
An overview of Education data in South Africa: an
inventory approach

CHRIS VAN WYK

Stellenbosch Economic Working Papers: 19/15

KEYWORDS: EDUCATION MANAGEMENT INFORMATION SYSTEMS, EDUCATION
DATA, PERFORMANCE DATA, DATA INTEGRATION, UNIQUE IDENTIFIERS,
DATA QUALITY
JEL: I20, I29

CHRIS VAN WYK
DEPARTMENT OF ECONOMICS
UNIVERSITY OF STELLENBOSCH
PRIVATE BAG X1, 7602
MATIELAND, SOUTH AFRICA
E-MAIL: CHRISVAN WYK@SUN.AC.ZA



UNIVERSITEIT
STELLENBOSCH
UNIVERSITY



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

An overview of Education data in South Africa: an inventory approach

CHRIS VAN WYK*

ABSTRACT

This paper provides an overview of the various datasets pertaining to education in South Africa that are informing or could inform policy making in education. The paper serves as an inventory for anyone interested in understanding what data is available, how it may be accessed, what the quality of the data is and in what formats it may be accessed.

The paper is divided into three parts. The first part provides a description of existing education datasets and the basic data elements contained in each of these datasets. When discussing each of the existing education datasets, the paper addresses the quality of the education data available in South Africa. The first part also refers to the policy implications and the important role that data plays in policy-formulation. No information system on its own is comprehensive enough to provide all the information needed in strategic decision-making. Hence, part two of this paper discusses the need for data integration as an important data management strategy. The third part examines the effectiveness of implementing a learner unit record system nationally in comparison with the EMIS system that is currently in place and that is based on aggregate or summary institution-level data.

Keywords: education management information systems, education data, performance data, data integration, unique identifiers, data quality
JEL codes: I20, I29

* 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.

Introduction

This paper provides a broad overview of education data sets in South Africa, predominantly data from the Department of Basic Education (DBE). Education data, specifically the datasets generated by the annual school survey in education management information system (EMIS) in most countries of the world, is often one of the most under-utilised data sources for both policy and research. Since its establishment in 1994, the Department of Basic Education (DBE) has been challenged with providing relevant and quality data as efficiently as possible to information users, particularly planners (for evidence-based planning) at all levels, decision makers and policy developers. A further challenge entailed the development of mechanisms for accountability and monitoring and evaluation of the education system. The realization of these challenges has been aided significantly with the establishment of Education Management Information Systems (EMIS) in 1995. The Education Management Information System (EMIS) is the unit in the national Department of Basic Education (DBE) and provincial departments that is responsible for the producing, managing and disseminating of education data. An important step in gaining an understanding of what is meant by EMIS within a national and provincial context is to consider a working definition. Many authors have attempted to define and describe EMIS (Carrizo, Sauvageot & Bella, 2003; Hau & Herstein, 2003) however; this paper will incorporate the following definition:

An Education Management Information System (EMIS) is a system for the collection, integration, processing, maintenance and dissemination of data and information to support decision making, policy-analysis and formulation, planning, monitoring and management at all levels of an education system. It is a system of people, technology, models, methods, processes, procedures, rules and regulations that function together to provide education leaders, decision makers and managers at all levels with a comprehensive, integrated set of relevant, reliable, unambiguous, and timely data and information to support them in completion of their responsibilities (Cassidy, 2005).

The objective of an education management information system (EMIS) should not only be to collect, store and process information but also to assist in education policy-making, by providing relevant and accessible information for research projects.

This paper serves as an inventory for anyone interested in understanding what education data is available in South Africa, how it may be accessed, what the quality of the data is and in what formats it may be available.

The paper is divided into three parts. The first part provides a description of existing education datasets as well as the basic data elements contained in each of these datasets. Proper management, planning and policy-formulation are contingent on quality data. When discussing each of the existing education datasets, this paper addresses the quality of education data available in South Africa in terms of completeness, relevance, accuracy and timeliness. How good is the quality of the EMIS data in South Africa? This paper draws from the author's experience in using and analysing the EMIS data and can report that the quality of EMIS data in South Africa is of acceptable standard and has improved over time. In analysing the education enrolment ratios Martin Gustafsson (2012) concurs that the quality of school census data is improving. The first part further refers to the policy implications and the important role that data plays in policy-formulation.

Part two of this paper discusses the need for data integration as an important data management strategy. This makes it imperative to use unique identifiers in the data integration process, as will be explained.

The third part examines the effectiveness of implementing a learner unit record system nationally in comparison with the EMIS system that is currently in place and that is based on aggregate or summary institution–level data.

Education Datasets in South Africa

Data is a necessary resource, produced by information systems and is a key building block to the management and decision-making in any organisation. An information system in the context of this paper includes the technology, the people, processes as well as the information. In South Africa different datasets exist that are generated by various information systems.

This section of the paper provides a description of the existing education datasets and the basic data elements contained in each of these datasets. The primary purpose is to provide a compilation of education datasets covering the broad field of South African education data. These datasets include a selection of data from many different systems and draws specifically on the outcomes of surveys and operational activities carried out by the Department of Basic Education (DBE).

The list of education datasets discussed in this paper is not exhaustive, but for the purpose of this paper it includes the data generated by education management information system (EMIS), the learning outcomes data, including international test data, such as PIRLS, TIMSS, SACMEQ, and key performance data in South Africa such as Annual National Assessment (ANA) and National Senior Certificate (NSC).

In the ensuing paragraphs we will look at the data elements, data quality, data integration and policy implications of these different datasets.

Master list of schools

The master list of schools in South Africa is a record of each school in the country. The list is maintained by provincial departments and regularly sent to DBE for updating. A key function of the master list is to uniquely identify each school in the country through a school identifier, generally called the “emis number”. Furthermore, the master list of schools has a specific number of key fields such as the school’s examination number which is specifically used to link school administrative data with the examination data. There are also other basic data fields in the school master list that could provide the means to answer some of the most frequently asked questions about the socio-economic status, learner enrolment, teachers and learner-teacher ratio of schools. **Table 1** provides a brief description of some of the data elements included in the master list of schools.

Table 1: *Data elements included in the master list of schools*

Data Element	Description
Natemis	A unique number assigned to each school. The importance and the use of a unique school identifier is discussed in this paper
Sector	Public or private

