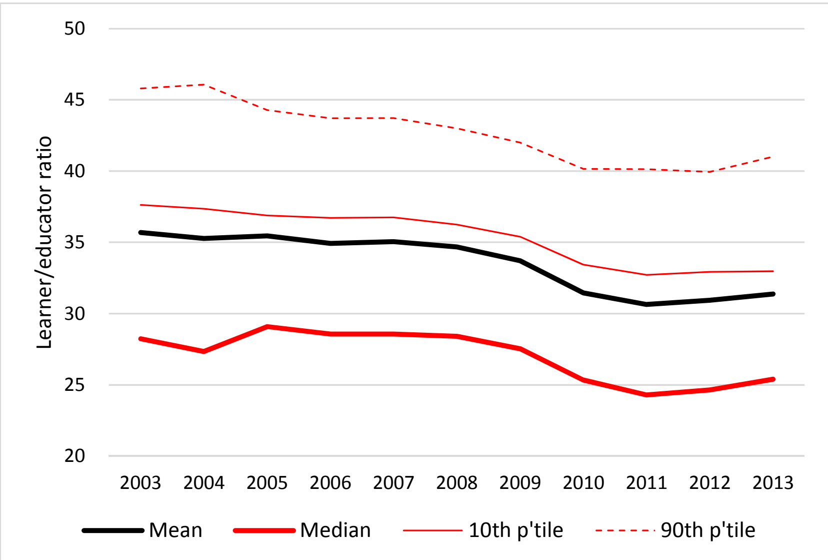
Note: Years in the column headings refer to the year of the educator count. There is no substantial change to the patterns seen in this table if the schools with Grade R-related data problems are excluded.

6.4 Learner/educator ratios over time

Above, learner/educator ratios in 2013 have been examined. Here the longer range trend, from 2003 to 2013, is examined. The starting point is 2003 as Snap Survey data (the only data used for the analysis presented here) for years before 2003 is clearly problematic, in particular as far as the numbers of educators are concerned, and the proportion of educators who are publicly paid.

For the following graph, which focusses on the primary level, only publicly employed educators were considered. Both the mean and the median point to a decline in the ratio up to 2011, followed by a slight rise. The mean is simply the overall number of learners divided by the number of educators, with the numerator and denominator drawing from the same schools. The median is weighted by learners so that small schools do not bias the statistic. One would therefore expect the median to be higher than the mean. The median can be thought of as the learner/educator ratio experienced by the median learner. By far, most of the changes up to 2011 in the ratios seen in Figure 28 are the result changes in the denominator, so the number of educators, not in the numerator, so the number of learners. Annual growth in the number of learners between 2003 and 2010 was 0.1%, against an annual growth of 1.6% with respect to educators. However, the post-2011 rise in the learner/educator ratio has been brought about largely by the enrolment increases already discussed in section 4. Inequality as measured by the gap between the 10th and 90th percentiles has not changed greatly over time, though it has declined slightly. In 2003 the gap was 17.6 learners, against 15.6 in 2013.

Figure 28: Learners/educator ratios at the primary level 2003-2013



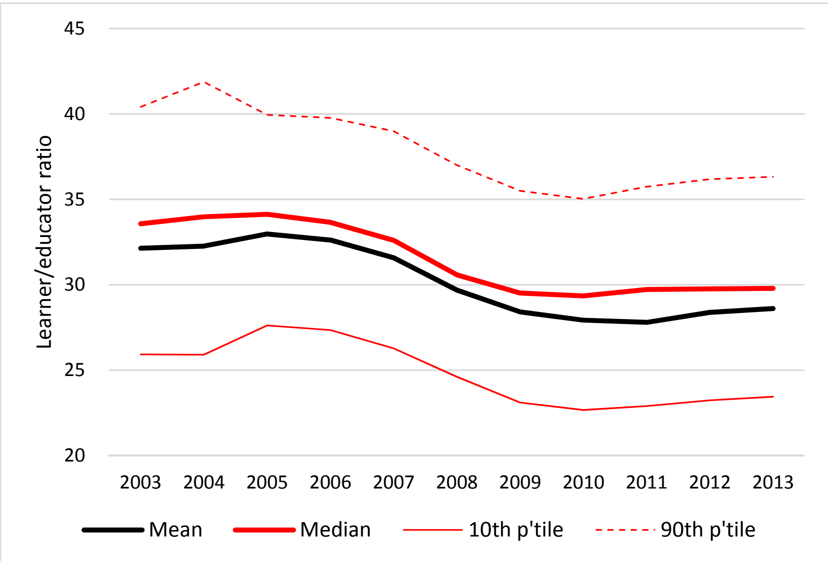
Source: Department of Basic Education, 2014.
Note: Only publicly paid educators are counted here and in the next graph. The graph covers public ordinary schools which were not covered in the next graph. In the schools covered in the above graph, 60% of learners are in schools offering grades 1 to 7, 16% in schools offering grades 1 to 9 and 3% in schools offering grades 1 to 12 (these are average percentages across years, and the presence of Grade R is ignored in the calculation of these percentages, though Grade R learners are counted in arriving at learner/educator ratios). The median, 10th percentile and 90th percentile values are weighted by each school's enrolment.

If privately paid educators working in public schools were included in the calculation, the picture seen in Figure 28 would change substantially, though the change would not look that different across the different years. In the years 2003 to 2005, counting privately paid educators reduced the mean learner/educator ratio by 2.1 learners. By 2011 to 2013, this reduction had become 2.4 learners.

For Figure 28 Grade R learners were counted. Learner counts in this regard are straightforward, but as discussed in a previous section, there are certain problems around the counting of Grade R teachers. A few tests were run to see whether these problems were affecting the picture unduly. It was found that they were not. Adjustments to correct the number of 'practitioners', meaning Grade R teachers (see section 6.1), were applied before Figure 28 was produced. Had this adjustment not occurred, the mean learner/educator ratios for the affected years, 2010 to 2013, would increase by 0.5 on average. However, the pre-2011 decline and post-2011 increase in the mean L/E ratio would remain essentially unchanged. For the years 2003 to 2007, when there was still a substantial number of schools without Grade R (over 50% in each year), an alternative set of curves were derived which excluded all schools offering Grade R. This exclusion resulted in a relatively large drop in the height of the mean L/E ratio curve (relative to what is seen in Figure 28). However, the downward slope remained the same. The drop in the L/E ratio could reflect problems around the counting of Grade R teachers, specifically the possibility that Grade R teachers were not counted in some schools in the Snap Survey. However, a more plausible explanation is that schools without Grade R tended to be historically more advantaged schools which found it easier to attract teachers into publicly funded posts. To conclude, it seems justified to consider Figure 28 a sufficiently accurate picture of the national trends in the L/E ratio at the primary level.

The following graph deals with the secondary level. As for the primary level, there has been a general reduction in the learner/educator ratio. If privately paid educators were counted the mean L/E ratio would be lower by 1.9 learners in 2003-2005, and 1.7 learners lower in 2011-2013. As for the primary level, there has been a small reduction in inequality. The gap between the 10^{th} and 90^{th} percentiles in Figure 29 declined from 14.5 learners in 2003 to 12.9 learners in 2013.

Figure 29: Learners/educator ratios at the secondary level 2003-2013



Source: Department of Basic Education, 2014.
Note: The graph covers public ordinary schools offering some schooling in the range Grade 10 to Grade 12, and no schooling below Grade 5. In the schools covered in the above graph, around 85% of learners are in grades 8 to 12 secondary schools, and 10% in grades 10 to 12 schools. See also the notes for the previous graph.

The following three graphs provide provincial versions of the national values discussed above. Certain erratic trends suggest data problems, for instance the exceptionally low L/E ratio for Northern Cape in 2004 in Figure 30. However, the trends seem consistent enough to allow for province-specific conclusions to be drawn. The general decline in the L/E ratio for the primary level seen at the national level is repeated across most provinces. However, a couple of provinces do not follow this trend. L/E ratios have clearly been rising in North West, whether one considers the mean (Figure 30) or the median (Figure 31). This province has moved from having one of the lowest provincial L/E ratios in 2003 to having one of the highest in 2013. Similarly, Free State did not see a decline in its L/E ratio, in fact the median increased. Yet in 2013 Free State remained a province with an exceptionally favourable (low) L/E ratio.

Figure 30: Mean learners/educator ratios at the primary level by province

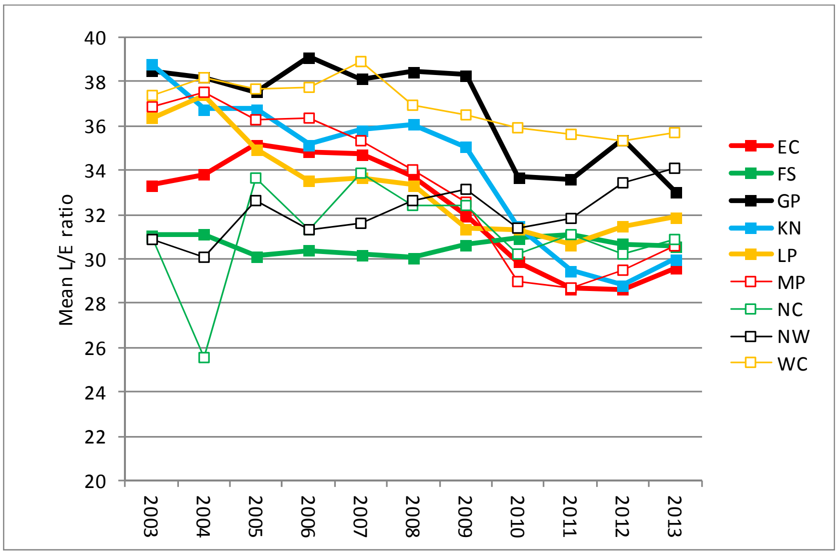
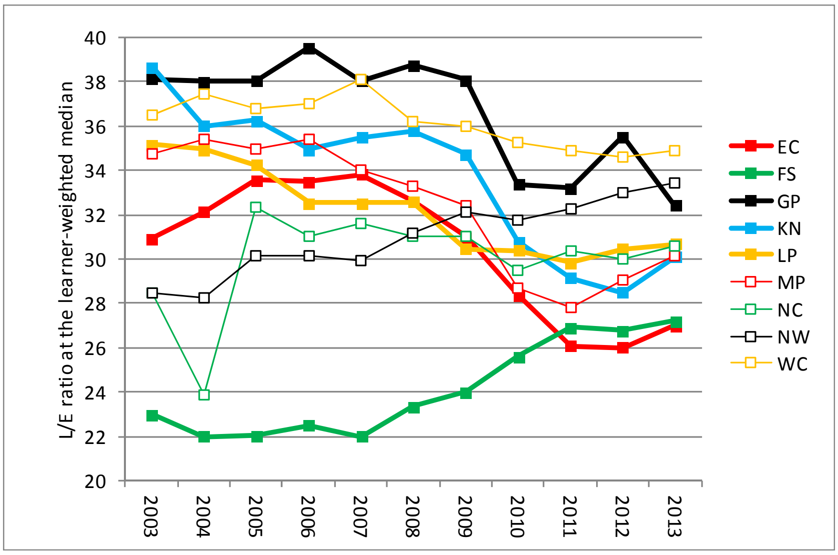


Figure 31: Median learner/educator ratios at the primary level by province



The finding referred to above that nationally reductions in the primary level L/E ratio were driven largely by the presence of more educators, holds true for a few provinces, for instance KwaZulu-Natal, but not for all (see Table 18 below). The North West increase in the L/E ratio was brought about by the fact that the number of educators declined faster than the number of learners, for instance.

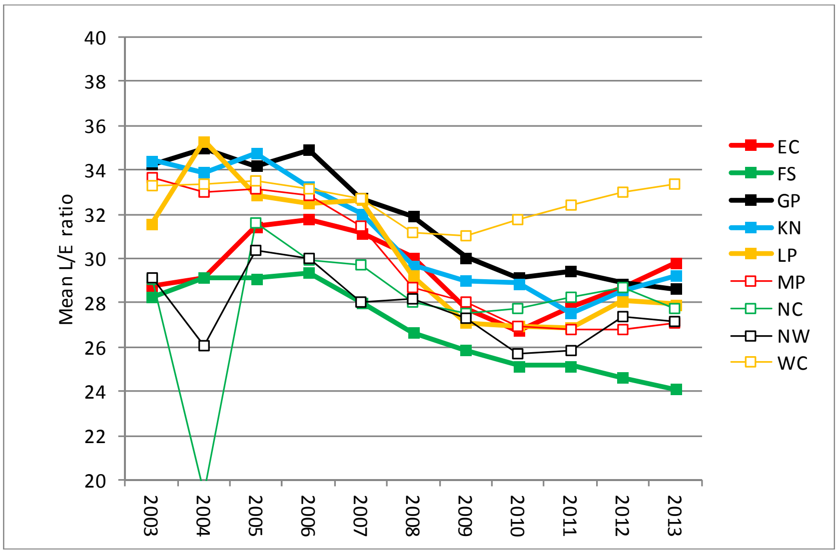
Table 18: Factors behind 2003-2013 L/E ratio changes

	Primary level			Secondary level		
	L/E ratio change	Enrolment change (%)	Educator change (%)	L/E ratio change	Enrolment change (%)	Educator change (%)
EC	-0.7	-1.7	0.4	-0.2	-0.2	0.5
FS	0.0	-0.1	-0.1	-0.6	-0.8	1.3
GP	-0.6	2.2	3.8	-0.7	2.1	4.4
KN	-1.0	0.1	3.1	-0.7	1.0	3.3
LP	-0.6	-1.4	0.4	-0.7	-0.1	2.3
MP	-0.9	1.3	4.1	-0.8	1.1	3.8
NC	0.1	7.0	6.9	0.1	6.8	5.9
NW	0.3	-1.1	-2.0	-0.2	-2.0	-1.1
WC	-0.3	0.9	1.7	-0.1	0.2	0.5
SA	-0.6	0.1	1.8	-0.6	0.6	2.4

Note: Figures represent annual changes, in the form of the steepness of the slope. The L/E ratio change in the annual change in the mean L/E ratio. The enrolment and educator change percentages are the slope divided by the mean across all years.

