

An overview of key data sets in education in South Africa

Chris van Wyk, University of Stellenbosch.*

Abstract

This paper provides an overview of the various data sets 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 data sets and the basic data elements contained in each of these data sets. When discussing each of the existing education data sets, 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

*Email address: chrisvanwyk@sun.ac.za.

South African Journal of Childhood Education | 2015 5(2): 146-170 | ISSN: 2223-7674 | © UJ

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 data sets generated by the annual school survey in the 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 entails the development of mechanisms for accountability and for monitoring and evaluation of the education system. The realization of these challenges was aided significantly with the establishment in 1995 of the EMIS unit, which is responsible for producing, managing and disseminating education data within the DBE and provincial education departments. 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 2005:27

The objective of an 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 data sets as well as the basic data elements contained in each of these data sets. Proper management, planning and policy-formulation are contingent on quality data. When discussing each of the existing education data sets, 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 data sets in South Africa

This section of the paper provides a description of the existing education data sets and the basic data elements contained in each of these data sets. The primary purpose is to provide a compilation of education data sets covering the broad field of South African education data.

These data sets include a selection of data from many different systems and draw specifically on the outcomes of surveys and operational activities carried out by the DBE.

The list of education data sets discussed in this paper is not exhaustive, but for the purpose of this paper it includes data generated by EMIS, learning outcomes data – including international test data such as the Progress in International Reading and Literacy Study (PIRLS); the Trends in International Mathematics and Science Study (TIMSS); and the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) – and key performance data in South Africa, such as the Annual National Assessment (ANA) and National Senior Certificate (NSC).

In the paragraphs that follow we will look at the elements, quality and integration of the data in these sets, as well as the policy implications of these different data sets.

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’ and referred to as ‘Natemis’ in the master list of schools (see Table 1). 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 (Source: RSA DBE 2015a)

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 National 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	This is an indication of the socio-economic status of the school. Schools are ranked according to poverty of the school community.

Figure 1 indicates the distribution of schools according to quintiles¹ 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.

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 (RSA DBE 2006).

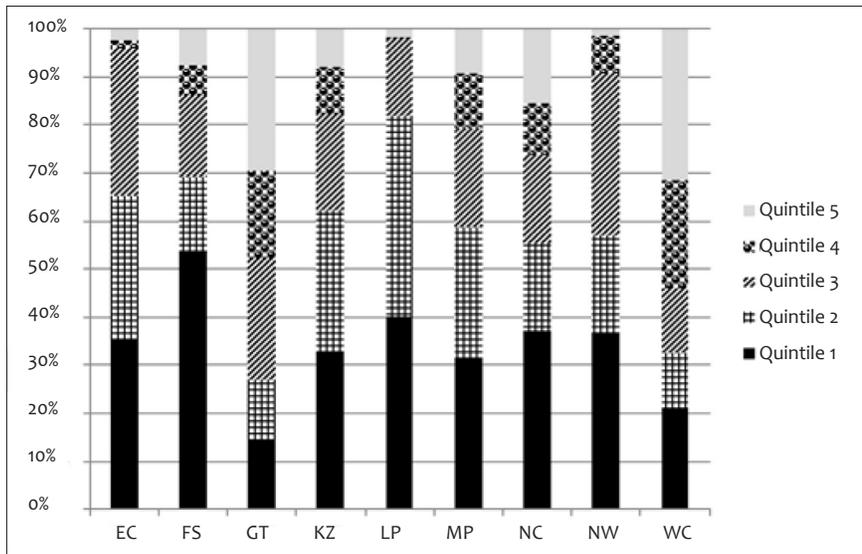


Figure 1: Distribution of schools by province and quintile (Source: Author’s calculation, based on Master list of of schools)

The master list of schools is a data table that is available through DBE and is accessible through their website (RSA 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 DBE website, 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 with previous master lists. The master list of schools should also be available earlier in the year. It is a useful data set for education planners and researchers and is even widely used in the private sector by 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 the 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 data sets 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 tenth school day of every year. The SNAP survey form issued by the DBE (RSA 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 (sections 59[1] and [2]) (RSA 1996) stipulates that

[e]very 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.