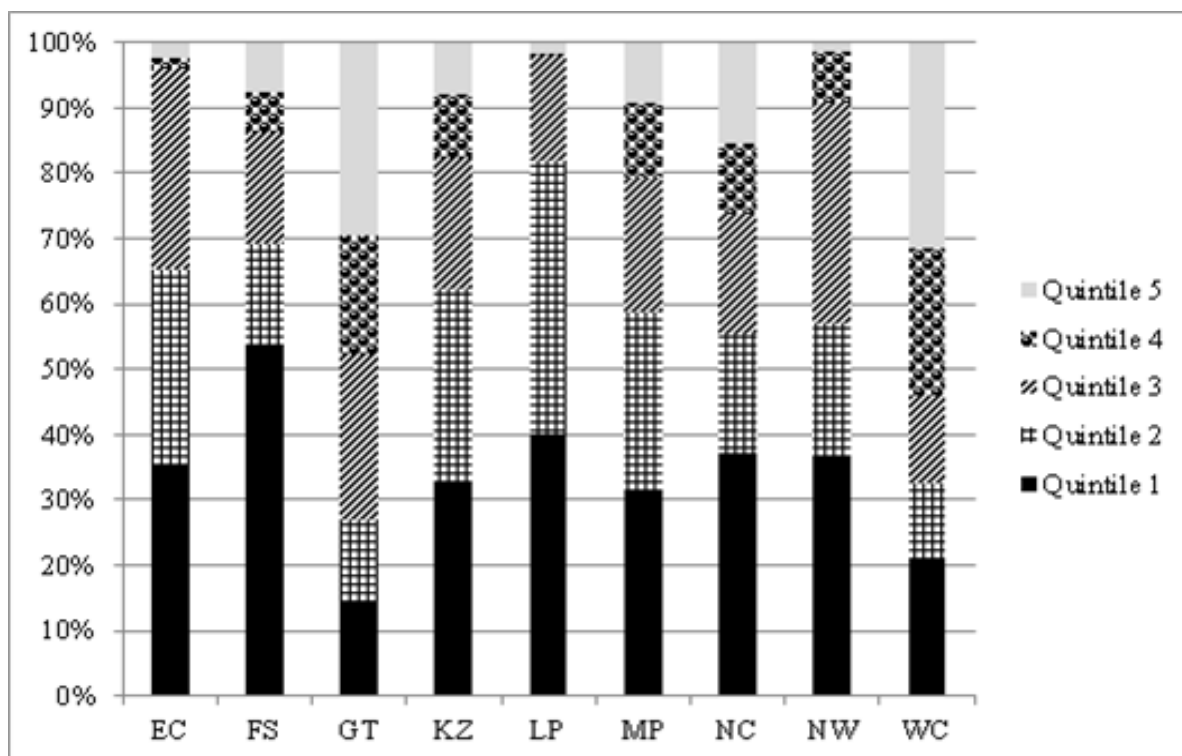
Phase	The type of school (e.g. primary)
ExDept	Indication of the Education Authority before 1994
PaypointNo	The unique number of an institution as assigned in the personnel salary system (PERSAL)
ComponentNo	Another unique number of an institution as assigned in the personnel salary system (PERSAL). This is the element used to link EMIS and PERSAL data. The quality of this data element in terms of completeness and accuracy is questionable
ExamNo	The unique number of an institution as assigned by Nation Senior Certificate Examinations. This is the data element used to link EMIS and NSC examination data. The quality of this data element in terms of completeness and accuracy is questionable.
GIS_Longitude	Geographical coordinates of the school
GIS_Latitude	Geographical coordinates of the school
Quintile ¹	This is an indication of the socio-economic status of the school. Schools are ranked according to poverty of the school community.

Source: Master list of schools 2015 (South Africa, DBE, 2015a).

Figure 1 below indicates the distribution of schools according to quintiles and shows the descriptive statistics possible from the master list of schools. Note that schools quintiles are often used in multivariate analyses as control measures for school performance. South African schools are divided into five categories (quintiles) based on the socio-economic status of the community in which the school is situated. Quintile 1 schools are the poorest while quintile 5 schools are the least poor.

Figure 1: *Distribution of schools by province and quintile*

1. One of five poverty-based categories to which public schools are allocated for purposes of non-personnel funding by provincial education departments. For example, quintile 1 schools are the poorest and quintile 5 schools are the least poor (South Africa, DBE, 2006).



Own calculation based on data in master list of schools

The master list of schools is a data table that is available through DBE and is accessible through their website (South Africa, DBE, 2015a). The master list is available per province for 2015, 2014 and 2013. The 2014 and 2013 lists are more comprehensive than the most recent one and include learner enrolment and teacher totals for a number of years, which makes the calculation of learner-teacher ratio per school possible.

Although the master list of schools is publicly available on the website of the Department of Basic Education, the quality of the data is of concern. The relevance and the completeness of the 2015 master list could increase with the inclusion of school enrolment and number of teachers for a few years, as was the case in previous master lists. The master list of schools should also be available earlier in the year. This is a useful dataset for education planners and researchers and it is even widely used in the private sector for those who regularly deal with schools. The master list of schools provides researchers with additional information about schools that is not always available in school performance data, such as socio-economic status of the school, the location of the school and the school size. The master list of schools is also used to match school data across years and link it with other datasets such as examination data, using the unique school identifier.

SNAP data

The Annual SNAP survey is the data about the schools as recorded on the 10th school day every year. The SNAP survey form issued by the DBE (2014a:1) clearly states that:

The Annual SNAP Survey for Ordinary Schools is the most important source of information regarding the situation at your school. The information collected will contribute to the Education Management Information System (EMIS) of the province and will be of assistance in the management, administration and governance of schools, including the supply of school resources to schools.

The South African Schools Act No. 84 of 1996 (section 59 [1] and [2]) states that:

Every school must supply such information about the school as is reasonably required by the Head of Education, and any person, who without just cause, fails to comply, shall be guilty of an offence (South African Schools Act, 1996).

The SNAP data is an important source for the allocation of funds per learner based on the National Standards for School Funding Norms; the allocation of teachers (post provisioning) to schools; and the annual publication of education data. The *Education Statistics in South Africa* for each year from 2008 to 2013 (South Africa, DBE, 2015b) are publications based on the data collected via the SNAP survey. **Table 2** indicates the important data elements included in the SNAP survey.

Table 2: *Data elements included in the SNAP survey*

Data Element	Grade	Gender	Race	Province	Public	Private	Governing Body	Permanent	Temporary
Enrolment	X	X		X	X	X			
Educators		X		X	X		X	X	X
Practitioners		X		X	X		X	X	X
Professional non-teaching staff		X		X	X		X	X	X
Admin. Staff		X		X	X		X	X	X
Support staff		X		X	X		X	X	X
Hostel staff		X		X	X		X	X	X

Source: SNAP survey questionnaire

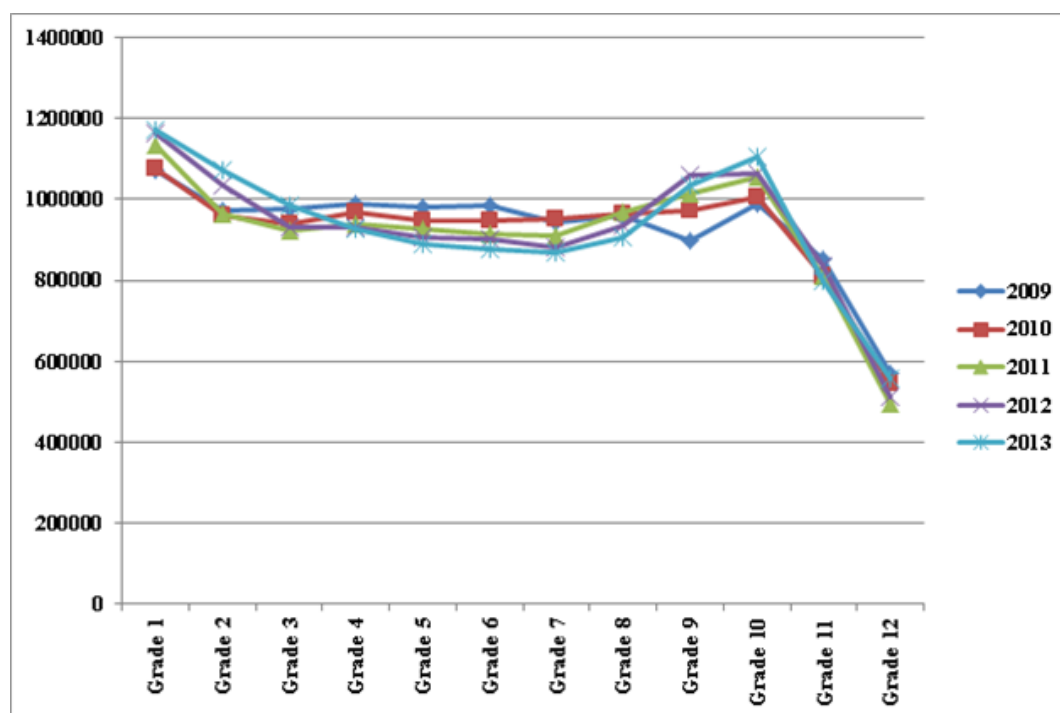
Table 3 provides a summary of all learners in the public education system as derived from the SNAP surveys by grade for the years 2002-2013. Some of the flow-through trends that can be observed in the South African Education system are graphically illustrated by **Figure 2** based on the data from the SNAP. Although the graph is a series of cross sections (it does not follow the same cohort of learners, but only gives the number of learners in each grade for each year), the overall picture gives a good indication of the trends and patterns in the entire education cycle in South Africa. The patterns appear to be quite stable as indicated by **Figure 2**. From this data, it is possible to determine the enrolment growth by grade for any given period. The data for the twelve grades (full education cycle) is shown, indicating how enrolments in each grade over the basic education cycle changed. During the first part of the education cycle, grade 1 to grade 10, the system seems to be successful in keeping the learners in school with a high dropout of learners after grade 10. This is typical of the school system of South Africa: successful in retaining learners in primary school with a high dropout rate in the secondary phase.