For the secondary level, the provincial trends for the mean L/E ratio (shown in Figure 32 below) are roughly similar to the trends for the median (not shown). According to Table 18, all provinces except for Northern Cape experienced an overall decline in their secondary-level L/E ratios, though from Figure 32 it is clear that several provinces did experience L/E ratio increases from around 2010.

Figure 32: Mean learners/educator ratios at the secondary level by province

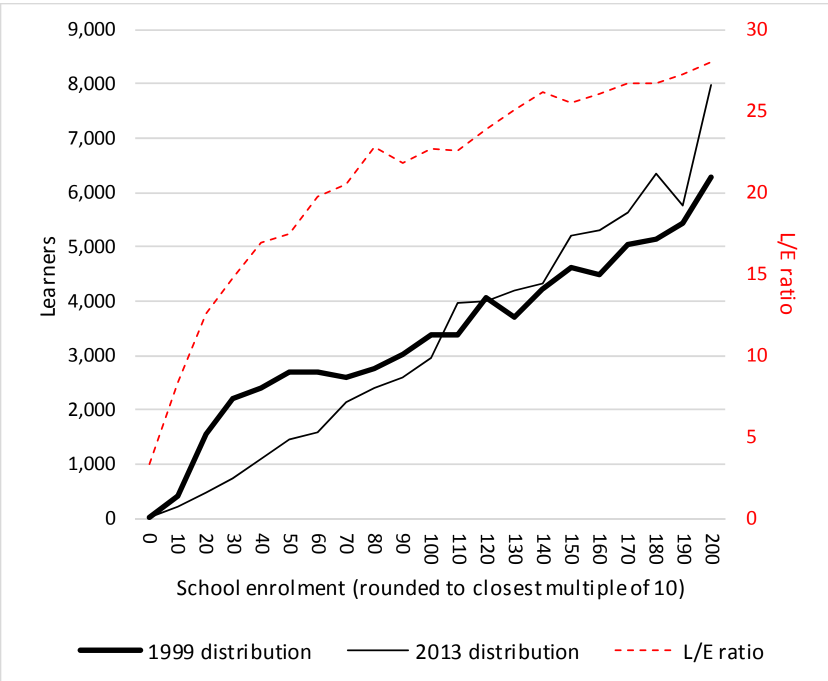


6.5 Urbanisation and economies of scale

A simple but rather telling analysis was undertaken to see whether a greater concentration of learners in larger schools was associated with a substantially reduced need for educators. In other words, a question of economies of scale was explored. One would expect some reduction in the need for educators in a more urbanised system with fewer learners in small and remote schools, because small schools inevitably require lower learner/educator ratios, even with multi-grade teaching (see section 5).

The following graph illustrates the logic of the approach. Snap Survey data from 1999 and 2013 were used. For each of the two years, schools were placed into school size categories, where each category was determined by the school's total enrolment rounded to the closest multiple of 10. Then total enrolment was made to be equal to 12 million learners in each year, through a proportional adjustment applied to all schools. The two enrolment distributions, for 1999 and 2013, are illustrated in Figure 33 below, up to school size 200. As one would expect, there are more learners in 1999 in smaller schools, compared to 2013. The 2013 actual learner/educator (L/E) ratio, counting just publicly paid educators, was then calculated for each category of school. As seen in the graph, L/E ratios are considerably lower for smaller schools, particularly for schools of a size below around 120 learners. The L/E ratio from 2013 was used to calculate a number of educators per school size category, for 2013 and then separately for 1999. This was done even for larger schools not illustrated in the graph. The total number of educators required in 2013 was indeed lower than for 1999, using the same total enrolment and the same L/E ratio per size category for both years. But the difference was small. In 1999, 375,326 educators were required, against 373,116 in 2013. The difference was thus only a small 0.6%. Assuming that all educators cost the same, this 0.6% difference represents the saving one can expect from fourteen years of urbanisation. This figure, combined with arguments made in section 5 around limited opportunities for closing schools, suggest that education planners should not view urbanisation as a major opportunity for greater economies of scale, at least as far as teacher provisioning is concerned. There could be specific regions where this could be achieved, but the evidence suggests that for the system as a whole large personnel cost savings are unlikely to be brought about in the near future as a result of a more urbanised schooling system.

Figure 33: Urbanisation and changing enrolment distributions



6.6 Indicators of within-province enrolment stability

Difficulties around pursuing the 'educators should follow learners' principle implicit in the post provisioning policy are clearly related to the degree to which learners move between schools. The less enrolment stability there is within a province, the more difficult it is to implement the post provisioning norms. Surprisingly, the degree of 'enrolment instability' within a province seems hardly to have been studied. The current section attempts to fill the knowledge gap.

The figures in the first column of Table 19 below were obtained by taking the Snap Survey enrolment data from the years 2012 and 2013, and converting enrolment values to weights, partly through the use of Annual Survey of Schools grades 10 to 12 subject figures (see Appendix B). The 2012 weights were then used to distribute 368,000 educators across all schools in the country, in a manner that resulted in whole educators allocated per school, with the minimum number being one educator per school. Thereafter 2013 weights were used to distribute a constant number of educators *per province* (the national total still being 368,000 educators). Changes in the number of educators per school between 2012 and 2013 were then analysed to conclude what percentage of educators should ideally be moved across schools between the two years. The figures in Table 19 indicate that Eastern Cape and North West would have experienced the greatest need to move educators between 2012 to 2013, as a result of increases or decreases in the relative enrolment weights of schools.

Table 19: Measures of enrolment stability by province

	% of educators who would move 2012-2013 (based on 368,000 educators simulation)	Indicator of between-school movements of learners					
		Within-province 2012-2013	Within-province 2010-2013	Within-province 2010-2013 just 'primary' (schools with Grade 1)	Within-province 2010-2013 just 'sec-ondary' (schools with Grade 12)	Within-province 2010-2013 Grade 12 only (using Snap)	Within-province 2010-2013 Grade 12 only (using exam-inations data)
EC	3.7	41.5	44.9	44.9	42.8	135.3	139.3
FS	2.8	30.1	28.3	28.3	33.5	119.5	117.2
GP	2.5	25.6	23.9	23.9	25.2	100.8	134.7
KN	2.9	31.3	29.2	29.2	30.6	123.6	127.6
LP	3.1	33.9	27.9	27.9	41.9	147.9	144.3
MP	2.8	30.1	24.4	24.4	33.4	121.8	121.9
NC	2.3	25.4	25.3	25.3	27.0	117.3	110.7
NW	4.0	44.0	31.1	31.1	43.5	133.1	124.7
WC	2.1	21.3	21.8	21.8	21.9	91.2	90.7
Avg.	2.9	31.5	28.5	28.5	33.3	121.2	123.5

Note: The figures in the bottom row are simple averages across the province rows.

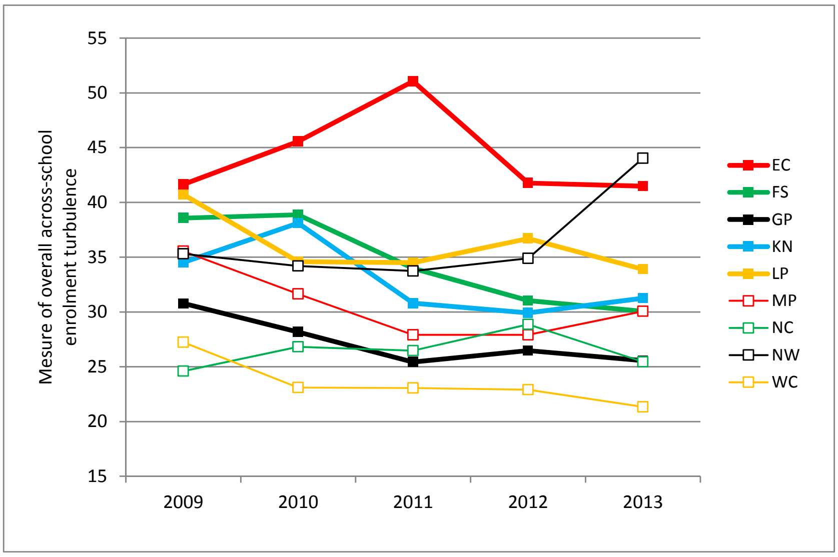
The remaining columns in Table 19 reflect values for an indicator of enrolment stability which was developed for this report. The aim was to develop an indicator which would gauge instability in the system based on the degree to which each school's share of overall enrolment changed between one year and the next. The formula for the indicator appears below. The indicator value *I* is calculated by adding *n* values, where *n* is the number of schools for which data exist across both periods of time *t*. Each of the *n* values is the absolute difference between the percentage of total enrolment in one year and in another. The final sum of *n* values is divided by two to provide a sense of the net movement across schools. One can think of the example of a two-school system where the enrolment shares are 50 and 50 in one year and 40 and 60 in the next. The absolute differences for the two schools would be 10 and 10. We would divide the sum 20 by 2 to arrive at a sense of shift in enrolment across the two years.

$$I = \left(\sum_{s=1}^{n} \left| \left(\frac{E_s}{\sum_{s=1}^{n} E_s} \right)_{t=2} - \left(\frac{E_s}{\sum_{s=1}^{n} E_s} \right)_{t=1} \right| \right) \Big/ 2$$

The second column of Table 19 provides values for *I* for each of provinces, using enrolments in 2012 and 2013. Note that the post provisioning weights were not used. Despite this, *I*, which is considerably easier to calculate than the post provisioning weights, provides an excellent approximation of the degree to which enrolment shifts influence the distribution of educator posts. Specifically, the correlation across the nine provinces of the values in the first column and the values in the second column of Table 19 is 0.996.

The remaining columns of Table 19 provide averages for *I* across three comparisons, 2010-2011, 2011-2012 and 2012-2013. As seen in Figure 34 below, values for *I* can change considerably over time. For instance, it appears that North West experienced far more enrolment instability between 2012 and 2013 than in previous years. Table 19 includes indicator values which draw just from certain schools or grades. Enrolments are slightly more stable at the primary than the secondary level, as seen in the lower values for *I* at the primary level. Though the stability of enrolment figures in Grade 12 do not affect post provisioning directly, as provisioning occurs for whole schools, calculating the indicator for just Grade 12 allowed for a comparison of Snap Survey data and Grade 12 examinations data. The two yield rather similar results. One would expect indicator values to be high for just Grade 12, given shifting strategies around promotion into Grade 12 from Grade 11.

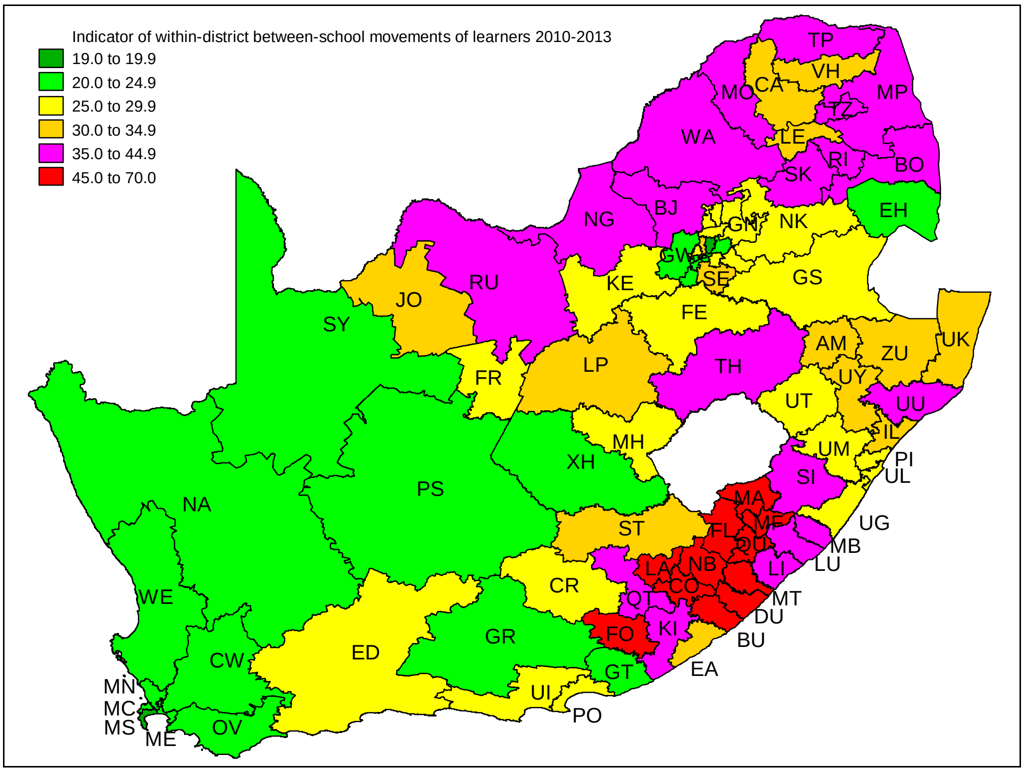
***Figure 34: Trends in the value of the enrolment stability indictor* 'I'**



Note: To illustrate, 2013 on the horizontal axis refers to the 2012 to 2013 comparison.