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 Report* for each year from 2008 to 2013 (RSA DBE 2015b) are publications based on the data collected via the SNAP survey. Table 2 shows the important data elements included in the SNAP survey.

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 to 2013. Some of the flow-through trends that can be observed in the South African education system are graphically illustrated in 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.

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 a spreadsheet at school level. The *School Realities* (RSA DBE 2015c) publication, based on the SNAP data, is a valuable data reference, but is only available on the DBE website as a PDF file, which is difficult to manipulate. If, for example, the DBE made enrolment data per school available in a more accessible file format on their website, it could help promote greater use of data in planning and decision-making. Here the DataFirst approach could be followed. DataFirst is a data service based at the University of Cape Town that gives researchers online access to African data and online help with data usage. It provides researchers access to South African enrolment data by publishing the SNAP data in a manipulable data file format on its data portal (DataFirst 2015).

Table 2: Data elements included in the SNAP survey (Source: RSA DBE 2014a)

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. Administration Staff		X		X	X		X	X	X
Support staff		X		X	X		X	X	X
Hostel staff		X		X	X		X	X	X

Table 3: Enrolment for public ordinary schools based on SNAP data (Source: Compiled from SNAP data, 2002-2013)

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

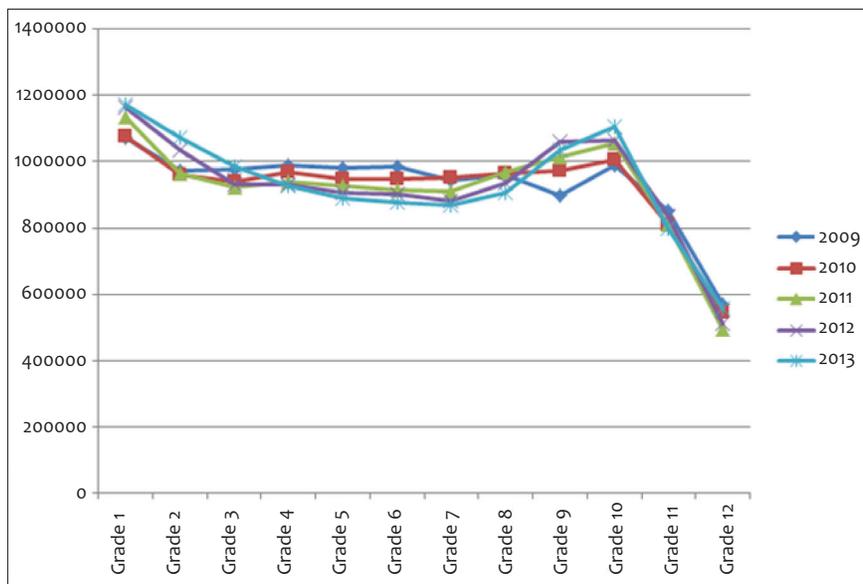


Figure 2: Enrolment in public schools by grade and year, 2009-2013 (Source: Author’s calculation, based on SNAP data, 2009-2013)

Annual School Survey (ASS)

The data available through the Annual School Survey (ASS) 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 one hundred learners that start Grade 1, fifty will drop out before Grade 12 (most of which happens in Grades 10 and 11); forty will pass the NSC examination; and twelve will qualify for university (Spaull 2013a). The availability and quality of the data from the ASS make it possible to answer key questions, for example, where in the system is the highest dropout and repetition?

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 ASS is a comprehensive survey of all public and independent schools in South Africa. It was designed to provide comparable information on the public and private sectors, as well as trend data over time. The ASS is completed by all schools in the country on a specific day, usually in March. The *Annual School Survey Form* (RSA DBE 2015e) and the *Dictionary of Education Concepts and Terms 2010* (RSA DBE 2015d) provide a comprehensive insight of the data that is available. Table 4 shows some of the key data elements included in the ASS.

Table 4: Data elements included in the Annual School Survey (Source: RSA DBE 2015e)

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		

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 to their age (Figure 3), but due to repetition (and perhaps also dropout) there is a much wider age range by Grade 10, indicating an average (repeater) problem in the higher grades.