Table 3: *Enrolment for public ordinary schools based on SNAP data*

Year	Grade 1	Grade 2	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8	Grade 9	Grade 10	Grade 11	Grade 12	Grand Total
2002	1264558	988046	924504	1051835	1118928	1015999	938073	912066	1064720	852976	653309	463223	11248237
2003	1266052	1099903	983064	934031	1015741	1080302	967759	953949	881813	1077854	715732	452798	11428998
2004	1270601	1080040	1008641	959378	891987	970320	1021425	978170	885057	1025839	798431	476410	11366299
2005	1207799	1094154	1054127	1038491	930745	878535	951489	1023329	905393	1043081	814589	511884	11453616
2006	1157170	1056042	1073683	1048398	1002756	898827	851511	992807	943509	1065101	864230	540532	11494566
2007	1141412	1022609	1040508	1064000	1009885	977269	874200	904411	931360	1086738	891243	593639	11537274
2008	1090254	1002847	989809	1023963	1016196	975832	939116	899097	877143	1047874	873125	566460	11301716
2009	1072993	972909	975070	991195	981907	985273	944125	958565	898019	988207	851525	570849	11190637
2010	1077532	958939	938796	970260	947994	948254	950675	965394	974521	1006549	808997	543487	11091398
2011	1133797	965334	921715	940845	925305	915487	910994	970854	1013033	1055790	812627	496593	11062374
2012	1162820	1033562	930238	931603	906325	904256	882079	934885	1058852	1065329	835939	512303	11158191
2013	1172878	1070954	984375	927562	889301	876862	871093	905569	1036555	1106913	797304	556445	11195811

Source: SNAP survey

Figure 2: Enrolment in public schools by grade and year, 2009-2013



Data on enrolment is important for many different purposes. However, it is not easily available (such as on the DBE's website) in a user friendly format, such as a database, a comma separated values (csv) file or in spreadsheet at school level. The *School Realities* (South Africa, DBE, 2015c) publication, based on the SNAP data, is a valuable data reference, but is only available on the website of DBE in a pdf-format that is difficult to manipulate. If the DBE for example makes enrolment data available per school in a more accessible file format on their website it could help to promote the greater use of data in planning and decision-making. Here the approach of DataFirst (2015) could be followed that publishes the SNAP data in an easily manipulated data file format on their data portal,

providing researchers access to South African enrolment data. DataFirst is a data service based at the University of Cape Town, South Africa that gives researchers online access to African data and online help with data usage.

Annual School Survey (ASS)

The data available through the Annual School Survey is a useful resource to determine overage, enrolment, repetition and dropout rates by gender and province. It is well-known that dropout in South Africa is extensive. A close inspection of school data shows that of the 100 pupils that start grade one, 50 will drop-out before Grade 12 (most of which happens in grade 10 and 11), 40 will pass the NSC examination and 12 will qualify for university (Spaull, 2013a). With the availability and the quality of the data from the Annual Schools Survey key questions can be answered such as: “Where in the system is the highest dropout and repetition?”

The Education Management Information System (EMIS) is the primary national process for collecting and reporting data related to schools in South Africa. An important function of EMIS is that education departments in provinces annually collect data from all schools through a comprehensive survey. The Annual School Survey (ASS) is a comprehensive survey of all public and independent schools in the South Africa. It was designed to provide comparable information on public and private sectors, as well as trend data over time. The Annual School Survey is completed by all schools in the country on a specific day, usually in March. The *Annual School Survey Form* (South Africa, DBE, 2015d) and the *Dictionary of Education Concepts and Terms 2010* (South Africa, DBE, 2015e) provide a comprehensive insight of the data that is available. The **Table 4** below shows some of the key data elements included in the Annual School Survey.

Table 4: *Data elements included in the Annual School Survey*

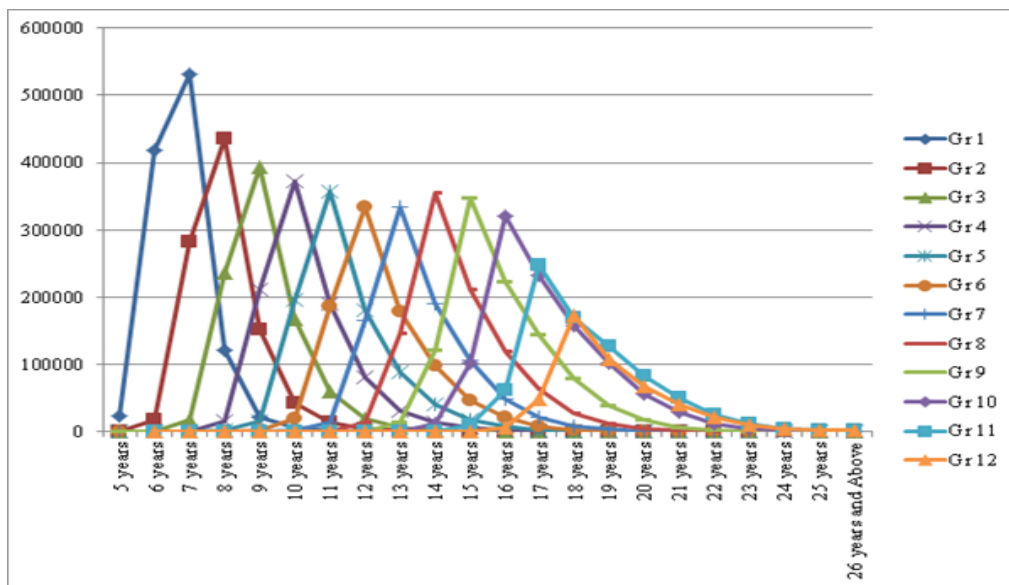
Data Element	Total	Province	Grade	Gender	Public	Private	Disability	Type
Computers	X	X						
Early Childhood Development (ECD)		X			X	X		
Enrolment		X	X	X	X	X	X	

Enrolment by Race		X	X	X	X	X		
Enrolment by Age		X	X	X	X	X		
Enrolment by Home Language		X	X	X	X	X		
Enrolment by Language of Learning and Teaching								
Grade 1 First time enrolment		X		X	X	X		
Gr 1 attended Gr R		X		X	X	X		
Repeaters		X	X	X	X	X		
School Fees		X	X		X	X		
Learner Migration		X			X	X		
Hostel Enrolment		X			X	X		
Learner Pregnancy		X	X		X	X		
Enrolment by Social Grant		X	X		X	X		
Enrolment by Parent(s) deceased		X	X	X	X	X		
Learner Mortality		X	X	X	X	X		X
Teacher Mortality		X		X	X	X		X
Learner Subjects by Phase and Race		X		X	X	X		

Source: Annual School Survey Questionnaire for 2014

Figure 3 is based on the data obtained from the Annual School Survey and shows the patterns of enrolment by grade and age. In lower grades, most learners are in the grade appropriate for their age (**Figure 3**), but due to repetition (and perhaps also drop-out) there is by grade 10 a much wider age range, indicating an overage (repeater) problem in the higher grades.