The next map illustrates the value of *I* per district, where only within-district shifts were taken into consideration. What seems surprising is how unstable enrolments are in more rural areas. Urban areas, such as Gauteng, reflect rather stable enrolment patterns. This is not what many would expect. One may think that enrolment instability would be highest in urban areas, where the proximity of schools is greatest and parents have greater leeway when it comes to switching from one school to another. Clearly, the map suggests this is not the case. In particular, the former Transkei area displays a high degree of enrolment stability, which would partly explain difficulties experienced in Eastern Cape with regard to the provisioning of teachers.

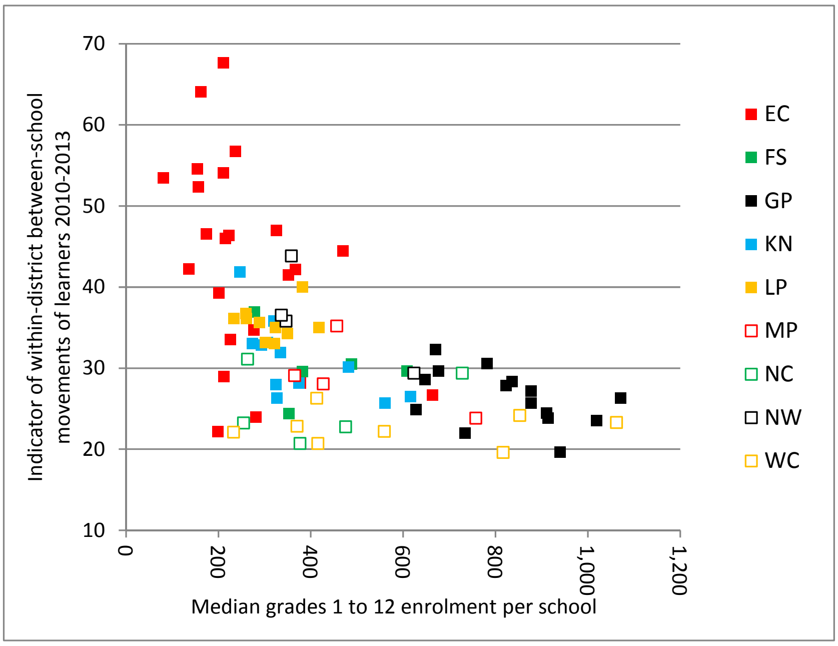
Figure 35: Enrolment stability within districts



Is there a correlation between the value of *I* and the prevalence of small schools? The next graph indicates that there is. However, districts in Eastern Cape still tend to have more enrolment stability than districts in other provinces with similarly sized schools. Eastern Cape's difficulties thus seem to be mix of structural and difficult-to-change factors, specifically the prevalence of small schools, and factors specific to Eastern Cape which encourage movement between schools. Some of the problem may be data problems, as opposed to actual learner movements across schools. In fact, some research points to a history of considerable 'gaming' of enrolment numbers by school principals in Eastern Cape[31].

[31] Gustafsson, 2015.

Figure 36: School size against enrolment instability by district



Note: Each point refers to one of the 86 districts.

7 A few policy implications of teacher movement patterns

This section focusses on what can be learnt from studying the movements of educators across schools, using payroll data. Very little analysis of this type has occurred in South Africa, partly due to issues around access to data. The analysis presented demonstrates that there is enormous potential for extracting important policy knowledge from the analysis.

One specific matter which is explored is the phenomenon of 'stepping up', in other words the tendency to move to a better performing schools. This phenomenon clearly exists in South Africa and what this means for policy is discussed.

By analysing which schools teachers choose to move to, when they move from one school to another, it is found that teachers tend, to some degree, to avoid moving to schools on the other side of a provincial boundary. The teacher labour market is thus rather bound by province. Given that teachers are trained according to a national curriculum it is concluded that the way provincial boundaries constitute barriers to movement reduces the efficiency of the overall educator allocation process. It ought to become easier for educators to find employment in other provinces.

The current section presents analysis which can be considered a starting point for further work in a potentially useful area, namely the movement of publicly paid employees across schools. Section 7.1 explains the degree to which employees in the payroll system can be linked to the geo-coordinates of schools. It also presents a few maps describing the distribution and movements of educators over the 2010 to 2012 period. Section 7.2 describes the patterns whereby educators tend to move to better performing schools when they change school. This is a pattern which has been observed in other schooling systems in the world, but has not been studied in South Africa previously[32]. More work needs to be done beyond what is presented here, partly so that a wider set of policy implications can be established. Section 7.3 does answer one important South African policy question, namely whether educators avoid moving into other provinces. It is found that this is indeed the case. This finding seems to confirm that one way the larger post provisioning system could be improved is by making it easier for educators to move across provincial boundaries, so that the matching of posts to educators can become more effective.

7.1 Data issues and the basic patterns

Linking schools-based staff in the payroll system (Persal) to school details, such as school geo-coordinates, contained in other datasets was partly resolved through a special link file which was obtained and which linked national EMIS school identifiers to Persal 'component numbers'. But the component number for the same institution can change over time, so an alternative and additional approach had to be employed. In this approach, the Persal numbers of employees were used to establish where components in different years, with different numbers, were in fact the same school. Specifically, if enough people in one component in one year were found to also be in a differently numbered component in another year, it was assumed that the component was the same place. Table 20 below provides an idea of how successful the linking of educators to school geo-coordinates was. In calculating these percentages, the denominator was all publicly paid schools-based educators according to the Snap Survey (counting public ordinary schools only). The numerator was those in the denominator whose school was found in Persal and whose school had geo-coordinates. The Persal data were from October 2011 and November 2012. The success rate of linking Persal to EMIS data was high, with 98% of educators being assigned geo-coordinates.

[32] The phenomenon for school principals in South Africa is to some extent dealt with in Wills (2015: 28).

Table 20: Percentage of educators with geo-coordinates

	2011	2012
EC	99	99
FS	99	99
GP	97	96
KN	98	98
LP	98	98
MP	97	97
NC	97	98
NW	95	95
WC	100	99
SA	98	98

Table 21 below presents a basic schema for educator movements between two consecutive years, and populates this schema using 2011 and 2012 Persal data. Under the category 'permanent' the group 'permanent on probation' has been included. The fact that there are so few joiners moving into ordinary schools (see 2,323) is indicative of the fact that educators joining schools tend first to be employed on a temporary basis. Educators not in ordinary schools could be in a variety places, including departmental offices, adult education centres, special schools, and teacher centres. The 'Other stayer combinations' row includes mostly educators who were clearly in a school in 2012, and probably in a different school in 2011, though the 2011 school EMIS number could not be confirmed. This row would also include educators moving from, say, an adult school to an ordinary school.

Table 21: Educator shifts 2011-2012

2011	2012	All educators	Permanent educators
Joiners		**21,997**	**2,673**
Outside the system	Ordinary school	16,016	2,323
Outside the system	Not an ordinary school	5,981	350
Stayers		**407,107**	**356,185**
Ordinary school	Same ordinary school	341,271	311,075
Ordinary school	Different ordinary school, same province	16,756	11,843
Ordinary school	Different ordinary school, different province	1,300	890
Ordinary school	Not an ordinary school	2,177	1,646
Not an ordinary school	Not an ordinary school	40,696	27,938
Other stayer combinations		4,907	2,793
Leavers		**25,306**	**13,335**
Ordinary school	Outside the system	20,246	11,959
Not an ordinary school	Outside the system	5,060	1,376
Total across both years		**454,410**	**372,193**

The figures from the above table point to a 5.1% joining rate (joiners over stayers plus leavers) and a 5.9% attrition rate (leavers over stayers plus leavers). The figures imply an annual net loss of 3,309 educators overall, and 4,230 for ordinary schools. Both the joining and attrition rates can be considered over-estimates if one's desired indicator is educators joining or leaving *for the first time*. In a separate analysis by the author of the current report the attrition rate has been found to be as low as 3.0% if one excludes educators who leave but then return within a period of four years.

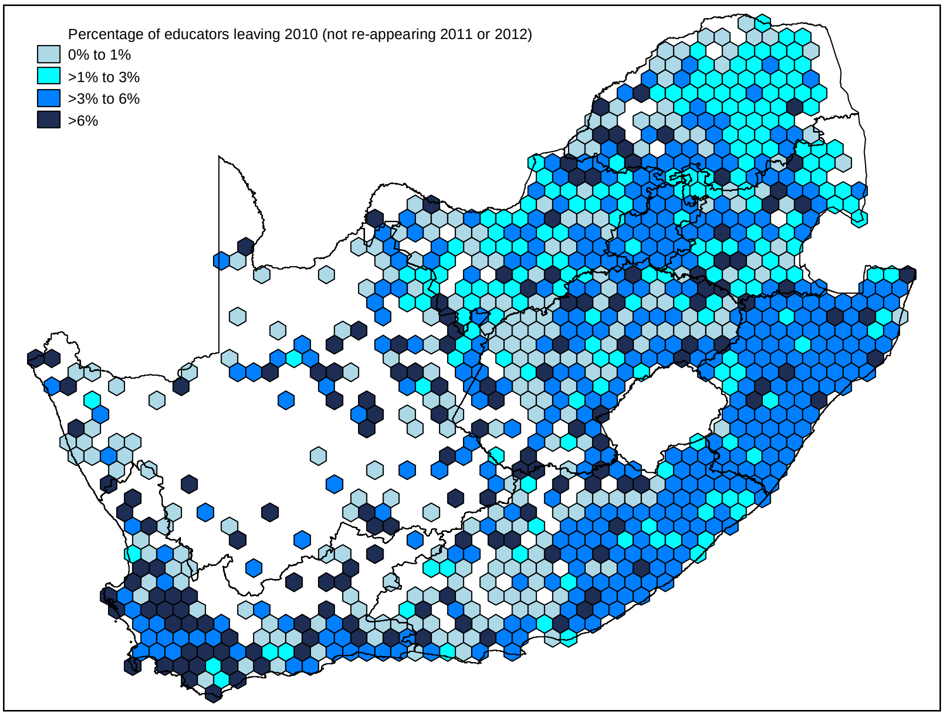
In the analysis that follows October 2010 Persal data, in addition to the 2011 and 2012 data referred to above, are used. The six descriptive statistics appearing in Table 22 below are discussed in relation to a series of maps which follow.

Table 22: Indicators of 2010 to 2012 movements

	Percentage of educators leaving 2010 (not re-appearing 2011 or 2012)	Percentage of educators moving school 2010-2012	Percentage gain or loss in all educators 2011-2012	Percentage gain or loss in permanent educators 2011-2012	Percentage of educators arriving 2012 (not appearing 2010 or 2011)	Average age in 2012 of schools-based educators
EC	3.5	-0.2	-5.8	-3.7	1.0	45.7
FS	3.6	0.0	1.9	0.1	5.6	44.9
GP	4.0	0.5	1.1	-0.6	5.8	44.8
KN	4.0	0.0	-1.7	0.5	4.1	42.4
LP	2.4	-0.1	4.2	4.8	1.5	46.8
MP	2.6	-0.2	0.3	4.1	3.1	45.3
NC	3.6	-0.4	1.2	-0.3	5.6	45.0
NW	3.0	-0.1	1.2	0.1	4.0	46.2
WC	5.1	0.1	-0.5	0.7	5.5	45.9
SA	3.6	0.0	-0.4	0.6	3.6	44.8

The first statistic is the percentage of schools-based educators leaving their school between 2010 and 2011, and not reappearing in 2012 in any school. The percentages could therefore reflect educators leaving a school and moving into a non-school location within the system. Figure 37 below illustrates that attrition from schools has been lowest in more remote parts of the country, such as the far north or Limpopo, rural Free State and the more sparsely populated and arid regions of the 'three Capes'. A likely explanation is that in these parts of the country there are fewer alternative employment opportunities for educators, in fact information and networks which might guide teachers to jobs elsewhere may be weak.