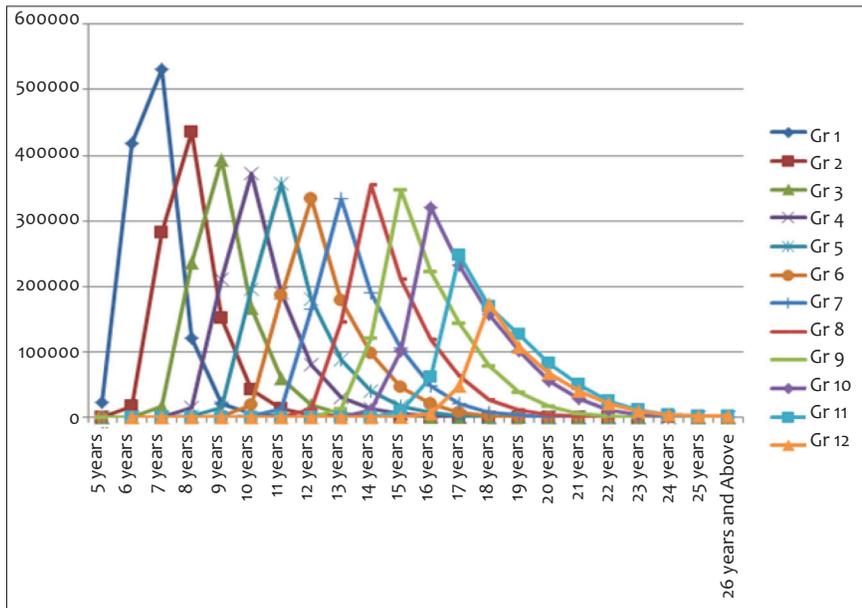


Figure 3: Enrolment by grade and age in 2011 (Source: ASS data for 2011)

Performance data

School outcomes are often used to determine the quality and efficiency of the education system. Standardised test scores are regarded as the best indicators to measure school performance. In the past, Grade 12 results were the only available measure in South Africa, but since 2011 the Annual National Assessment (ANA), a test in literacy and numeracy for Grades 1 to 6 and Grade 9, is also available.

Comprehensive learning outcomes data now exists, including international test data such as PIRLS, TIMSS and SACMEQ, and key performance data for South Africa, such as the ANA and NSC. In addition, in the Western Cape Systemic Tests in language and mathematics are conducted each year in Grades 3, 6 and 9. These examinations and tests are important diagnostic assessment tools for schools, educators and management, as they provide specific information on performance.

South Africa participates in three main international tests of educational achievement, namely TIMSS, PIRLS and SACMEQ. These tests show how the performance of South African learners has changed over time relative to earlier

Table 5: Performance data by year

Performance System	National	International	Provincial	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
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	X
Annual National Assessment (Grade 1 to 6) and Grade 9	X																			X	X	X	X
Systemic Test 1			X								X		X		X		X		X	X	X	X	X
Grade 3																							
Systemic Test			X											X		X		X	X	X	X	X	X
Grade 6																							
Systemic Test																							
Grade 9			X																	X	X	X	X
SACMEQ (Grade 6)		X							X							X							
TIMSS (Gr 9)		X		X				X															
PIRLS (Grade 5)		X													X								
PRE-PIRLS (Grade 4)	X																						

cohorts, and relative to other countries participating in these studies (Spaull 2013b). Table 5 indicates the availability of performance data in South Africa.

The National Senior Certificate (NSC)

The Senior Certificate (commonly referred to as ‘matric’) exam results are available per subject for every learner who wrote the examination since 1994. There is a growing body of research on student achievement and school performance in South Africa, which is of particular importance for this paper (Crouch & Mabogoane 2001; Taylor 2014; Simkins 2010; Van der Berg & Shepherd 2010; Van der Berg 2007).