Figure 3: Enrolment by grade and age in 2011



Source: ASS data for 2011

Importance of the School Survey Data

Enrolment-driven data management is a central focus of South African government's redress efforts, mainly because of state based accountability policies, such as the South African Schools Act and broader education policy. The enrolment-driven nature of the education system is reflected in the allocation of school funds based on the National Norms and Standards for School Funding (South Africa, DBE: 2006), the post-provisioning of teachers, the grading of primary and secondary schools and many more. The Department of Basic Education has recognised the important role that education statistics can play in the education

National Senior Certificate (Gr 12)	X			X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	
Annual National Assessment (Grade 1 to 6) and Grade 9	X																			X	X	X	X
Systemic Test ² Grade 3			X						X		X		X		X		X	X	X	X	X	X	X
Systemic Test Grade 6			X								X		X		X		X	X	X	X	X	X	X
Systemic Test Grade 9			X														X	X	X	X	X	X	X
SACMEQ (Grade 6)		X						X							X								X
TIMSS (Gr 9)		X		X				X			X												X
PIRLS (Grade 5)		X											X										X
PRE-PIRLS (Grade 4)	X																						X

The National Senior Certificate (NSC)

The Senior Certificate results are available per subject for every learner that wrote the examination since 1994. There is a growing research on student achievement and school performance in South Africa and it is of particular importance for this paper (Taylor, S., 2014; Van der Berg & Shepherd, 2010; Simkins, 2010; Van der Berg, 2007; Crouch & Mabogoane, 2001).

The *Schools Report* (South Africa, DBE, 2015e) is a publication of the performance of individual schools in the National Senior Certificate for the last 3 years, 2012 to 2014. The report includes data elements such as official school number, quintile, number of students that wrote and number of student that passed for individual schools. The *School Subject Report* (South Africa, DBE: 2015f) is another useful publication and an important data source on the National Senior Certificate results. The report includes key subjects with data elements such as year of examination, quintile of the school, number that wrote and number that passed for specific key subjects. Such data accessibility and availability on school and subject performance are extremely useful and could contribute to growing use of the National Senior Certificate (NSC) results, as it allows analysts to determine patterns and trends in school performance. **Table 6** and **Figure 4** indicate the general trend for NSC examination. With the exception of 2011, where fewer Matrics passed Matric than in the previous year partly due to fewer Matrics writing the Matric exam, the number of individuals passing Matric has generally been on the increase over the past five years. In fact, the proportion of Matrics passing the national Matric examination has increased by a staggering 17.9 percentage points, from 60.3% in 2009 to 78.2% in 2013.

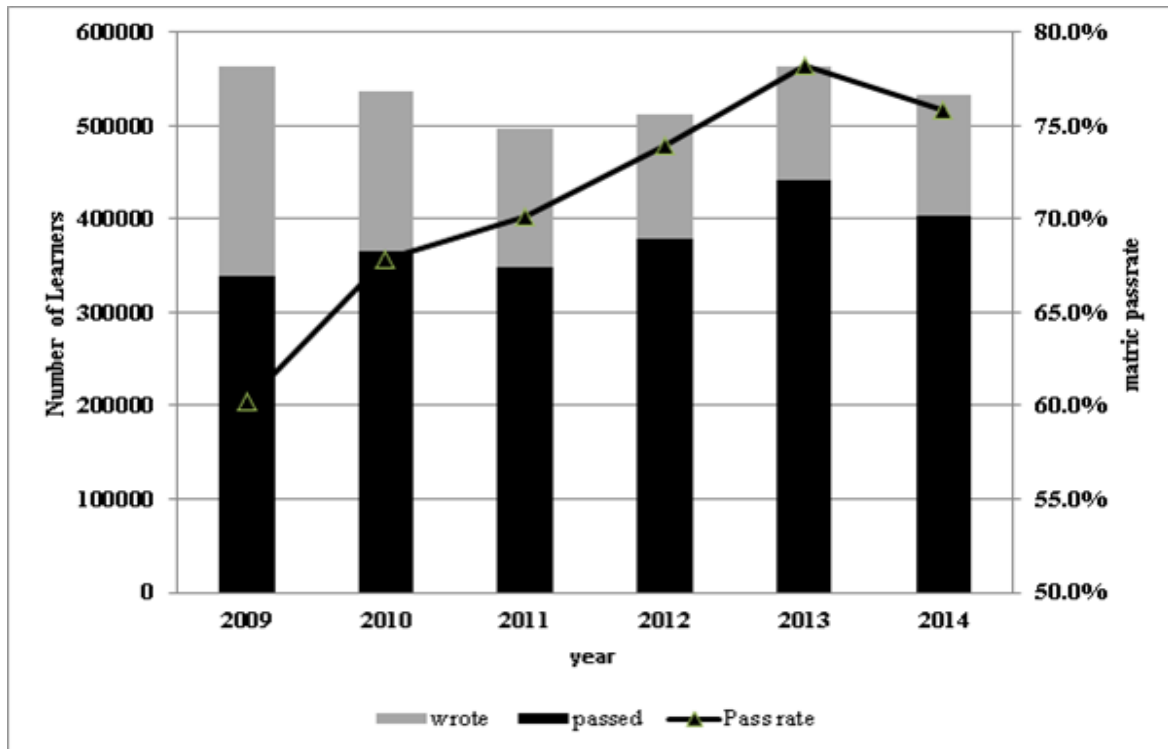
Table 6: NSC results (2009-2013) in all schools

Year	Wrote	Passed	Pass rate
2009	562750	339144	60.3%
2010	537543	364513	67.8%

2. Systemic Tests are standardised tests in the Western Cape for Language and Mathematics in Grades 3, 6 and 9.

2011	496090	348117	70.2%
2012	511152	377829	73.9%
2013	563462	440877	78.2%
2014	532860	403874	75.8%

Figure 4: NSC results (2009-2013) in all schools



Source: Own calculations from matriculation data

A further factor that could contribute to greater use of Matric examination data would be if it were available in a more user friendly data format, such as a database, spreadsheet or even a comma delimited text file on the DBE's website that could be downloaded, which is not currently the case.