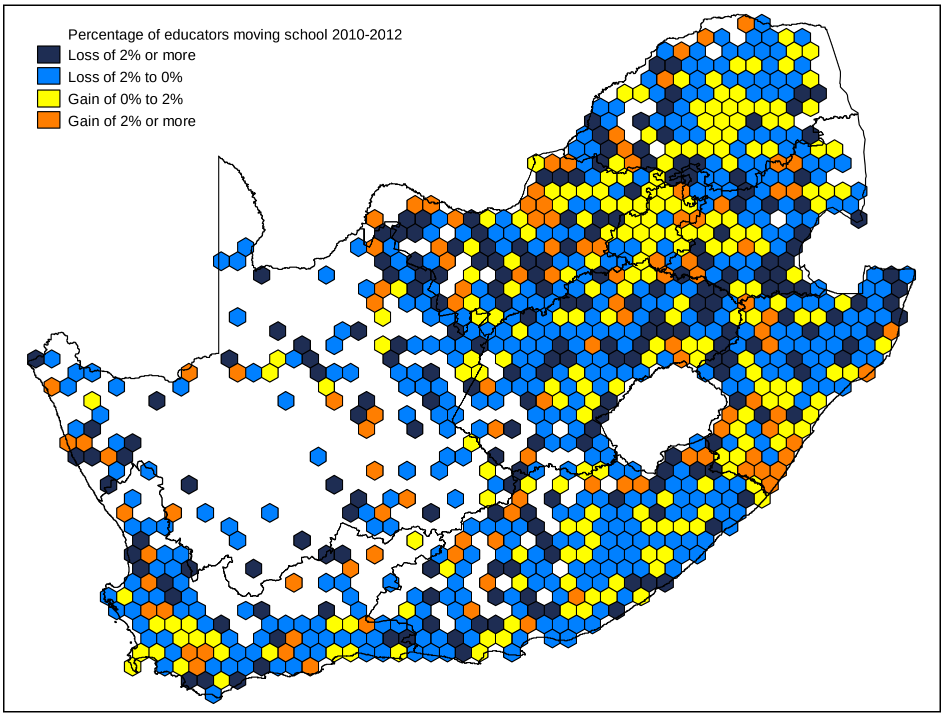
Figure 37: Publicly paid educators leaving schools



Source: Here and in the maps that follow the source is Persal data for October 2010, October 2011 and November 2012, linked to the geo-coordinates of schools.

The second statistic is the percentage of schools-based educators changing their school during the 2010 to 2012 period. Specifically, an educator loss percentage was calculated by expressing the number of schools-based educators who left a school for another school between 2010 and 2011 or between 2011 and 2012 as a percentage of all schools-based educators, and then dividing the percentage by two to obtain and annual figure. An educator gain percentage was calculated along similar lines. The final statistics reflected in Figure 38 below are the net loss or gain per cell. The national statistic would always be zero (see Table 22). The general patterns seems to be one of movement towards more urban areas. Even within some provinces, there appears to be a movement towards provincial capitals, for instance in Limpopo and North West. In KwaZulu-Natal there appears to be a clear movement from the north to the south of the province.

Figure 38: Between-school movement of educators 2010-2012



The previous map focussed only on gains and losses involving movements across schools. The third statistic, reflected in the next map, is the overall gain or loss of schools-based educators, taking into account educators arriving in schools and leaving schools altogether. Here the national statistic is a net loss of -0.4% between 2011 and 2012. How should the previous map be read together with the next one? To illustrate, in KwaZulu-Natal there was a southward movement of educators, but at the same time in most parts of the province the number of educators was declining. The implication is that the southward movement occurred partly because people applied for educator posts left vacant by departing educators. The dramatic 2011 to 2012 decline in educator numbers in the Transkei area of Eastern Cape is striking. However, this would to a large extent be a reflection of the province's management of temporary educators. A more stable picture emerges if one focusses only on permanent educators, as in Figure 40. Here a loss would occur if a permanently employed educator left a school. A gain could be either the arrival of a permanently employed educator into a school, or the promotion of an educator within a school from a temporary to permanent status between 2011 and 2012 (the latter pattern in fact accounts for 62% of gains). Note that the following two maps could have been produced using Snap Survey data, whilst the other maps in this section all require Persal data as the necessary individual-level data are not available in the Snap Survey.

Figure 39: Gain or loss of all educators 2011-2012

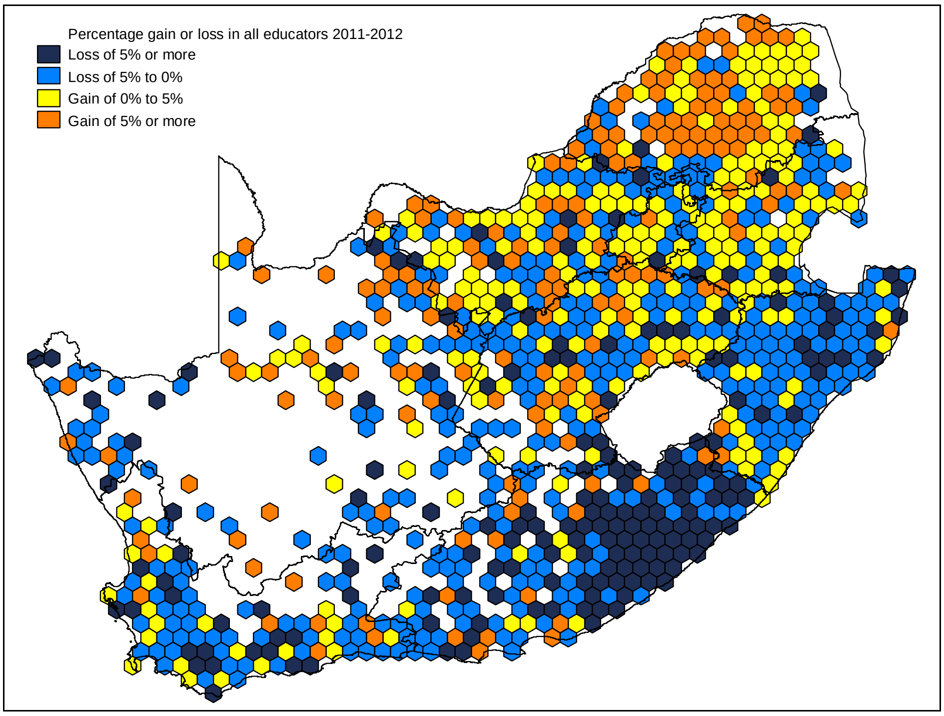
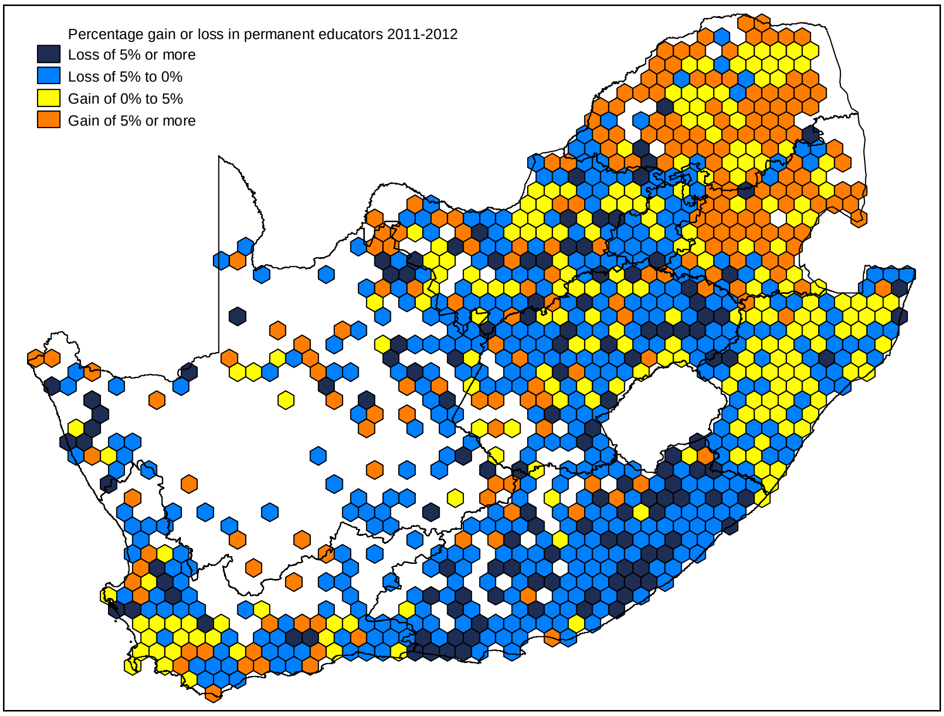
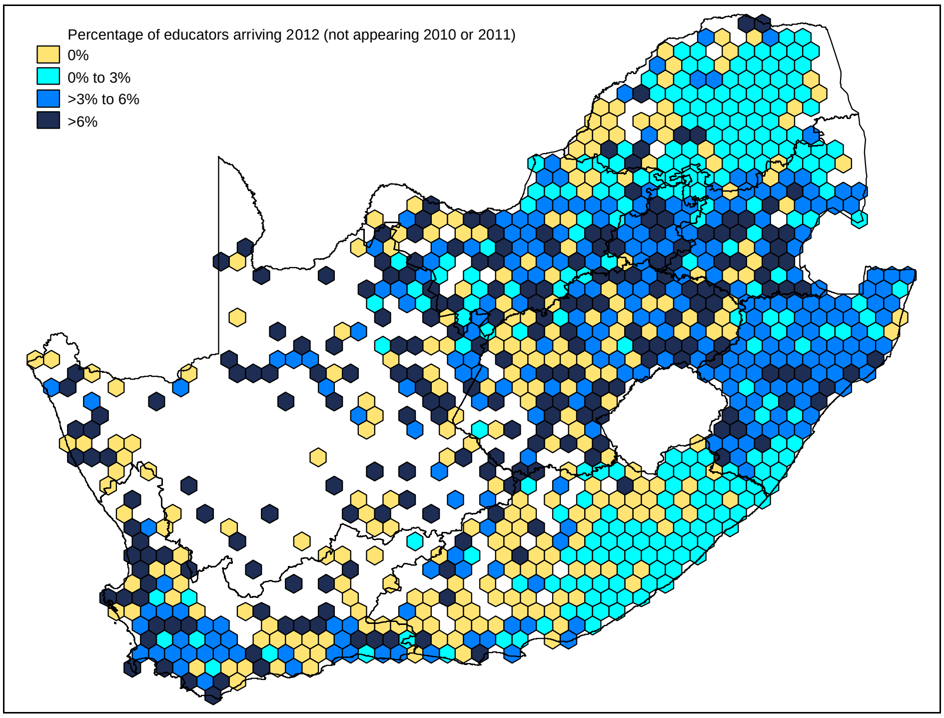


Figure 40: Gain or loss of permanent educators 2011-2012



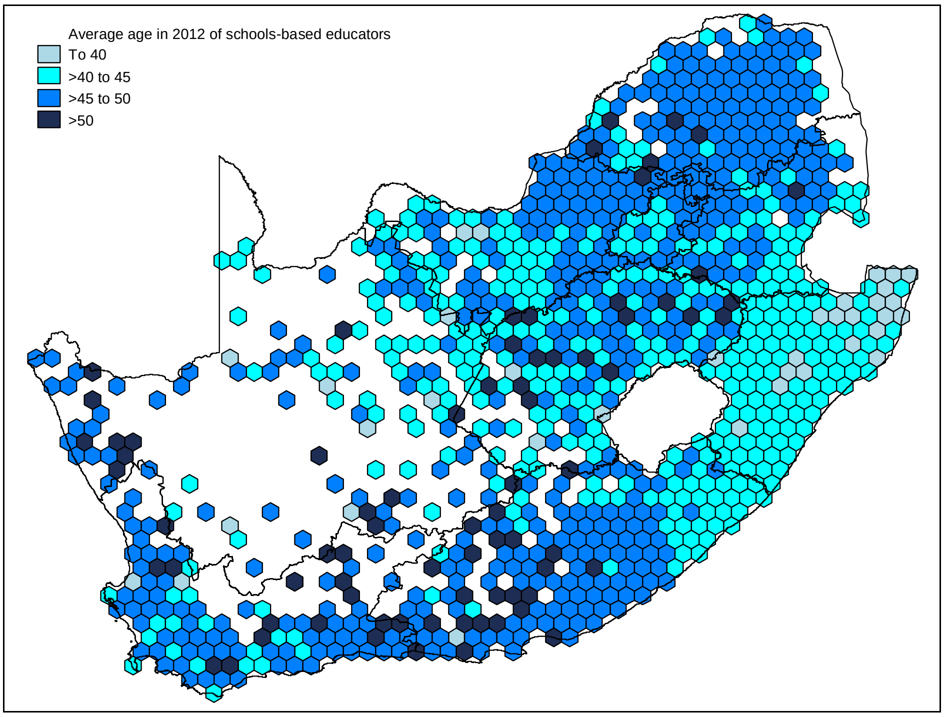
The fifth statistic is illustrated in Figure 41 below. It is clear that levels of joining are particularly low in Limpopo and Eastern Cape. In Limpopo this is understandable given the low levels of leaving amongst schools-based educators (see Figure 37). In Eastern Cape educator departures would be at least partly justified by ongoing declines in enrolment numbers (see Figure 3 in section 4).