The *School Performance Report* (RSA DBE 2015f) is a publication of the performance of individual schools in the NSC exam for the years 2012 to 2014. The report includes data elements such as the official school number, quintile, the number of students who wrote the exam, and the number who passed for individual schools. The *School Subject Report* (RSA DBE 2015g) is another useful publication and an important data source on NSC results. The report includes key subjects, with data elements such as the year of examination, the quintile of the school, the number of students who wrote the exam, and the number who passed for specific key subjects. Such accessibility and availability of data on school and subject performance is extremely useful, and could contribute to growing use of the NSC results, as it allows analysts to determine patterns and trends in school performance. Table 6 and Figure 4 show the general trend for the NSC exam. With the exception of 2011 – where fewer learners passed matric than in the previous year, partly due to fewer of them taking the exam – the number of individuals passing matric has generally been on the increase in recent years. In fact, the proportion of Grade 12’s passing the national matric examination 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 (Source: Compiled from NSC results, 2009-2014)

Year	Wrote	Passed	Pass rate
2009	562750	339144	60.3%
2010	537543	364513	67.8%
2011	496090	348117	70.2%
2012	511152	377829	73.9%
2013	563462	440877	78.2%
2014	532860	403874	75.8%

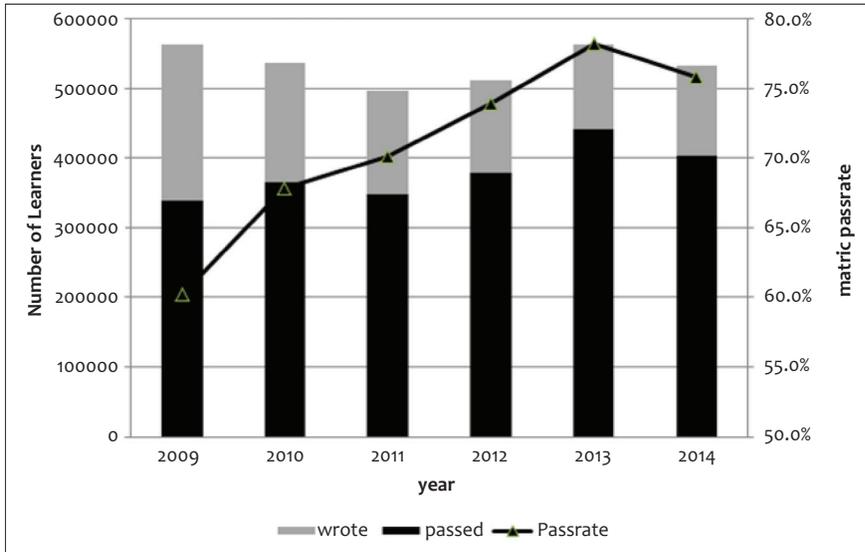


Figure 4: NSC results (2009-2013) in all schools (Source: Author's calculations from matriculation data)

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

Annual National Assessment (ANA)

The Annual National Assessments (ANAs) (RSA DBE 2015h) are standardised national tests in languages and mathematics in the Senior Phase (Grades 7-9) and Intermediate Phase (Grades 4-6), and in literacy and numeracy for the Foundation Phase (Grades 1-3). The question papers and marking memoranda are supplied by the DBE, but schools conduct the tests and manage the marking and internal moderation themselves – 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 NSC. As a result, it was not really possible to determine the performance of primary schools, hold schools accountable for learning performance, or provide support to schools where it was needed most.

The DBE has released reports for the ANA 2012 (RSA DBE 2012); ANA 2013 (RSA DBE 2013); and ANA 2014 (RSA DBE 2014c) outlining their rationale, methodology, curriculum interventions and results.

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

SACMEQ is an international non-profit developmental organization of fifteen 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 2015). SACMEQ's rich data sets provide new possibilities for investigating relationships between educational outcomes, socio-economic status (SES), learner and teacher characteristics, school resources and school processes (Van der Berg 2006), and there is a growing body of literature using SACMEQ data in empirical studies to evaluate learner performance in South Africa (Gustafsson 2005; Moloi & Chetty 2010; Shepherd 2015; Spaul & Taylor 2012; Spaul 2014, 2013b, 2012; Van der Berg & Louw 2007; Van der Berg 2008).