Annual National Assessment (ANA)

The Annual National Assessments (ANA) (South Africa, DBE, 2015g) are standardised national assessments for languages and mathematics in the senior phase (grades 7 - 9), intermediate phase (grades 4 - 6) and in literacy and numeracy for the foundation phase (grades 1 - 3). The question papers and marking memoranda (exemplars) are supplied by the national Department of Basic Education and the schools manage the conduct of the tests as well as the marking and internal moderation itself, something that has raised some issues regarding the quality of such data. Up until 2011 the only standardised national examination that existed in the country was the National Senior Certificate. As a result it was not really possible to determine the performance of primary schools, or to hold schools accountable for learning performance or to provide support to schools where it was needed most.

The Department of Basic Education has released reports for the ANA 2012 (South Africa, DBE, 2012), ANA 2013 (South Africa, DBE, 2013) and ANA 2014 (South Africa, DBE,

2014c) results which outline their rationale, methodology, curriculum interventions and results.

The Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ)

The Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ, 2015) is an international non-profit developmental organization of 15 Ministries of Education in Southern and Eastern Africa that decided to work together to share experiences and expertise in developing the capacities of education planners to apply scientific methods to monitor and evaluate the conditions of schooling and the quality of education. SACMEQ's rich data sets provided new possibilities for investigating relationships between educational outcomes, socio-economic status (SES), pupil and teacher characteristics, school resources and school processes (Van der Berg, 2006).

There is a growing body of literature where SACMEQ data is used in empirical studies to evaluate the learner in South Africa (Gustafsson, M, 2005; Moloi & Chetty, 2011; Shepherd, 2015; Spaul, 2014, 2013b, 2012; Spaul & Taylor, S., 2012; Van der Berg, 2008; Van der Berg, S. & Louw, M. 2007).

Progress in International Reading and Literacy Study (PIRLS)

PIRLS (2015) is an international comparative study aimed at benchmarking literacy/reading levels across countries. It tests the reading literacy of grade four children from 49 countries. This world-wide assessment and research project is designed to measure trends in children's reading literacy achievement and collect information about policy and practices related to learning to read and reading instruction. The PIRLS results can be used for issues related to LOLT and to track reading performance over time. It a useful tool to do international comparisons. Various local researchers have analysed the PILRLS datasets and provided valuable insight of the state of education in the country (Shepherd, 2011, 2013; Taylor, S. & Yu, 2009).

prePIRLS is a stepping stone to participating in PIRLS and provides a way to assess reading at the end of the primary school cycle for a range of developing countries. prePIRLS reflects the same conception of reading as PIRLS, except it is less difficult. Participation in prePIRLS prepares countries for moving toward participation in PIRLS. The prePIRLS assessment is intended to provide valuable diagnostic information about strengths and weaknesses in students' reading skills and important policy information about the necessary steps to improve students' reading (prePIRLS, 2015).

There is comprehensive supporting documentation on PIRLS such as the User Guide for the International Database, the codebooks and the data files are publicly available on the web site (PIRLS, 2015).

Trends in International Mathematics and Science Study (TIMSS)

TIMSS tests mathematics and science achievement of grade 8 and 9 learners in South Africa. The test is useful to monitor and evaluate the quality of schooling in specific areas of the curriculum at given points in time. South Africa has participated in TIMSS assessments: 1995, 1999 (grade 8 only), 2002 (grade 8 and 9) and 2011 (grade 9 only). The next round of TIMSS in South Africa will be administered in 2015.

The new TIMSS Numeracy assessment has been designed to be administered at the Grade 4, 5 or 6 level to measure children's numeracy learning outcomes. TIMSS Numeracy was

conducted for the first time in South Africa in 2014 at the Grade 5 level (TIMSS SA, 2015), but results have not yet been released.

There is comprehensive supporting documentation on TIMSS such as the User Guide for the International Database, the codebooks and the data files are publicly available on the web site (TIMSS, 2015). The contribution of TIMSS data in understanding education in South Africa is widely researched (Reddy, Zuze, Visser, Winnaar, Juan, Prinsloo, Arends & Rogers 2015; Reddy, Van der Berg, Janse van Rensburg & Taylor S. 2012).

The importance of performance data

The availability of performance data makes it possible, among other, to determine if learners are learning the content at the appropriate levels. Performance data also forms the basis for accountability. Furthermore, it assists to compare learners nationally, and also internationally, and helps to determine intervention strategies for improving instruction.

Van der Berg (2007:876) affirms the importance of quality data on learner outcomes when he states that:

The education authorities have a paucity of information for decision making. Identifying under-performing schools in order to take remedial action requires understanding school performance. Educational policy improvements need evidence on what works and what does not, and '(d)eveloping such evidence means that regular high quality information about student outcomes must be generated'.

Data Integration

The EMIS datasets, such as the annual school survey (ASS), or the performance datasets, such as the National Senior Certificate (NSC) examination results, on their own are not comprehensive enough for sound education planning, monitoring and policy formulation. There is also not one information system in the Department of Basic Education (DBE) containing all the relevant information to promote the utilization of data in planning and evidence-based decision-making. Therefore relevant information has to be extracted from different information systems. This illustrates the necessity and value of integrating different data sources to the decision- and policy-making process.

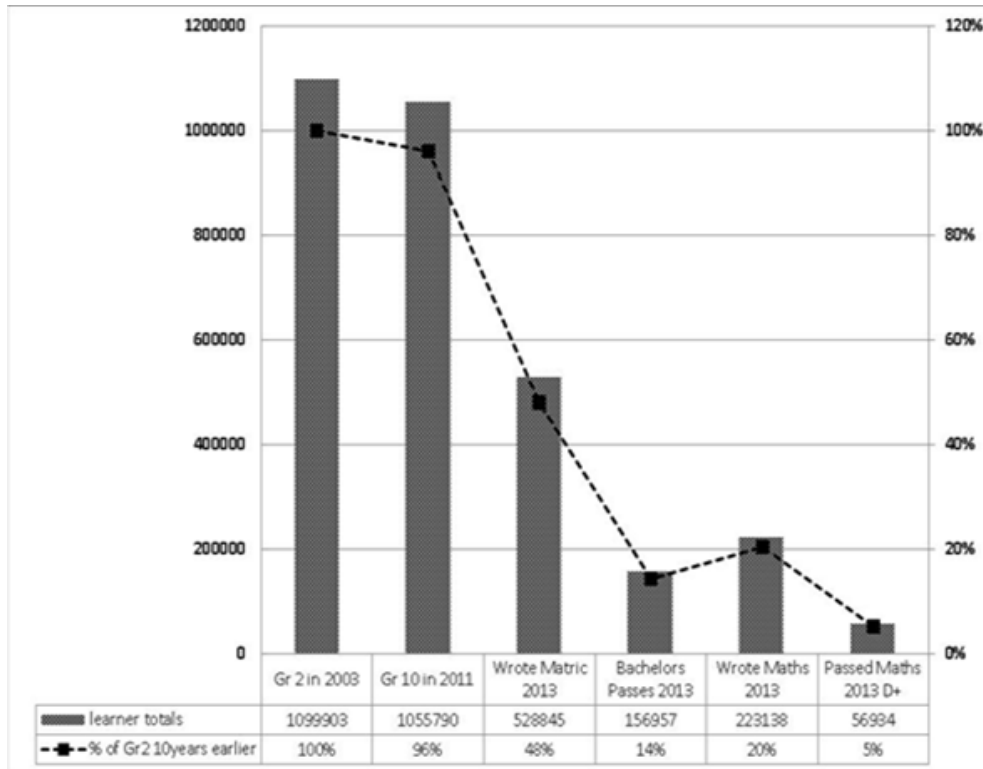
Data integration means that data from multiple sources (EMIS, Examinations, etc.) can be linked, integrated, or merged through the use of a common field across a collection of data sources. Data integration is intended to add value to the data that is already collected and available in various systems. All the datasets mentioned will remain disparate databases, fragmented islands of data and will exist in isolation from other data if they are not linked. This limits their potential use for informing management decision making.