Figure 41: Publicly paid educators joining schools



Finally, Figure 42 below deals with educator age. What is very clear is the concentration of younger teachers in KwaZulu-Natal, and the absence of any cell in this province where the average age exceeds 45 (in 2012). The same can be said of the far east of Eastern Cape.

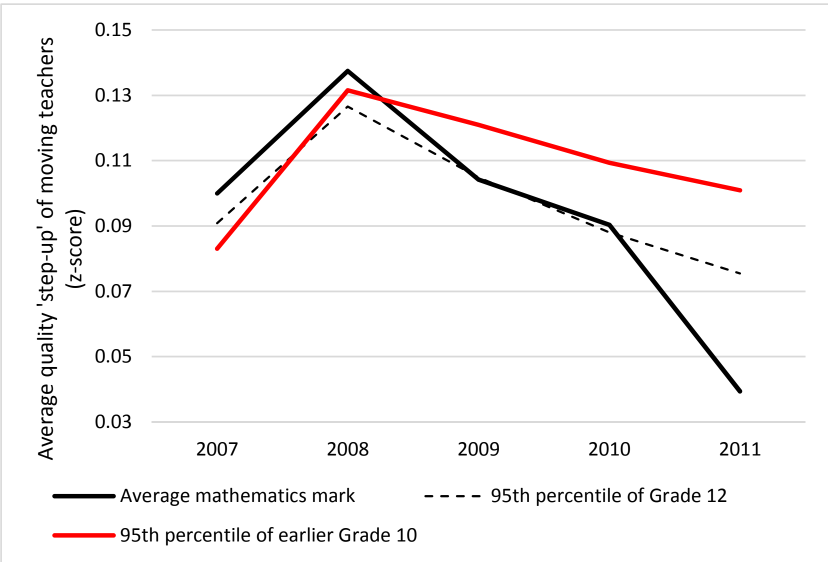
Figure 42: Average age of schools-based educators



7.2 Movements towards better quality schools

This section presents some basic yet important analysis establishing that educators, when they move, tend to move to better schools. Moreover, the size of the quality 'step-up' experienced by moving teachers is quantified. The following graph draws from the data of 3,929 educators who moved school between 2011 and 2012. Only educators moving from a schools with Grade 12 to another school with Grade 12 were considered. Moreover, both the previous and the new school had to have Grade 12 mathematics results for all the years from 2007 to 2011. The horizontal axis refers to the year whose examination results were used. If one focusses on the average mathematics mark, it is clear that on average moving educators moved to a better performing school. Probably only a minority of moving educators were mathematics educators, yet using mathematics results seems appropriate as this subject's results tend to be indicative of the performance of a school as a whole. Although one is looking at movements between 2011 and 2012, it seems that results from 2008 were most likely to inform the school choices of educators (this is assuming that teachers were in general searching for better schools). One might expect a bit of a lag. The information educators have about other schools may not be the latest, and applying for a post and obtaining it can be a lengthy process. The fact that 2008 values in Figure 43 are much higher than 2007 values is likely to be the outcome of the fact that in 2007 examinations according to an outgoing curriculum were written, a curriculum which would become less familiar to educators over time. Using the 2008 average score point, we see that on average educators moved to schools whose average mark was 0.14 standard deviations above the previous school. A difference of 0.14 is not very large if one considers that the difference between schools at the 10th and 90th performance percentiles was in 2008 2.6 standard deviations. Yet for 10% of moving educators the 'step-up' was greater than 1.34 of a standard deviation. Moreover, 44% of educators 'stepped down' in the sense that they moved to worse performing schools.

Figure 43: Size of the quality 'step-up' for 2011-2012 moving teachers



The other two curves in Figure 43 refer to indicators of school performance which would be less easily observed by educators. The indicator '95th percentile of earlier Grade 10' means the 95th percentile of mathematics performance in Grade 12 using an approach where earlier Grade 10 enrolment is counted, so learners dropping out before Grade 12 is taken into account. This indicator has been found to be a particularly stable indicator for ranking the performance of schools[33]. The fact that the 2008 value for this indicator is lower than the value for the average mark can be seen as indicative of the fact that the average mark provides a partially inaccurate picture of which schools are better performing. The fact that by 2011 the '95th percentile of earlier Grade 10' indicator value is substantially higher than the average mark value is indicative of the fact that on the whole educators make fairly accurate moves, in the sense of moves which maximise the 'step-up'. Inaccuracies within the school measures which educators easily see, such as the average mark, mean that to some degree educators make the wrong school choices, but these inaccuracies are not serious enough to offset the general pattern of an actual and real process of 'stepping up'.

Table 23 breaks across-school movements down by the historical apartheid-era administration type. The 'step-up' values are standard deviations using 2008 average mathematics marks. Much of the 'stepping up' occurs between schools which are both former homeland schools. The average 'step-up' here, across 1,381 educators, is 0.16 standard deviations. Movements into historically Indian (HoD) and white schools are in general associated with a large 'step-up'.

Table 23: 2011-2012 movements across ex-department

New➔ Old🡇	Homeland	DET	HoR	HoD	White	New	Total
Homeland	1,381 (0.16)	224	18 (1.00)	35 (1.44)	19 (1.89)	185 (0.26)	1,862
DET	167	591 (0.06)	26 (0.37)	26 (1.18)	54 (0.93)	54	918
HoR	12 (-0.35)	18	94	5 (1.29)	29 (0.74)	9 (-0.62)	167
HoD	14	17 (-1.32)	4	49	20 (0.39)	3	107
White	10	33 (-1.11)	21 (-1.02)	9	255 (0.12)		328
New	170	68	7	4 (0.92)	10 (1.90)	95	354
Total	1,754	951	170	128	387	346	3,736

Note: Here and in the following table numbers represent educators of any rank, permanently employed or not, moving school between 2011 and 2012. In brackets are 'step up' values, or the average difference in the average Grade 12 mathematics score between the new and old school (new minus old, using z-scores). Only 'step up' values which are statistically significantly different from zero, using the 5% level of significance, are shown. Homeland and 'DET' are historically black African, 'HoR' is historically coloured, 'HoD' is historically Indian.

The next table illustrates movements across quintiles. Given that learners with a higher socio-economic status and greater home background advantages perform better, it is not surprising that on average the move is towards a higher quintile. Of all movers, 36% move to a richer quintile, whilst 24% move to a poorer quintile. The remainder stay within the same quintile. If one takes, for each mover, the quintile of the new school and subtracts from this the quintile of the old school, the average one obtains across all movers is 0.22 of a quintile.

[33] See report accompanying the current report and titled *Treating schools to a new administration: Evidence from South Africa of the impact of better practices in the system-level administration of schools.*

Table 24: 2011-2012 movements across quintile

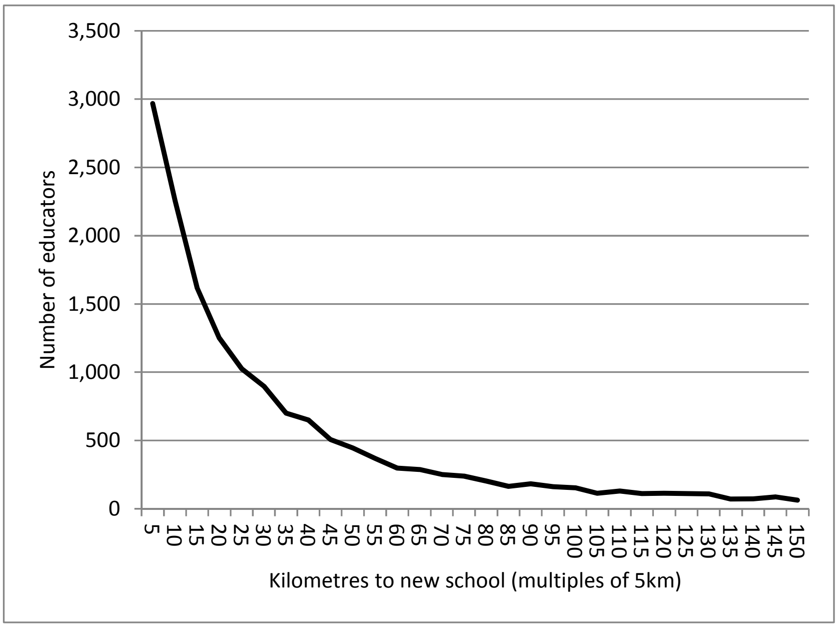
New➔ Old🡻	Q1	Q2	Q3	Q4	Q5	Total
Q1	360 (0.19)	222 (0.15)	224	89 (0.32)	30 (0.77)	925
Q2	190	261 (0.14)	249 (0.21)	101 (0.31)	35 (0.74)	836
Q3	146	159	464 (0.18)	196 (0.35)	90 (0.75)	1,055
Q4	40	59 (-0.33)	157	158 (0.13)	136 (0.53)	550
Q5	8	13 (-1.31)	52 (-0.64)	82 (-0.73)	282 (0.15)	437
Total	744	714	1,146	626	573	3,803

What does the above analysis of the 'step-up' imply for policy and further research? It would be good to understand what educators are really looking for when they attempt to 'step up'. Is it better performing learners, who generally make teaching easier, or is it the better resourcing associated with better performing schools? Are the teacher who 'step up' exceptionally good teachers? To what proportion of teachers is 'stepping up' not something they are interested in, even if an easy opportunity for this exists (considering that so many educators move without 'stepping up')? Turning to policy, the 'stepping up' phenomenon strengthens arguments for special interventions to keep teachers, in particular better performing teachers, in poorer schools and schools which struggle academically. Such interventions would include incentives to teach in schools which teachers would otherwise tend to avoid. Such incentives need not be monetary. They could take the form of special further education opportunities, for instance. The phenomenon also strengthens the argument for targeting more disadvantaged schools when it comes to the in-service training of teachers, and teacher accountability programmes, in particular if these schools tend to be left with teachers who need most support. Just inserting information on the 'stepping up' phenomenon into the public discourse on schooling could stimulate debate around innovative ways in which teachers could be encouraged to remain in less advantaged schools, or even seek to teach in such schools.

7.3 Provincial boundaries as barriers to teacher movement

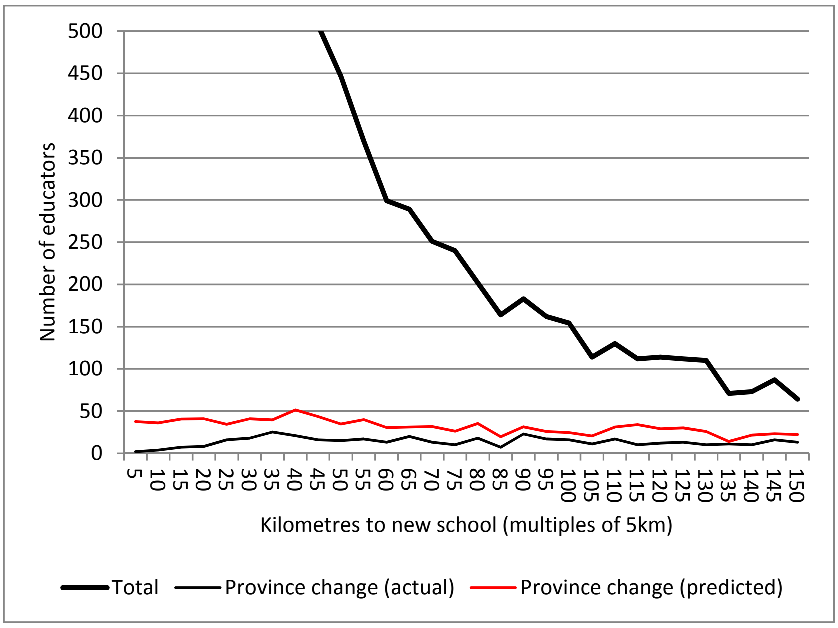
Figure 44 below uses data on 17,289 educators moving school between 2011 and 2012 (a few of these educators would have data lying beyond the right-hand limit of the horizontal axis). The mean distance between an educator's old and new schools was 61 km, and the median 23km. Just 15% of educators moved more than 100 km. In general, then, educators do not move very far. Distances are of course straight line distances as only the geo-coordinates of schools were used in the analysis, not any other spatial data dealing with elements such as roads.

Figure 44: Distance moved by educators 2011-2012



Data on actual moving can be used to gain a rough idea of the degree to which educators avoid moving to other provinces, possibly because this is administratively difficult. In the case of each moving educator the following was done. The distance actually moved was considered. For instance, an educator may have moved 100 kilometres. Thereafter all schools roughly that distance from the school of departure were considered. A threshold of 15 kilometres was used, meaning that schools between 85 and 115 kilometres from the school of departure were considered. If all these schools were within the same province as the school of departure, then moving to another school within the same province was taken as inevitable. However, if say half the schools were in another province, then the probability of moving to a school in another province was considered 50%. Probabilities were added, and compared to actual figures, to produce the following graph, which zooms into the previous graph and features two additional curves. As one might expect, if probabilities (or predictions) are used, a slightly higher level of across-province movement is found. Overall, only 7% of 17,289 educators who moved to a different school between 2011 and 2012 moved to a different province. In the predicted scenario, 12% moved to a different province. Roughly, around 850 more schools-based educators would move each year to another province if the province of one's new school was not an issue. What this analysis indicates is that there are indeed barriers which discourage movement across provinces.