Progress in International Reading and Literacy Study (PIRLS)

PIRLS (TIMSS & PIRLS 2015a) is an international comparative study aimed at benchmarking literacy/reading levels across countries. It tests the reading literacy of fourth graders from forty-nine countries. This worldwide assessment and research project is designed to measure trends in children's reading literacy achievement and to collect information about policy and practices related to reading and reading instruction. The PIRLS results can be used for issues related to the language of learning and teaching (LOLT) and to track reading performance over time. It is a useful tool for international comparisons. Various local researchers have analysed the PIRLS data sets and provided valuable insight into the state of education in the country (Shepherd 2011, 2013; Taylor & Yu 2009).

The preProgress in International Reading Literacy Study (prePIRLS) is a stepping stone to participation in PIRLS and provides a way of assessing reading at the end of the primary school cycle for a range of developing countries. The prePIRLS reflects the same conception of reading as PIRLS, but is easier. 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 (TIMSS & PIRLS 2015b). There is comprehensive supporting documentation on PIRLS, such as the *User Guide for the International Database*, code books and data files, all of which are available on the website (TIMSS & PIRLS 2015a). South Africa participated in PIRLS up to 2014, but then changed to prePIRLS, on the grounds that this measure provides more detailed information on the strengths and weaknesses of South African learners.

Trends in International Mathematics and Science Study (TIMSS)

TIMSS tests the mathematics and science achievement of Grade 8 and Grade 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 participated in the TIMSS assessments in 1995; 1999 (Grade 8 only); 2002 (Grades 8 and 9); and 2011 (Grade 9 only). The next round of TIMSS will be administered in South Africa in 2015.

The new TIMSS numeracy assessment was designed to be administered at the Grade 4-, 5- or 6-level to measure children’s numeracy learning outcomes. TIMSS numeracy tests were conducted for the first time in South Africa in 2014 at the Grade 5-level (TIMSS SA 2015), but the results have not yet been released.

There is comprehensive supporting documentation on TIMSS, such as the *User Guide for the International Database*, the code books and the data files, which are available on the website (TIMSS & PIRLS 2015a). The contribution of TIMSS data to understanding education in South Africa has been widely researched (Reddy, Van der Berg, Janse van Rensburg & Taylor 2012; Reddy, Zuze, Visser et al 2015).

The importance of performance data

The availability of performance data makes it possible, among other things, to determine if learners are learning the content at the appropriate levels. Performance data also forms the basis for accountability. Furthermore, it assists in comparing learners on a national and international level 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

[t]he education authorities have a paucity of information for decision-making. Identifying under-performing schools in order to take remedial action requires understanding school performance.

Data integration

The EMIS data sets (such as the ASS) or performance data sets (such as the NSC examination results) on their own are not comprehensive enough for sound education planning, monitoring and policy formulation. There is also no single information system in the 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, NSC, 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 data that has already been collected and is available in various systems. All the data sets mentioned will remain disparate databases – fragmented islands of data that exist in isolation from other data – if they are not linked. This limits their potential use for informing decision-making.

Figure 5 is a typical example of data from the NSC combined with 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 rather a reflection of high grade repetition in Grades 10 and 11.