All these separate datasets from the different information systems or even from different years need to be seamlessly brought together, integrated through a unique school or learner identifier.

Figure 5 is a typical example of data from the National Senior Certificate (NSC) combined with the EMIS data to provide an overview of the education system. It is clear from the graph that there is a lot of attrition between grade 10 and grade 12. For example, there were 1 055 790 grade 10 enrolments in 2011 but only 528 845 grade 12 enrolments two years later

in 2011 – roughly half. Some of this, however, would not be attrition but a reflection of high grade repetition in Grade 10 and 11.

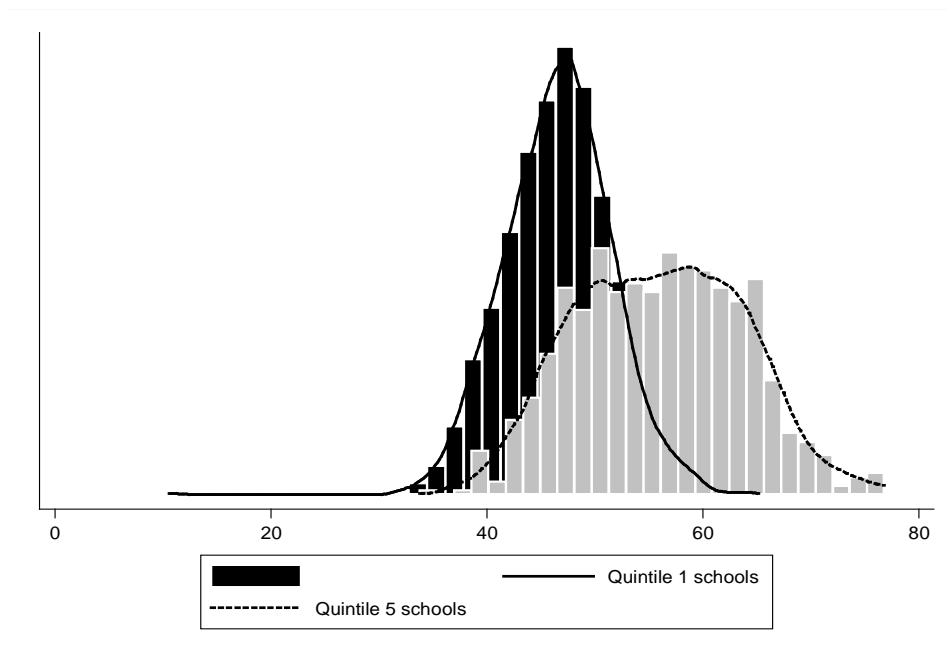
Figure 5: The ratio of grade 2 enrolments to matric performance in 2013 (expressed as a percentage) in public schools



Source: Own calculations from Matriculation and EMIS data

Performance data (NSC, ANA, SACMEQ, PIRLS, TIMSS and Western Cape Systemic Tests) clearly indicates the bimodal nature of the South African education system. **Figure 6**, a distribution of the NSC examination results of 2013 by school quintile, shows that one part of the system represents the well performing schools while the second part of the system represents the distribution of under-performing schools. This is adequately reported in the literature (see Gustafsson, 2005; Fleisch, 2008). The value of data integration is emphasised in **Figure 6**. When combining the NSC data with the EMIS data it is possible to split the matric performance by school quintiles (available in the EMIS data).

Figure 6: Distribution of 2013 NSC examination by average school aggregate mark



Own calculations from NSC data and EMIS data

Data integration is also necessary to create a longitudinal dataset. Longitudinal data coverage is a key requirement for tracking individual learners through the education system. This involves creating a dataset that contains information of the same learners or schools from year to year. With this longitudinal data coverage one can determine exactly how many learners of a specific cohort dropped out of the system, how many progressed through the system without any repetition and how many are still in the system with one or more repetitions.

A unique learner identifier

In creating a longitudinal data system it is necessary to link the different datasets that have been collected for individual learners or schools for each year through the use of a common field across these datasets. To be able to do this, unique identification codes must be assigned to every learner or every school. It is important that this identifier is consistent and accurate over time. A unique identifier is a single, non-duplicated number that is assigned to, and remains with, a learner throughout his or her education career irrespective of whether the learner changes schools. The unique identifier (in the case of the ANA the learner's ANA number, or for schools the school's emis number) is used to link the datasets of the different years to identify all learners or schools.

It is then possible through a learner identifier to follow the progress of each learner over time, and across schools or districts within the country. The lack of commonly used unique identifiers that allow linkage across data systems contributes to the unavailability of integrated information systems or longitudinal datasets. Linking matric data for several years by using the school's emis number, for instance, we can observe outcomes for the same school at different times.

Learner unit record data systems in South Africa

This section examines the prospects of implementing a learner unit record system nationally to replace the EMIS system that is currently in place based on aggregate/summary institution-level data.

Aggregate data

Aggregate data refers to data collected at the school level typically through surveys such as the SNAP and Annual School Survey. Aggregated or summary data collection in South Africa provides information on issues at institutional level, such as enrolment by grade and age, enrolment by gender and language. Since the establishment of a single national Education Management Information System (EMIS) in 1995, summary or aggregated data was collected at institutional-level through survey questionnaires. The completion of surveys has become a specialized, separate and additional responsibility for schools that might become cumbersome and result in a reluctance to comply or cooperate. However, the collection of summary data through surveys is a well-established practice in South Africa that currently produces data quality of an acceptable standard.

Systems producing learner unit-records

Learner unit record data refers to the data collected for each learner through a school administration and management system. The ideal is that data collection processes and activities should be institutionalized to such an extent that they become part of the day-to-day running of the institution and not an add-on, separate or additional process. In other words, all data should be generated through the operations (day-to-day running) of the institution based on the core business (learner and learning) of the school. In such instances the national or provincial EMIS data then should come mainly from operational systems at institutional level where the main data collection mode should be at learner unit-level and not at an aggregated (summary) level.

The DBE is already moving towards data collection at the learner unit-level through an electronic school administration and management system. The South African School Administration and Management System (SA-SAMS) is a comprehensive school administration and management software solution developed, maintained and enhanced free of charge by the national DBE.

Collection of data through a School Administration and Management System

SA-SAMS is an off-line (desktop) school administration and management system that has been widely distributed and piloted in all provinces. The rollout plan of SA-SAMS and the training of staff are managed by each province. The software caters for a number of modules to manage school administration according to policy prescripts, such as General School Information, Learner and Parent Information, Human Resource Information, Learner Listing, Governance, Financial, Curriculum, Timetabling, Physical Resources, Attendance – Learner and Educator, etc. SA-SAMS can be phased-in according to specific modules and expanded as the needs of the institution increase. This phasing-in of modules over time in a specific institution is an approach that is highly recommended. The Free State has implemented the system and all their data collection processes (SNAP and ASS) take place via this method. The Eastern Cape is another Province where this method is successfully implemented. The availability of such learner-unit record data makes it possible to analyse the flow through patterns in terms of repetition, dropout and progression, which is not possible with aggregated datasets.