Figure 45: Educator movements 2011-2012 and provincial boundaries



References

African National Congress (2013). *53rd National Conference Resolutions*. Johannesburg. Available from: <http://www.anc.org.za/docs/res/2013/resolutions53r.pdf> [Accessed June 2015].

Brophy, T.S.L., Daniels, R.C. & Musundwa, S. (2014). *gpsbound: Routine for importing and verifying geographical information from a user provided shapefile*. Cape Town: SALDRU.

Centre for Development and Enterprise (2014). *Cities of hope: Young people and opportunity in South Africa's cities*. Johannesburg.

Deloitte (2013). *National implementation of post provisioning.* Johannesburg. [Unpublished report produced for the DBE.]

Department of Basic Education (2014). Snap Survey of ordinary schools 1997-2013 (dataset). Pretoria. Available from: <http://www.datafirst.uct.ac.za> [Accessed June 2014].

Department of Basic Education (2015). *Annual performance plan 2015-2016*. Pretoria. Available from: <http://www.education.gov.za> [Accessed September 2015].

Department of Education (2005). *Report of the Ministerial Committee on Rural Education: A new vision for rural schooling*. Pretoria.

Department of Transport (2009). *Final draft: National scholar transport policy*. Pretoria.

Gastner, M.T. & Newman, M.E.J. (2006). Optimal design of spatial distribution networks. *Physical Review E*, 74(016117).

Gustafsson, M. (2015). Enrolment ratios and related puzzles in developing countries: Approaches for interrogating the data drawing from the case of South Africa. *International Journal of Educational Development,* 42: 63-72.

Gustafsson, M. & Taylor, S. (2013). *Treating schools to a new administration: The impact of South Africa's 2005 provincial boundary changes on school performance*. Stellenbosch: University of Stellenbosch.

Habermann, H. & Padrutt, P. (2011). Growth diagnotics: Strengths and weaknesses of a creative analytical framework to identify economic growth constraints in developing countries. *Journal of Knowledge Management, Economics and Information Technology*, 7: 1-25.

Hausmann, R., Klinger, B. & Wagner, R. (2008). *Doing growth diagnostics in practice: A 'mindbook'*. Cambridge: Center for International Development.

Heeks, R. (2003). *Most eGovernment-for-development projects fail: How can risks be reduced?* Manchester: University of Manchester.

Jerven, M. (2013). *Poor numbers: How we are misled by African development statistics and what to do about it.* Claremont: UCT Press.

Kok, P. & Collinson, M. (2006). *Migration and urbanisation in South Africa*. Pretoria: Stats SA.

Laldaparsad, S. (2013). *Urban and rural trends in South Africa*. Pretoria: Stats SA.

National Planning Commission (2012). *National development plan 2030: Our future - make it work*. Pretoria.

Spaull, N. (2015). *Education quality in South Africa and Sub-Saharan Africa: An economic approach*. Stellenbosch: University of Stellenbosch. Available from: <http://resep.sun.ac.za/wp-content/uploads/2015/04/Nic-Spaull-Thesis.pdf> [Accessed May 2015].

Statistics South Africa (2013). *Census 2011 Municipal factsheet*. Pretoria.

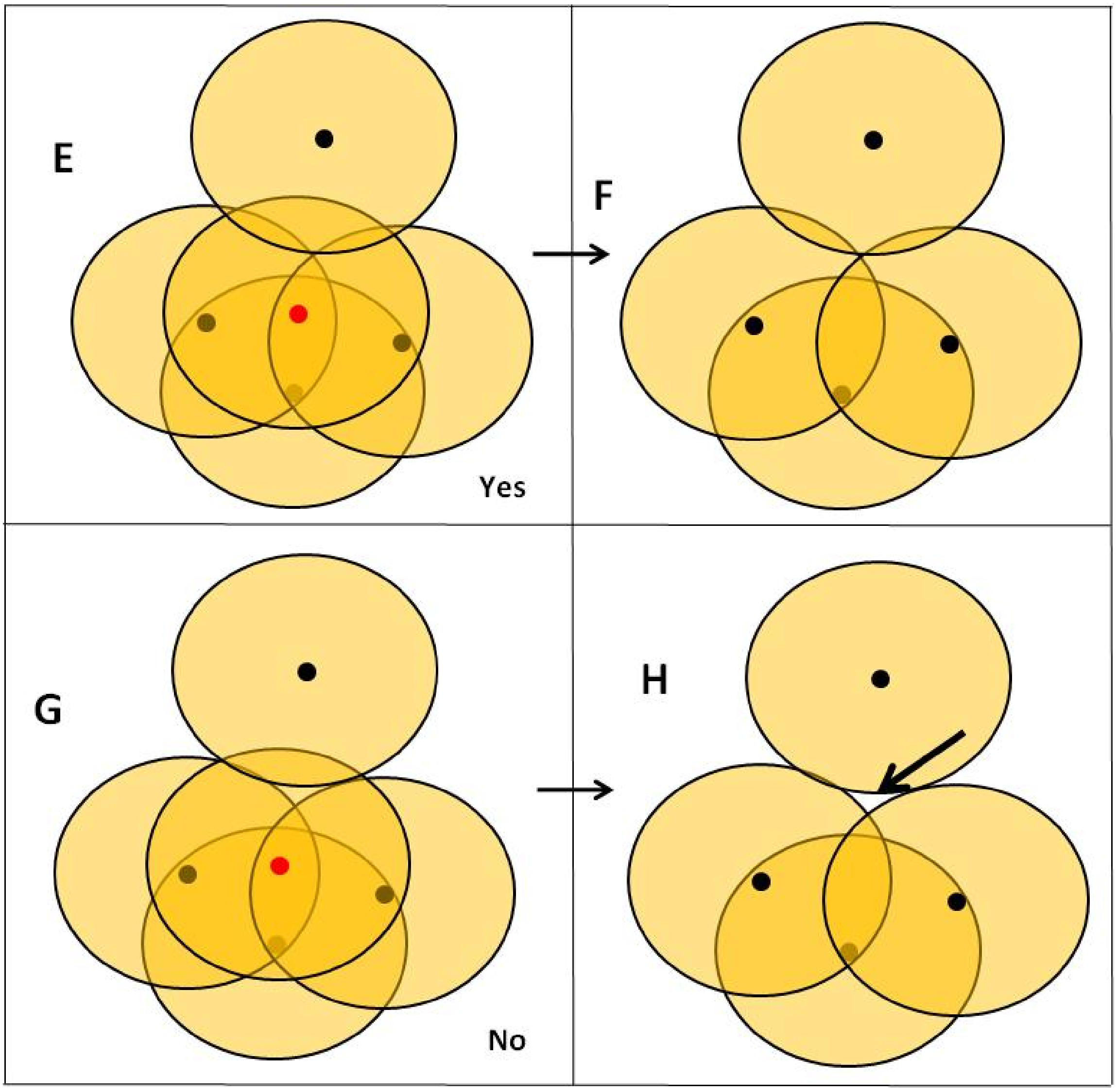
United Nations: Population Division (2014). *World urbanization prospects: The 2014 revision*. New York.

Wills, G. (2015). *A profile of the labour market for school principals in South Africa: Evidence to inform policy*. Stellenbosch: University of Stellenbosch. Available from: <https://ideas.repec.org/p/sza/wpaper/wpapers245.html> [Accessed February 2016].

Appendix A: Details on the analysis of small schools

This appendix provides details on the analysis into the closure of small schools introduced in section 5. One kind of geographical scenario is not covered in the approach presented in the code. This is the one in Panel E below, which should result in the removal of the red dot school, given that its removal does not result in the loss of any covered area (compare panels E and F). However, this kind of scenario seems impossible to deal with unless one creates a raster grid across the area, something that would not be impossible to program, but would be very time-consuming. Part of the complexity is that one would need to distinguish between a situation such as that of Panel E, which can lead to a school closure, and a somewhat different situation such as that of Panel G, where the school should not be closed given that doing this would leave a blank, uncovered area (see area indicated by large arrow in Panel H). Thus what the code below does is that it limits itself to situations where nearby schools are not further than the critical distance (2 km) from the reference school.

Figure 46: Further school location scenarios



The Stata code begins on the next page.

```
****************************************************************************************************************
set more off
* Setting of the enrolment threshold:
local minpupils = 83 // This is the minimum number of pupils needed for the school to be non-small.
* Setting of the distance threshold:
local maxdist = 2.0 // This is maximum distance, in km, that one would want pupils to travel.
use "temp15", clear
* Variables are the following:
** natemis - school identifier (numeric)
** learners - number of pupils
** gis_long - longitude in decimal format
** gis_lat - latitude in decimal format
* No values should be blank.
gen learners2 = learners // New variable will contain the post-merging enrolments.
gen merged = 0 // A value of 1 will show the school was merged.
save "tempschools", replace
quietly keep if learners<`minpupils'
sort learners
save "tempsmallschools", replace
local nsmallschools = _N
* Looping through schools below the enrolment threshold:
forvalues ismallschools = 1 / `nsmallschools' {
  use "tempsmallschools", clear
  local mylon = gis_long[`ismallschools']
  local mylat = gis_lat[`ismallschools']
  local mysmallschool = natemis[`ismallschools']
  local mysmalllearners = learners[`ismallschools']
  use "tempschools", clear
  quietly drop if natemis==`mysmallschool' | merged==1
* Identifying nearby schools:
  gen templat = `mylat'
  gen templon = `mylon'
  quietly vincenty templat templon gis_lat gis_long, vin(tempdist) inkm // Here the distance between the reference small school and all other schools is calculated.
  quietly keep if tempdist<=`maxdist'
  if _N>1 { // One needs at least two nearby schools to get close to a situation where all of the circumference of the possibly removed school is covered.
* Determining orientation of nearby schools relative to base school:
    quietly vincenty templat templon gis_lat templon, vin(tempadjacent) inkm
            * From this point forward calculations are made as if the Earth was flat, which seems acceptable given one is looking at very small distances.
    gen tempangle = acos(tempadjacent / tempdist) * (180 / _pi) // Variable tempangle is the bearing of the other school to the nearby school, with north being 0 and counting eastwards (clockwise).
```

```
          * The 180 * _pi bit above converts radian values to degrees.
    quietly replace tempangle = 180 - tempangle if gis_lat<templat & gis_long>templon
          quietly replace tempangle = 180 + tempangle if gis_lat<templat & gis_long<templon
    quietly replace tempangle = 360 - tempangle if gis_lat>templat & gis_long<templon
* Establishing which part of the circumference is now covered by this nearby school:
    gen maxdist = `maxdist'
    gen tempangle2 = acos((tempdist / 2) / maxdist) * (180 / _pi)
    * Variable tempangle2 above is angle between line to nearby school and line to the point at which the two circles defined by maxdist overlap.
    gen minpoint = tempangle - tempangle2
          quietly replace minpoint = 360 + minpoint if minpoint<0
          gen maxpoint = mod(tempangle + tempangle2, 360)
* Checking whether nearby schools cover the entire circumference of the smaall school:
    sort minpoint
          gen maxpoint360 = maxpoint
          quietly replace maxpoint360 = 360 + maxpoint if maxpoint<minpoint
          local gapfound = 0
          local nnearby = _N
          local maxoverall = 0
    forvalues inearby = 1 / `nnearby' {
           if `inearby'!=_N {
             if maxpoint360[`inearby']>`maxoverall' {
                    local maxoverall = maxpoint360[`inearby']
                  }
             if `maxoverall'<minpoint[`inearby' + 1] {
                    local gapfound = 1
                    continue, break
                  }
           }
           else {
             if `maxoverall'<360 & maxpoint[`inearby']<360 {
                    local gapfound = 1
                  }
           }
           if `maxoverall'>=360 {
             continue, break
           }
          }
* Merging the small school to nearest school, if possible:
    if `gapfound'==0 {
           sort tempdist
```

```
            local mynatemis = natemis[1]
            use "tempschools", clear
            quietly replace merged = 1 if natemis==`mysmallschool'
            quietly replace learners2 = 0 if natemis==`mysmallschool'
            quietly replace learners2 = learners2 + `mysmalllearners' if natemis==`mynatemis'
            quietly save "tempschools", replace
            local remainder = `nsmallschools' - `ismallschools'
            display "School `mysmallschool' merged. `remainder' more small schools to analyse."
           }
  }
}
* Producing a summary of results:
use "tempschools", clear
quietly summ natemis, det
local numberschools = r(N)
quietly summ learners, det
local totlearners = r(sum)
quietly summ merged, det
local mergedschools = r(sum)
quietly keep if merged==1
quietly summ learners, det
local totlearnersclosed = r(sum)
use "tempschools", clear
quietly keep if learners2>0 & learners2!=learners
quietly summ natemis, det
local numberreceivers = r(N)
use "tempsmallschools", clear
quietly summ natemis, det
local numbersmallschools = r(N)
display "Number of schools analysed: " `numberschools'
display "Total enrolment in these schools: " `totlearners'
display "Distance threshold used (km): " `maxdist'
display "Enrolment threshold used: " `minlearners'
display "Schools with enrolment below the threshold: " `numbersmallschools'
display "Number of schools closed: "  `mergedschools'
display "Number of learners in closed schools: " `totlearnersclosed'
display "Number of schools receiving learners from closed schools: " `numberreceivers'
*****************************************************************************************************************
```

Appendix B: Analysis of learner/educator ratios

The following three tables reflect information of relevance to section 6.1, which dealt with the distribution of teachers in 2013. Table 25 below provides descriptive statistics relating to the school-level data used for the subsequent regression analysis. Of the twelve school-level weights used to calculate the entitlement to educators posts, all but two (the basic school and poverty weights) are applicable to only some schools, depending on the grades offered by the school. Clearly the five first weights are the largest. Three weights use data from the 2013 Annual Survey of Schools, namely the grades 10 to 12 subject-learner weight, the weight reflecting a second language of instruction at the primary level and the weight doing the same for the secondary level. For all weights data from the 2013 Snap Survey were used. A few schools, specifically 65 of them, were excluded from the regression analyses because their quintile was missing from the data and it was thus not possible to calculate the poverty weight. A further six schools were excluded because their staff data were missing in the Snap Survey dataset. One-educator schools are indicated because certain provinces have, in addition to the specifications of the post provisioning policy, introduced the rule that two should be the minimum number of educators in a school (this matter is discussed below).