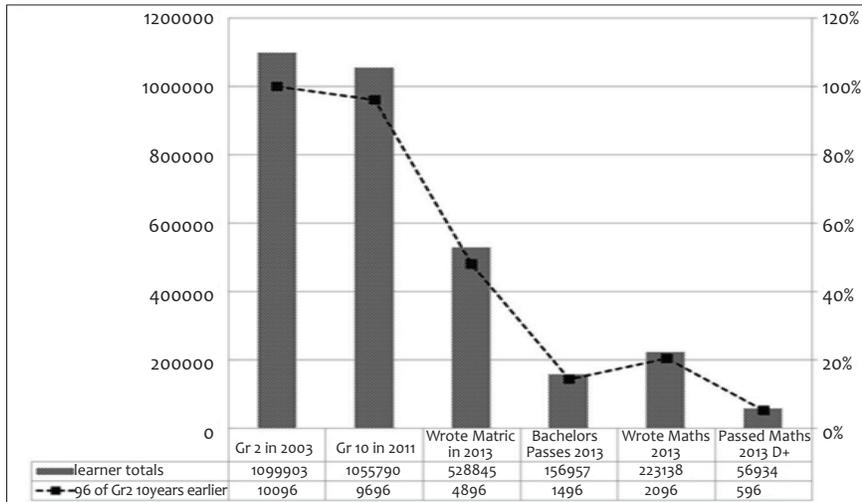


Figure 5: The ratio of Grade 2 enrolments to matric performance in 2013 (expressed as a percentage) in public schools (Source: Author’s calculations, based on NSC and EMIS data)

Performance data (NSC, ANA, SACMEQ, PIRLS, TIMSS and the Western Cape Systemic Tests) clearly show the bimodal nature of the South African education system. Figure 6, a distribution of the NSC examination results of 2013 by school quintile, illustrates this phenomenon at the end of Grade 12. 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.

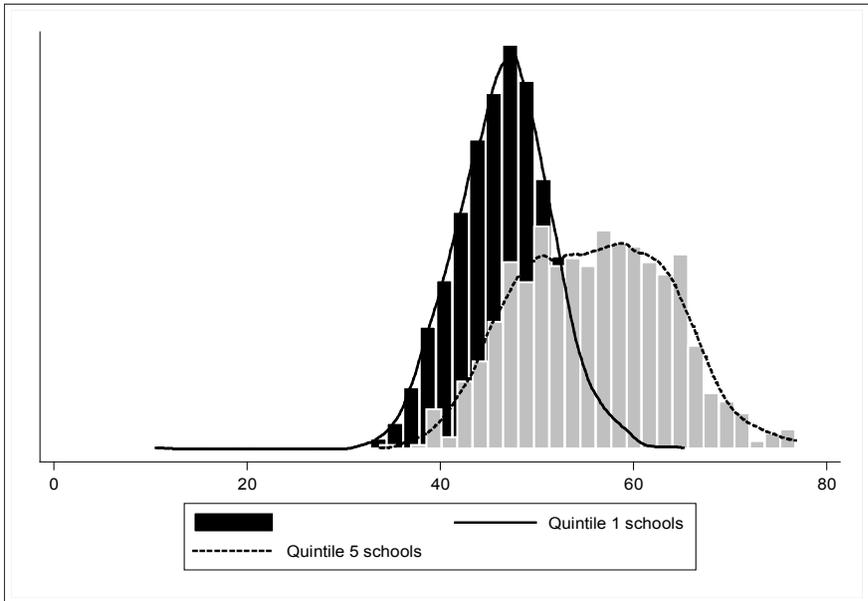


Figure 6: Distribution of 2013 NSC examination by average school aggregate mark (Source: Author's calculations, based on NSC and EMIS data)

Data integration is also necessary to create a longitudinal data set. Longitudinal data coverage is a key requirement for tracking individual learners through the education system. This involves creating a data set that contains information of the same learners or schools from year to year. With this longitudinal data coverage it is possible to determine exactly how many learners from 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

When creating a longitudinal data system it is necessary to link the different data sets that have been collected for individual learners or schools for each year through the use of a common field across these data sets. 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. It will then be possible to follow the progress of each learner over time through a learner identifier, 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 data sets.

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 or summary institution-level data.

Aggregate data

Aggregate data refers to data collected at the school level, typically through surveys such as the SNAP or ASS. Aggregate or summary data collection in South Africa provides information on issues at the institutional level, such as enrolment by grade and age or by gender and language. Since the establishment of a single national EMIS in 1995, summary or aggregate data have been collected at institutional level by means of 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 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, a 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 (learners and learning) of the school. In such instances the national or provincial EMIS data 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 aggregate (summary) level.

Collection of data through a school administration and management system

The DBE is moving towards data collection at the learner unit level by means of 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 DBE.