Learner Unit Record Tracking System (LURITS)

LURITS is a standardized system that assigns a unique national identifier to learners in South Africa in public schools. LURITS is the tracking module of the school administration and

management system that assigns each learner a unique number. The system enables the Department of Basic Education (DBE) to store the unit record level data of learners and educators in a central national database and enables the Department to register learners and to track the movement of learners from school to school throughout their school career.

The main purpose of this system is to track the movement of learners from school to school and to provide accurate enrolment numbers and learner profile data for planning and strategic decision-making purposes. It should also make it possible to track learner performance over time in the ANA.

Centralized Education Management Information System (CEMIS)

Western Cape has a web-enabled system called Centralized Education Management Information System (CEMIS). CEMIS is not a school management and administration system in the true sense of the word. It is mainly used as a learner registration and tracking system. The system is centrally developed and managed by the provincial department. The main functionality of the system is to register learners and to track and monitor individual learners in the province: registration of learners, transfers of learners between schools, examination passes, etc. Because it is centrally managed the WCED has access to the information of each learner in all schools. The SNAP and Annual School Survey (ASS) are populated through this system at provincial level and sent to the national department once a year. It seems that the system works well, based on the high quality of the annual data that is submitted. The CEMIS dataset now makes it possible for the WCED through longitudinal cohort analysis to determine flow through patterns: progression of learners through the education system without repetition, repetition of learners who nevertheless remain in the system and dropping out of learners. Linking the CEMIS data with the data of Systemic Tests will further make it possible not only to follow learners through the education system, but also to track the performance of individual learners over time.

Issues Related to the Development of a Unit Record System

With the availability of unit-level learner records key questions can be answered such as: “What is the profile of the learners who dropped out of the system, or what is the profile of the learners who progressed without any repetition?” If we have a panel dataset, e.g. with the introduction of the Annual National Assessments, a longitudinal dataset can be created where there is outcomes data for the same learner at multiple points in time.

The suggestion further is with regard to learner unit school administration and management system to build the system incrementally. There is no doubt that a unit record system is necessary and should be implemented in all schools in the country. The recommendation is that a school administration and management system, such as SA-SAMS should be phased in according to specific modules based on the capacity of the school and that the survey method still be continued until the SA-SAMS is well-established in all institutions in a province. An important question at this stage is: Should a unit record system be developed and data stored at unit record level at the national department? There are specific technical and capacity challenges relating to this question. For example, taken into account the number of learners in the country one database table with millions of learner records could easily exist. The national system should therefore not be a replication of the operational unit record system of the institutional and provincial level. The only learner-unit level records at national level should be records in the national learner identifier system for learner registration and learner tracking purposes.

Conclusion and recommendations

This paper provided an overview of types of data that should be considered for use in decision making and school improvement strategies, as well as which are available and ready for analysis. **Table 7** provides a summary of the main large-scale datasets and where to access them.

Table 7: *Main large-scale education datasets*

DATA SYSTEM	LEVEL	ACCESS DETAIL
MASTER LIST OF SCHOOLS	School	http://www.education.gov.za/EMIS/EMISDownloads/tabid/466/Default.aspx (accessed June 2015)
SNAP SURVEY	School	http://www.education.gov.za/EMIS/tabid/57/Default.aspx (accessed June 2015) https://www.datafirst.uct.ac.za/dataportal/index.php/catalog (accessed October 2015)
ANNUAL SCHOOL SURVEY	School	http://www.education.gov.za/EMIS/tabid/57/Default.aspx (accessed June 2015)
MATRIC EXAMINATION	Learner	http://www.education.gov.za/dobeinternet/Home/tabid/36/Default.aspx (accessed June 2015)
ANNUAL NATIONAL ASSESSMENT	Learner	http://www.education.gov.za/Examinations/AnnualNationalAssessmentsANA/tabid/569/Default.aspx (accessed June 2015)
SACMEQ	Learner	http://www.sacmeq.org/?q=about-us (accessed June 2015)
PIRLS	Learner	http://timss.bc.edu/ (accessed June 2015)
TIMSS	Learner	http://timss.bc.edu/ (accessed June 2015)
TIMSS SA	Learner	http://www.timss-sa.org.za/ (accessed June 2015)

Recent years have seen a considerable expansion in the availability and in some cases also in the quality of data available for policy and decision making, and also for research. This paper contains not only an overview of the data, but also some suggestions regarding data improvement and accessibility of data in a form that it could be better utilized for both policy and analysis. Based on the foregoing and the information presented earlier in this paper, the following recommendations are made:

- Include the enrolment of various years as additional fields in the master list of schools,
- Improve the data quality of the master list by updating all the missing fields, specifically in key fields such as quintile, examno, and others,
- Increase the accessibility of the SNAP and ASS data by making it available in a user friendly file format, such as a database, spreadsheet or even a comma delimited text file that could be downloaded,
- Make the current ANA and NSC data, already accessible on the DBE's website in a pdf format, available in a more user friendly data file format, such as a database, spreadsheet or even a comma delimited text file that could be downloaded. Follow the best practice example of TIMSS and PIRLS about the availability of the international database by year with the relevant documentation, such as the User Guide, codebooks and the data files.

It is hoped that the information provided in this paper about the quality, relevance and accessibility of education data in South Africa will aid policy researchers and policy makers in their selection and use of education data.

References

Carrizo, L., Sauvageot, C. & Bella, N. 2003. Information tools for the preparation and monitoring of education plans. Paris, France: UNESCO. Retrieved from <http://unesdoc.unesco.org/images/0013/001323/132306e.pdf> (accessed October 2015).

- Cassidy T. 2005. Education Management Information System (EMIS) Development in Latin America and the Caribbean: Lessons and Challenges. September 2005.
- Crouch, L, & Mabogoane, T. 2001. No magic bullets, just tracer bullets: The role of learning resources, social advantage, and education management in improving the performance of South African schools. *Social Dynamics*, 27(1), 60-78.
- DataFirst.2015.<https://www.datafirst.uct.ac.za/dataportal/index.php/catalog> (accessed October 2015).
- Fleisch, B. 2008. Primary Education in Crisis: Why South African schoolchildren underachieve in reading and mathematics. Cape Town. : Juta & Co.
- Gustafsson, M. 2005. The Relationships Between Schooling Inputs and Outputs in South Africa: Methodologies and Policy Recommendations Based on the 2000 SACMEQ Dataset. Paper to SACMEQ Invitational Conference, 28–30 September 2005. International Institute for Educational Planning, Unesco, Paris.
- Gustafsson, M. 2012. 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 (2015) 63–72.
- Hau H. & Herstein J. 2003. Education Management Information Systems (EMIS): Integrated Data and Information Systems and Their Implications in Education Management. Paper presented at the Annual Conference of Comparative and International Education Society, New Orleans, March 2003.
- Moloi, M. & Chetty, M. 2011. The SACMEQ III Project in South Africa: A Study of the Conditions of Schooling and the Quality of Education: South Africa: Country Report. Department of Basic Education Ordinary Schools, 2014. Retrieved from <http://www.education.gov.za/EMIS/tabid/57/Default.aspx> (accessed June 2015)
- PIRLS. 2015. Retrieved from <http://timss.bc.edu/> (accessed June 2015)
- Pre-PIRLS, 2015. Retrieved from <http://timss.bc.edu/pirls2011/prepirls.html> (accessed June 2015)
- Reddy, V., Van der Berg S., Janse van Rensburg D., Taylor S. 2012. Educational outcomes: Pathways and performance in South African high schools. *S Afr J Sci.* 2012; 108(3/4), Art. #620, 8 pages. [http:// dx.doi.org/10.4102/sajs.v108i3/4.620](http://dx.doi.org/10.4102/sajs.v108i3/4.620)
- Reddy, V., Zuze, T., Visser, M., Winnaar L., Juan, A., Prinsloo, C., Arends, F., & Rogers, S. 2015. Beyond Benchmarks. What 20 years of TIMSS data tell us about South African education. Retrieved from http://www.timsssa.org.za/?page_id=41 (accessed June 2015)
- SACMEQ. 2015. The Southern and Eastern Africa Consortium for Monitoring Educational Quality. Retrieved from <http://www.sacmeq.org/?q=about-us> (accessed June 2015)
- South Africa, DBE. 2004. National Education Information Policy. Pretoria: Department of Education, August 2004
- South Africa, DBE. 2006. South African Schools Act, 1996 (Act No 84 of 1996), Amended National Norms And Standards For School Funding. Pretoria: Department of Education, 31 August 2006
- South Africa, DBE. 2012. Annual National Assessment 2012: Report on the ANA of 2012. Pretoria: Department of Basic Education, December 2012. Retrieved from <http://www.education.gov.za/Examinations/AnnualNationalAssessmentsANA/tabid/569/Default.aspx> (accessed June 2015)
- South Africa, DBE. 2013. Annual National Assessment 2013: Report on the ANA of 2013. Pretoria: Department of Basic Education, December 2013. Retrieved from