Table 25: Post provisioning weights

	EC	FS	GP	KN	LP	MP	NC	NW	WC	SA
% of learner-weighted schools with non-zero weight values										
Gr 1 to 4 learner	74	65	61	59	55	62	68	64	66	63
Gr 5 to 6 learner	73	64	61	57	54	61	67	63	65	62
Gr 7 learner	70	59	60	56	53	57	62	59	64	60
Gr 8 to 9 learner	61	47	39	47	46	46	49	37	43	47
Gr 10 to 12 subjects	27	36	38	42	45	39	34	33	36	37
School	100	100	100	100	100	100	100	100	100	100
Combined school	3	7	1	4	1	7	7	5	2	3
Grade within 1 to 7	75	71	63	62	57	68	73	71	66	66
Grade within 8 to 12	72	52	41	48	46	49	52	43	43	50
Second lang. prim.	6	31	19	10	9	53	19	6	21	16
Second lang. seco.	2	13	5	3	2	30	11	3	13	7
Poverty	100	100	100	100	100	100	100	100	100	100
Mean school-level weight values for non-zero schools (no learner weighting)										
Gr 1 to 4 learner	176	271	581	269	254	341	293	319	391	301
Gr 5 to 6 learner	67	108	219	105	98	134	116	119	150	112
Gr 7 learner	41	78	117	60	54	84	73	68	89	60
Gr 8 to 9 learner	119	311	556	266	263	329	245	305	425	184
Gr 10 to 12 subjects	511	553	889	505	377	584	472	501	708	270
School	15	15	15	15	15	15	15	15	15	15
Combined school	15	15	15	15	15	15	15	15	15	1
Grade within 1 to 7	13	12	13	13	13	12	13	12	13	9
Grade within 8 to 12	5	7	9	9	10	8	7	7	8	4
Second lang. prim.	13	11	16	12	11	18	20	12	22	3
Second lang. seco.	25	17	26	15	15	15	23	19	26	1
Poverty	21	31	58	30	27	37	32	31	43	51
Schools with quintile	5,539	1,325	2,035	5,926	3,922	1,762	551	1,546	1,452	24,058
Schools including missing quintile	5,551	1,327	2,056	5,936	3,924	1,768	553	1,550	1,458	24,123
Educators actual (with quintile)	59,141	22,459	53,580	87,719	54,240	32,216	8,245	23,647	27,195	368,442
% one-educator schools	2.4	18.5	0.0	0.2	0.5	1.5	0.2	0.3	0.1	1.8
% one-educator schools*	0.3	0.5	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1

*Note: Row headings referring to weights should be read in combination with the earlier Table 5. * refers to learner-weighted schools.*

Table 26: Regression outputs for post provisioning simulation (2005 policy)

	EC	FS	GP	KN	LP	MP	NC	NW	WC	SA
Slope coefficients										
Gr 1 to 4 learner★	0.019***	0.024***	0.021***	0.025***	0.026***	0.024***	0.021***	0.022***	0.021***	0.022***
Gr 5 to 6 learner★	0.026***	0.028***	0.021***	0.021***	0.021***	0.027***	0.025***	0.030***	0.020***	0.025***
Gr 7 learner★	0.023***	0.028***	0.031***	0.036***	0.039***	0.022***	0.023***	0.020***	0.033***	0.025***
Gr 8 to 9 learner★	0.021***	0.020***	0.019***	0.024***	0.020***	0.024***	0.020***	0.023***	0.019***	0.023***
Gr 10 to 12 subjects	0.018***	0.027***	0.027***	0.023***	0.027***	0.024***	0.021***	0.024***	0.021***	0.023***
Combined school	-0.121***	-0.108***	-0.141***	-0.039**	-0.045	-0.017	-0.046	-0.010	-0.066**	-0.034***
Grade within 1 to 7	0.020*	0.005	0.032	0.080***	0.003	0.028	-0.044	-0.096***	0.067**	0.049***
Grade within 8 to 12	0.358***	0.460***	0.264***	0.089***	0.126***	0.182***	0.089	0.019	0.149***	0.153***
Second lang. prim.	0.026**	0.002	0.014	-0.002	0.002	0.004	0.028	-0.001	0.009	0.009***
Second lang. seco.	0.021	-0.059***	0.019	0.012	0.049**	0.004	0.035	-0.025	0.017**	0.012**
Poverty	0.001	0.023***	0.002	-0.004	-0.020***	0.003	0.032***	-0.009	0.011***	0.000
Intercept	1.7***	1.2***	0.6***	0.9***	1.9***	1.4***	1.5***	3.1***	0.2***	1.4***
Adjusted R^2	0.888	0.968	0.949	0.935	0.913	0.943	0.929	0.928	0.975	0.930
Predicted educators	59,141	22,459	53,580	87,719	54,240	32,291	8,245	23,647	27,195	368,514
'Misplaced educators'	3,733	925	1,491	3,902	2,689	1,251	242	1,084	612	18,944
Misplaced as % of all	6.3	4.1	2.8	4.4	5.0	3.9	2.9	4.6	2.3	5.1
Coefficient of variation for ★	0.134	0.153	0.235	0.247	0.330	0.085	0.100	0.183	0.282	0.063

Note: The dependent variable in the regression is the number of publicly paid educators per school in 2013, according to the Snap Survey. Staff considered 'practitioners' were excluded from this variable.

Table 27: Regression outputs for post provisioning simulation (2002 policy)

	EC	FS	GP	KN	LP	MP	NC	NW	WC	SA
Slope coefficients										
Gr 1 to 4 learner★	0.020***	0.024***	0.022***	0.025***	0.027***	0.024***	0.022***	0.023***	0.021***	0.023***
Gr 5 to 6 learner★	0.027***	0.029***	0.022***	0.021***	0.021***	0.027***	0.025***	0.030***	0.020***	0.025***
Gr 7 learner★	0.024***	0.029***	0.031***	0.037***	0.040***	0.023***	0.024***	0.021***	0.034***	0.025***
Gr 8 to 9 learner★	0.021***	0.020***	0.020***	0.025***	0.020***	0.024***	0.021***	0.023***	0.019***	0.023***
Coefficient of variation for ★	0.137	0.171	0.207	0.257	0.341	0.071	0.079	0.163	0.300	0.048
Best fit year	2005	2005	2002	2005	2005	2002	2002	2002	2005	2002
Poverty slope coefficient	-0.001	0.028***	0.001	-0.002	-0.019***	0.004	0.023**	-0.005	0.009***	0.000
Best fit year		2005					2005		2005	

Table 26 above reflects the outputs of ten separate school-level regressions, one for each province and one for the country as a whole. The weights calculated according to the 2005 policy explain the number of publicly paid educators per school to a high degree. In the case of Western Cape 98% of the variation in educator numbers is explained by the weights. The lowest figure is that of Eastern Cape, at 89%. An alternative and more intuitive indicator of predictability, apart from the R^2 value, was calculated, namely the percentage of 'misplaced' educators. For this, a threshold of 0.5 educators was used. If the actual number of educators in a school exceeded the number of educators predicted by the regression plus 0.5, then this excess was counted within a category 'misplaced educators'. Thus if a school had 10 educators, and the model predicted it should have 9.1 educators, then 0.4 would be the excess, or the number of 'misplaced educators'. These educators come to between 6.3% (Eastern Cape) and 2.3% (Western Cape) of all educators.

If the around 8,300 schools offering Grade R, but reporting to have no 'practitioners' in the 2013 Snap Survey are excluded from the Table 26 regression, the regression outputs remain surprisingly unchanged. Only in four provinces does the R^2 value increase by more than 0.001, the largest increase being seen for Mpumalanga (here R^2 increased from the 0.943 seen in Table 26 to 0.947). One might have expected to see larger differences, given that educators teaching Grade R are not determined by the post provisioning model, at least not in the model as implemented in the current analysis[34]. The slope coefficients do not also not change in any noteworthy way if the 8,300 schools are excluded. The conclusion one can draw is that though the data problem inherent in the insufficient distinction between 'educator' and 'practitioner' is unfortunate, it seems not to influence the policy conclusions presented in the current report to any serious extent.

Table 27 provides outputs for two regressions using parameters from the 2002 post provisioning norms. In the first of the two regressions, the four variables indicated by ★ were replaced so that 2002 weights were used (the grades 10 to 12 variable was not switched as the 2002 version of this variable is not applicable to the new curriculum introduced between 2006 and 2008 in these grades). The analysis allows one to assess whether the actual distribution of educators in 2013 was closer to the 2002 or 2005 weights (as far as grades 1 to 9 are concerned). The closer the four coefficients are to each other, the closer the actual distribution of educators follows the policy weights. If a province followed, say, the 2005 weights perfectly in allocating educators, one would expect the four coefficients to equal each other. Mpumalanga is the province that gets closest to having coefficients equalling each other, in both of the above tables, and hence its coefficient of variation is low in both tables. But for Mpumalanga, it is lowest in Table 27, indicating that the distribution of educators follows the 2002 weights more closely than the 2005 weights. In all other provinces, there is greater deviation from the policy weights than in Mpumalanga, whether one considers the 2002 or 2005 weights. Yet if one compares the coefficients of variation in the last two tables, one can conclude that for a further three provinces (Gauteng, Northern Cape, North West) the 2002 weights are the most consistent predictors of education allocation, whilst for the remaining five provinces, the 2005 weights emerge as stronger. Neither the 2002 nor 2005 norms allow for provincial variation in the grades 1 to 9 weights, but the figures in the above two tables suggest that all provinces other than perhaps Mpumalanga either did apply somewhat different weights, or experienced greater problems in some grades than others in filling posts. Specifically, in all provinces either the grades 5 to 6 or the Grade 7 coefficient is the largest, suggesting that these grades, which get the lowest weights in the policy, are most likely to be adjusted upwards by provinces (whether the 2002 or 2005 coefficients were used here depended on which of the two years produced the lowest coefficient of variation).

[34] The model as stated in the policy provides for the option to generate Grade R posts through the model, but this option is reportedly not used by any province. Moreover, this option has not been used in simulations presented here.

The second 2002 regression switched only one variable, the variable dealing with the quintile weightings. Whether one uses the 2005 or 2002 weights, just three provinces emerge as having positive and statistically significant coefficients, namely Free State, Northern Cape and Western Cape. In all three provinces, the 2005 weights emerge as more statistically significant than the 2002 weights (using *p* values, which are not shown in the tables). What all this means is that these three provinces are the only ones which very clearly succeed in realising the pro-poor distribution of educators specified in the policy, and the three provinces follow the 2005 pro-poor weights more closely than the 2002 weights. Limpopo has a *negative* and statistically significant coefficient for the poverty variable (whether one uses the 2002 or 2005 policy parameters), suggesting that this province ends up resourcing better off schools better than poorer schools. With respect to the other five provinces, the patterns are inconclusive.

Table 28 below confirms that provinces approach the staffing of very small schools rather differently, and by implication depart in several instances from the national policy. Only 0.4% of learners nationally are in the one- or two-educator schools described in this table. Yet these schools often receive considerable policy attention because they so exceptionally difficult to operate well. To obtain the enrolment threshold seen in the second-last column the following was done. A first variable was calculated where 0 meant a one-educator school and 1 meant a two-educator school. A second variable was also calculated where 0 meant enrolment was *x* or fewer learners and 1 meant more than *x* learners. *x* was varied until the highest possible correlation between the two 0-1 variables was obtained. This analysis was repeated for each province. The second-last column displays *x*. The last column displays the correlation. A key finding is that Eastern Cape and Free State, two provinces with high numbers of small schools, use fairly high enrolment thresholds for a school to qualify for a second educator (26 and 23, respectively). KwaZulu-Natal, on the other hand, uses a very low threshold, resulting in very few one-educator schools. In fact, the average enrolment for two-educator schools in KwaZulu-Natal is about equal to the average enrolment for *one*-educator schools in Eastern Cape.