SA-SAMS is an offline (desktop) system that has been widely distributed and piloted in all provinces. The roll-out plan of SA-SAMS and the training of staff are managed by each province individually. 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 accounting, curriculum, timetabling, physical resources, learner and educator attendance, etc. SA-SAMS can be phased in using selected 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 data collection processes (SNAP and ASS) take place via this method. The Eastern Cape is another province where this method has been 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 data sets.

Learner Unit Record Tracking System (LURITS)

LURITS is a standardized system that assigns a unique national identifier to learners in public schools in South Africa. It is the tracking module of the school administration and management system that assigns each learner a unique number. The system allows the 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 careers.

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.

Centralized Education Management Information System (CEMIS)

The Western Cape has a web-enabled system called the Centralized Education Management Information System (CEMIS). CEMIS is not a school management and administration system in the true sense of the word, and is mainly used as a learner registration and tracking system. The system is centrally developed and managed by the provincial department. The main function 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. As it is centrally managed, the Western Cape Education Department (WCED) has access to the information of each learner in all schools. The SNAP and ASS are populated through this system at provincial level and sent to the national department once a year. Based on the high quality of the annual data that is submitted it seems that the system works well. The CEMIS data set makes it possible for the WCED, through longitudinal cohort analysis, to determine flow-through patterns: the progression of learners through the education system, the repetition of learners who remain in the system, and the dropping out of learners. Linking the CEMIS data with that of the 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.

Conclusion and recommendations

This paper provides an overview of types of data on the South African school system which are available for use in policy making, instructional leadership, administration and research. Table 7 provides a summary of the main large-scale data sets and where to access them.

Table 7: Main large-scale education data sets

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 an overview of the data, and makes suggestions regarding improvements aimed at facilitating use of the available information. In particular, 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, exam number, and others;
- Increase the accessibility of the SNAP and ASS data by making it available for download in a user-friendly file format, such as a database, spreadsheet or even comma-delimited text file; and
- Make the current ANA and NSC data, now only accessible on the DBE website in portable document format (PDF), available for download in a more user-friendly data file format, such as a database, spreadsheet or comma-delimited text file. 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*, code books and 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: United Nations Educational, Scientific and Cultural Organization (UNESCO). Retrieved from <http://unesdoc.unesco.org/images/0013/001323/132306e.pdf> (accessed 14 October 2015).
- Cassidy T. 2005. *Education Management Information System (EMIS) development in Latin America and the Caribbean: Lessons and Challenges*. Work document. Study prepared for the 8th Regional Policy Dialogue Meeting, Inter-American Development Bank Education Network, Integration and Regional Programs Department and Sustainable Development Department. Washington, DC: Inter-American Development Bank.
- 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. *Data Portal*. Retrieved from <https://www.datafirst.uct.ac.za/dataportal/index.php/catalog> (accessed 14 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 data set*. Paper presented at the SACMEQ Invitational Conference, Paris, 28-30 September 2005. Paris: International Institute for Educational Planning, UNESCO.
- 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. 2010. *The SACMEQ III Project in South Africa: A Study of the conditions of schooling and the quality of education. South Africa: Country Report*. Pretoria: Department of Basic Education. Retrieved from <http://www.education.gov.za/EMIS/tabid/57/Default.aspx> (accessed 16 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. *South African Journal of Science*, 108(3/4):Art no 620. Retrieved from <http://dx.doi.org/10.4102/sajs.v108i3/4.620> (accessed 12 June 2015).

- 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 16 June 2015).
- RSA (Republic of South Africa). 1996. South African Schools Act, 1996 (Act No 84 of 1996). *Government Gazette*, Vol 377, No 17579, 15 November 1996. Pretoria: Government Printers.
- RSA DBE (Republic of South Africa. Department of Basic Education). 2004. *National Education Information Policy*. Pretoria: Department of Basic Education.
- RSA 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.
- RSA 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 18 June 2015).
- RSA 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 18 June 2015).
- RSA 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 18 June 2015).
- RSA 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 18 June 2015).
- RSA 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 18 June 2015).
- RSA DBE. 2015a. *School Masterlist Data*. Pretoria: Department of Basic Education. Retrieved from <http://www.education.gov.za/EMIS/EMISDownloads/tabid/466/Default.aspx> (accessed 18 June 2015).
- RSA 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> (accessed 18 June 2015).
- RSA DBE. 2015c. *School Realities Report*. Pretoria: Department of Basic Education. Retrieved from <http://www.education.gov.za/EMIS/StatisticalPublications/tabid/462/Default.aspx> (accessed 18 June 2015).
- RSA DBE. 2015d. *Dictionary of Education Concepts and Terms*. Pretoria: Department of Basic Education. Retrieved from <http://www.education.gov.za/EMIS/>