- <http://www.education.gov.za/Examinations/AnnualNationalAssessmentsANA/tabid/569/Default.aspx> (accessed June 2015)
- South Africa, DBE. 2014a. Annual Snap Survey for Ordinary Schools 2014. Pretoria: Department of Basic Education: Retrieved from <http://www.education.gov.za/EMIS/EMISDownloads/tabid/466/Default.aspx> (accessed June 2015)
- South Africa, DBE. 2014b. National Guidelines on: Completing the SNAP Survey for Ordinary Schools 2014. Pretoria: Department of Basic Education, December 2014. Retrieved from <http://www.education.gov.za/LinkClick.aspx?fileticket=u84UbV7FyJM%3d&tabid=57&mid=1751> (accessed June 2015)
- South Africa, DBE. 2014c. Annual National Assessment 2014: Report on the ANA of 2014. Pretoria: Department of Basic Education, December 2014. Retrieved from <http://www.education.gov.za/Examinations/AnnualNationalAssessmentsANA/tabid/569/Default.aspx> (accessed June 2015)
- South Africa, DBE. 2015a. School Masterlist Data. Pretoria: Department of Basic Education: Retrieved from <http://www.education.gov.za/EMIS/EMISDownloads/tabid/466/Default.aspx> (accessed June 2015)
- South Africa, DBE. 2015b. Education Statistics In South Africa Report. Pretoria: Department of Basic Education: Retrieved from <http://www.education.gov.za/EMIS/StatisticalPublications/tabid/462/Default.aspx>
- South Africa, DBE. 2015c. School Realities Report. Pretoria: Department of Basic Education: Retrieved from <http://www.education.gov.za/EMIS/StatisticalPublications/tabid/462/Default.aspx> (accessed June 2015)
- South Africa, DBE. 2015d. Dictionary of Education Concepts. Pretoria: Department of Basic Education: Retrieved from <http://www.education.gov.za/EMIS/StatisticalPublications/tabid/462/Default.aspx> (accessed June 2015)
- South Africa, DBE. 2015e. Annual Survey: Ordinary Schools. Pretoria: Department of Basic Education: Retrieved from <http://www.education.gov.za/EMIS/tabid/57/Default.aspx>
- South Africa, DBE. 2015e. NSC Examinations 2014: School Performance Report. Pretoria: Department of Basic Education. Retrieved from <http://www.education.gov.za/LinkClick.aspx?fileticket=4JVOD%2f6ncds%3d&tabid=358&mid=1325> (accessed June 2015)
- South Africa, DBE. 2015f. NSC Examinations 2014: School Subject Report. Pretoria: Department of Basic Education. Retrieved from <http://www.education.gov.za/Examinations/ANA2014testsandmemos/tabid/569/Default.aspx> (accessed June 2015)
- South Africa, DBE. 2015g. Annual National Assessments. Pretoria: Department of Basic Education. Retrieved from <http://www.education.gov.za/Examinations/AnnualNationalAssessmentsANA/tabid/569/Default.aspx> (accessed June 2015)
- South African Schools Act. 1996. South African Government, Pretoria: Government Printers
- Shepherd, D. 2011. Constraints to school effectiveness: what prevents poor schools from delivering results? *Stellenbosch Working Paper Series No. WP05/2011*
- Shepherd, D. 2013. A question of efficiency: decomposing South African reading test scores using PIRLS 2006. *Stellenbosch Working Paper Series No. WP20/2013*

- Shepherd, D. 2015. Learn to teach, teach to learn: A within-pupil across-subject approach to estimating the impact of teacher subject knowledge on South African grade 6 performance. *Stellenbosch Working Paper Series No.01/2015*
- Simkins, C. 2010. Interpreting the mathematics, science and English performance of candidates in the November 2008 National Senior Certificate examination. Johannesburg: Centre for Development and Enterprise.
- Spaull, N. & Taylor, S. 2012. "Effective enrolment" - Creating a composite measure of educational access and educational quality to accurately describe education system performance in sub-Saharan Africa. *Stellenbosch Economic Working Papers 21/2012*
- Spaull, N. 2013a. South Africa's Education Crisis: The quality of education in South Africa 1994-2011. Report Commissioned by Centre for Development and Enterprise, October 2013
- Spaull, N. 2013b. The effects of rapidly expanding primary school access on effective learning: The case of Southern and Eastern Africa since 2000. *Stellenbosch Working Paper Series No. WP01*.
- Spaull, N. 2014. What do we know about primary teachers' mathematical content knowledge in South Africa? An analysis of SACMEQ 2007. *Stellenbosch Working Paper Series No. WP13/2014*
- Spaull, N. 2012. Poverty & privilege: Primary school inequality in South Africa. *International Journal of Educational Development 2012*; (2012): 1-12.
- Taylor, S & Yu, D. 2009. The importance of socio-economic status in determining educational achievement in South Africa. *Stellenbosch Working Paper Series No. WP01/2009*
- Taylor, S. 2014. Reviewing the Language Compensation Policy in the National Senior Certificate. *Perspectives in Education 2014*; 32(1): 124-141.
- TIMMS SA. 2015. Retrieved from <http://www.timss-sa.org.za/> (accessed June 2015)
- TIMSS. 2015. Retrieved from <http://timss.bc.edu/> (accessed June 2015)
- Van der Berg, S & Shepherd, D. 2010. Signalling performance: Continuous assessment and matriculation examination marks in South African schools. *Stellenbosch Working Paper Series No. WP28/2010*
- Van der Berg, S. & Louw, M. 2007. Lessons learnt from SACMEQII: South African student performance in regional context. *Stellenbosch Working Paper Series No. WP16/2007*
- Van der Berg, S. 2006. How effective are poor schools? Poverty and educational outcomes in South Africa. *Stellenbosch Economic Working Papers: 06/06*.
- Van der Berg, S. 2007. Apartheid's Enduring Legacy: Inequalities in Education. *Journal of African Economies*; 16 (5), 849-880.
- Van der Berg, S. 2008. How effective are poor schools? Poverty and educational outcomes in South Africa. *Studies in Educational Evaluation*; 34(3) September: 145-154