Table 28: Enrolment thresholds for one- and two-educator schools

	One-educator schools		Two-educator schools				
	Number	Mean enrolment	Number	Mean enrolment	% of all learners in these schools	Enrolment threshold between one- and two-educator schools	Correlation coefficient
EC	135	38	259	45	1.0	26	0.382
FS	245	13	110	32	1.1	23	0.588
GP	0		3	78	0.0		
KN	9	27	138	39	0.2	5	0.481
LP	21	32	91	45	0.3	12	0.416
MP	26	21	59	55	0.4	19	0.656
NC	1	31	37	41	0.6	31	0.228
NW	5	52	43	42	0.3		
WC	2	22	105	41	0.5	5	0.704
SA	444	23	845	42	0.4	17	0.541

What appears in the following table are the actual L/E ratios used for calculating the gap against equity-based norms seen in Table 12 above. Thus staff classified as publicly paid educators in the Snap Survey were used for 'E', whilst 'L' is the total of grades 1 to 12 enrolments per school.

Table 29: L/E ratios by school level in 2013

	Quintile 1	Quintile 2	Quintile 3	Quintile 4	Quintile 5	Overall
Schools with learners anywhere in grades 10 to 12 and nowhere in grades 1 to 4 (5,834 schools)						
EC	31.6	27.6	28.5	30.0	31.1	29.2
FS	24.4	25.0	25.6	24.6	25.8	25.1
GP	31.1	28.9	28.7	28.9	27.4	28.7
KN	29.6	29.8	30.9	29.3	29.6	29.9
LP	27.5	27.6	28.0	31.5	28.6	28.0
MP	27.7	28.7	28.2	27.8	26.5	28.1
NC	29.2	27.9	29.7	30.7	29.0	29.4
NW	28.3	28.9	27.8	28.2	28.6	28.3
WC	34.4	32.8	32.8	34.0	33.1	33.3
SA	28.6	28.6	29.1	29.6	29.3	29.0
All other schools, generally primary and combined primary-secondary schools (18,186)						
EC	28.8	25.3	30.4	33.7	32.7	28.8
FS	27.3	30.0	30.4	29.0	30.7	29.1
GP	36.0	37.0	35.6	37.0	37.9	36.6
KN	27.4	28.5	31.0	32.0	31.9	29.7
LP	28.1	28.7	29.0	31.5	33.7	28.8
MP	30.1	31.6	30.9	31.3	31.7	31.0
NC	31.5	32.1	33.1	33.6	34.3	32.7
NW	30.3	31.4	32.8	33.4	36.8	31.8
WC	31.3	35.2	34.9	35.6	37.3	35.4
SA	29.1	29.7	31.6	33.9	34.9	31.0

Source: Department of Basic Education, 2014.

In order to examine the relationship between the L/E ratio advantage of schools and learner performance, measures of learner performance were regressed on the L/E ratio *dis*advantage, meaning any value greater than zero, and a number of other variables. Table 30 below provides details of the five regression analyses run for Free State. Similar analyses were run for all provinces, and the country. A summary of all results appears in Table 31 below. Three different combinations of the explanatory variables were used (models A, B and C). In the first model, a linear relationship between the L/E ratio disadvantage and performance was assumed. Only positive L/E ratio disadvantage values were used, such as 3.4 meaning the L/E ratio was 3.4 learners higher than what the *provincial* application of the post provisioning model would predict (this was after the ideal per school educator count had been converted to an integer, as explained in section 6.1). Only positive values were used in the end because the aim of the analysis was to examine the relationship between *under*-staffing (relative to the norms) and learner performance. Other variables used in the first model were dummy variables for the quintiles, the average grade group size (total school enrolment divided by number of grades offered) to control for school size, and average grade group size squared. In a second model, the square of the L/E ratio disadvantage variable was added as it was possible that the relationship with learner performance was not linear. Here the joint significance of the untransformed and the squared versions of the variable was tested, using the 10% level of significance as the threshold. In the third model, distance to the closest school with a similar grade offering was introduced[35]. Model C was repeated for just quintiles 1 to 3 schools, and for just quintiles 4 to 5 schools. Table 30 indicates that for Free State in four of the five columns (all except the last) a statistically significant (conditional) correlation between the L/E ratio disadvantage and learner performance, in this instance in Grade 6 mathematics in the 2013 Annual National Assessments, was found. In all instances, a high L/E ratio (so under-staffing) is associated with lower than average performance in ANA, a relationship one would expect. The last row of the table indicates the change in ANA performance, in terms of a fraction of a national learner-level standard deviation, associated with having an L/E ratio

[35] Specifically, the other school considered had to have at least one grade in common with the reference school.

disadvantage of 4.0. The magnitudes are substantial if one considers that in South Africa learners gain roughly 0.3 standard deviations of knowledge in a year[36]. Thus one could say that the -0.14 values seen in models B and C point to a level of under-performance, associated with an L/E which exceeds the norm by 4.0, of half a year of learning. One cannot from this analysis conclude what causes what. The dynamics are complex. Under-performing schools perform poorly for a number of reasons, and teachers are likely to avoid such schools if possible, making it difficult for the province to fill posts in these schools. On the other hand, under-staffing is likely to be one of several factors which contributes to under-performance in a school.

Table 30: Regression of ANA results on L/E ratio disadvantage for Free State

	All quintiles			Quintiles 1 to 3	Quintiles 4 to 5
	Model A	Model B	Model C	Model C	Model C
L/E disadvantage	-0.446**	-0.725	-0.743	-0.571	-1.695
above squared		0.019	0.019	0.003	0.111
Above two jointly significant?	N/A	Yes	Yes	Yes	No
Is quintile 2	0.250	0.220	0.590	-0.500	
Is quintile 3	0.694	0.711	1.302	0.379	
Is quintile 4	11.162***	11.127***	11.804***		-12.503***
Is quintile 5	21.818***	21.842***	22.420***		
Average grade group size	-0.101***	-0.103***	-0.090***	-0.064**	-0.218**
above squared	0.000***	0.000***	0.000**	0.000*	0.001
Distance to closest school			0.300	0.400	0.330
Constant	45.192***	45.808***	44.058***	42.584***	76.200***
Adjusted R^2	0.212	0.211	0.210	0.044	0.295
Observations	362	362	359	293	66
Performance difference for L/E ratio disadvantage = 4.	-0.09	-0.14	-0.14	-0.12	

*Note: The dependent variable for all regressions shown here is the average Grade 6 mathematics score in the Annual National Assessments in 2013. Only schools with an L/E ratio difference value greater than zero were analysed, meaning only under-staffed schools were included. As was the case in other analyses, for instance that of Table 10, outlier schools with an L/E ratio difference exceeding 20 were excluded from the analysis. *** indicates that the estimate is significant at the 1% level of significance, ** at the 5% level, and * at the 10% level.*

The following summary of findings after similar regressions were run for all provinces, and the country, point to a couple of important patterns. Firstly, only in four provinces is the expected pattern of a higher L/E ratio associated with lower learner performance found. Secondly, this relationship emerges to a greater extent at the Grade 6 level.

[36] Spaull, 2015: 51.

Table 31: Summary of L/E ratio regression analyses

	All quintiles			Quintiles 1 to 3	Quintiles 4 to 5
	Model A	Model B	Model C	Model C	Model C
Dependent variable is school-level Grade 6 mathematics ANA average 2013					
EC					
FS	-0.09	-0.14	-0.14	-0.12	
GP	-0.08	-0.02	-0.03		
KN					
LP					
MP	-0.06	0.02			
NC					
NW					
WC					0.25
SA		-0.01	-0.01	-0.02	
Dependent variable is school-level Grade 12 mathematics performance 2013					
EC					0.95
FS					
GP	-0.27				
KN					
LP					
MP					
NC					
NW					
WC				-1.23	
SA					

Note: The 10% level of significance is the threshold used for reporting values in this table. The Grade 12 mathematics performance indicator used is the 95th percentile of the score in mathematics, but using total Grade 10 enrolment two years previously (so in 2011) as the reference population and assigning a score of zero to students not taking mathematics in 2013. This indicator has emerged as a particularly robust indicator of Grade 12 performance, and is discussed in a 2015 report titled 'Treating schools to a new administration: Evidence from South Africa of the possible impact of better practices in the system-level administration of schools', produced, like this report, within the PSPPD project. The striking WC Grade 12 figure of -1.23 emerges from a regression including just 45 schools. It is likely to be driven by just one or two unusual schools in the dataset, as opposed to a widespread phenomenon. The exclusion of extreme outliers discussed in the note for the previous table was applied to all the provincial regression analyses.

Turning to the matter of lags between enrolment patterns and educator distributions (see section 6.3 above), the following table shows 21 adjusted R^2 values obtained from 21 regressions, where the actual number of publicly paid educators in one year across schools was regressed on enrolment-based post provisioning weights from the same year, or from a previous year. Within each column, the highest R^2 value is marked in bold. A clear pattern emerges suggesting that there is a lag of two years between enrolment patterns and educator distributions in KwaZulu-Natal. Put crudely, educators do follow learners, but with a delay of two years. The lags only become trustworthy from the 2010 column onwards, given that the historical pattern is a lag of two years. The lag of one year in the 2009 column is almost certainly a result of the fact that 2007 enrolment data were not used in the analysis.

Table 32: Post provisioning lags analysis for KwaZulu-Natal

Educators ➔ Enrolment 🡻	2008	2009	2010	2011	2012	2013
2008	**0.9042**	**0.9074**	**0.9541**	0.8495	0.9395	0.9186
2009		0.8930	0.9435	**0.8503**	0.9539	0.9347
2010			0.9263	0.8499	**0.9654**	0.9521
2011				0.8447	0.9637	**0.9616**
2012					0.9492	0.9520
2013						0.9378
Lag in years	[0]	[1]	2	2	2	2

Appendix C: District codes used in maps

The following district codes were used in, for instance, the maps appearing in section 6.1.

Table 33: District codes used in the report

Prov.	Dist. code	District	Prov.	Dist. code	District
KN	AM	Amajuba	FS	MH	Motheo
NW	BJ	Bojanala	WC	MN	Metro North
MP	BO	Bohlabela	LP	MO	Mogalakwena
EC	BU	Butterworth	LP	MP	Mopani
LP	CA	Capricorn	WC	MS	Metro South
EC	CO	Cofimvaba	EC	MT	Mthatha
EC	CR	Cradock	NC	NA	Namakwa
WC	CW	Cape Winelands	EC	NB	Ngcobo
EC	DU	Dutywa	NW	NG	Ngaka Modiri Molema
EC	EA	East London	MP	NK	Nkangala
WC	ED	Eden and Central Karoo	WC	OV	Overberg
MP	EH	Ehlanzeni	KN	PI	Pinetown
GP	EN	Ekurhuleni North	EC	PO	Port Elizabeth
GP	ES	Ekurhuleni South	NC	PS	Pixley Ka Seme
FS	FE	Fezile Dabi	EC	QT	Queenstown
EC	FL	Mt Fletcher	EC	QU	Qumbu
EC	FO	Fort Beaufort	LP	RI	Riba Cross
NC	FR	Frances Baard	NW	RU	Dr Ruth Segomotsi Mompati
GP	GE	Gauteng East	GP	SE	Sedibeng East
GP	GN	Gauteng North	KN	SI	Sisonke
EC	GR	Graaff-Reinet	LP	SK	Sekhukhune
MP	GS	Gert Sibande	EC	ST	Sterkspruit
EC	GT	Grahamstown	GP	SW	Sedibeng West
GP	GW	Gauteng West	NC	SY	Siyanda
KN	IL	Ilembe	FS	TH	Thabo Mofutsanyana
GP	JC	Johannesburg Central	GP	TN	Tshwane North
GP	JE	Johannesburg East	LP	TP	Tshipise Sagole
GP	JN	Johannesburg North	GP	TS	Tshwane South
NC	JO	John Taolo Gaetsewe	GP	TW	Tshwane West
GP	JS	Johannesburg South	LP	TZ	Tzaneen
GP	JW	Johannesburg West	KN	UG	Ugu
NW	KE	Dr Kenneth Kaunda	EC	UI	Uitenhage
EC	KI	King Williams Town	KN	UK	Umkhanyakude
EC	LA	Lady Frere	KN	UL	Umlazi
LP	LE	Lebowakgomo	KN	UM	Umgungundlovu
EC	LI	Libode	KN	UT	Uthukela
FS	LP	Lejweleputswa	KN	UU	Uthungulu
EC	LU	Lusikisiki	KN	UY	Umzinyathi
EC	MA	Maluti	LP	VH	Vhembe
EC	MB	Mbizana	LP	WA	Waterberg
WC	MC	Metro Central	WC	WE	West Coast
WC	ME	Metro East	FS	XH	Xhariep
EC	MF	Mt Frere	KN	ZU	Zululand

A few relatively small districts in Gauteng do not have district codes attached to them in the country maps, but their positions are shown in the following map, which zooms into this province.

Figure 47: Map of Gauteng districts