- StatisticalPublications/tabid/462/Default.aspx (accessed 18 June 2015).
- RSA DBE. 2015e. *Annual Survey: Ordinary Schools*. Pretoria: Department of Basic Education. Retrieved from <http://www.education.gov.za/EMIS/tabid/57/Default.aspx> (accessed 18 June 2015).
- RSA DBE. 2015f. *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 18 June 2015).
- RSA DBE. 2015g. *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 18 June 2015).
- RSA DBE. 2015h. *Annual National Assessments*. Pretoria: Department of Basic Education. Retrieved from <http://www.education.gov.za/Examinations/Annual-NationalAssessmentsANA/tabid/569/Default.aspx> (accessed 18 June 2015).
- SACMEQ (Southern and Eastern Africa Consortium for Monitoring Educational Quality). 2015. *About us*. Retrieved from <http://www.sacmeq.org/?q=about-us> (accessed 16 June 2015).
- Shepherd, D. 2011. *Constraints to school effectiveness: What prevents poor schools from delivering results?* Stellenbosch Working Paper Series No WPO5/2011. Stellenbosch: Department of Economics, Stellenbosch University.
- Shepherd, D. 2013. *A question of efficiency: decomposing South African reading test scores using PIRLS 2006*. Stellenbosch Working Paper Series No WP20/2013. Stellenbosch: Department of Economics, Stellenbosch University.
- Shepherd, D. 2015. *Learn to teach, teach to learn: A within-learner across-subject approach to estimating the impact of teacher subject knowledge on South African Grade 6 performance*. Stellenbosch Working Paper Series No WPO1/2015. Stellenbosch: Department of Economics, Stellenbosch University.
- 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 WP21/2012. Stellenbosch: Department of Economics, Stellenbosch University.
- Spaull, N. 2012. *Poverty & privilege: Primary school inequality in South Africa*. *International Journal of Educational Development* 2012:1-12.
- Spaull, N. 2013a. *South Africa's Education Crisis: The quality of education in South Africa, 1994-2011*. Report commissioned by the Centre for Development and Enterprise, October 2013. Johannesburg: Centre for Development and Enterprise.
- 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 WP01/2013. Stellenbosch: Department of Economics, Stellenbosch University.
- 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. Stellenbosch: Department of Economics, Stellenbosch University.
- 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. Stellenbosch: Department of Economics, Stellenbosch University.
- Taylor, S. 2014. Reviewing the Language Compensation Policy in the National Senior Certificate. *Perspectives in Education* 2014, 32(1):124-141.
- TIMMS SA. 2015. *Welcome to TIMMS SA*. Retrieved from <http://www.timss-sa.org.za> (accessed 16 June 2015).
- TIMSS & PIRLS. 2015a. *Latest Publications from TIMSS and PIRLS*. Retrieved from <http://timss.bc.edu> (accessed 16 June 2015).
- TIMSS & PIRLS. 2015b. *PIRLS 2011*. Retrieved from <http://timss.bc.edu/pirls2011/prepirls.html> (accessed 16 June 2015).
- 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. Stellenbosch: Department of Economics, Stellenbosch University.
- 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. Stellenbosch: Department of Economics, Stellenbosch University.
- Van der Berg, S. 2006. *How effective are poor schools? Poverty and educational outcomes in South Africa*. Stellenbosch Economic Working Paper Series No WP06/2006. Stellenbosch: Department of Economics, Stellenbosch University.
- 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):145-154.