RESEP

Research on Socio-Economic Policy

Department of Economics, University of Stellenbosch

Education datasets in South Africa

Chris van Wyk*

RESEP Policy Brief

The working paper on which this policy brief is based reviews education datasets available in South Africa, mostly from the Department of Basic Education (DBE). The working paper serves as an inventory for anyone wanting to know what data are available in South Africa, how the data may be accessed, what the quality is like and in what formats the data may be accessed.[1] This brief describes these datasets and the data elements of each, explains the need for data integration as a data management strategy, and offers recommendations for improving the quality and accessibility of datasets for both researchers and policymakers.

1. The importance of education data

Datasets generated by the education management information system (EMIS) annual schools survey are a valuable but often neglected data source for both research and policymaking. Since its establishment in 1994, the DBE has been tasked with providing good, relevant data as efficiently as possible to information users, particularly planners and decision makers.

Education datasets that inform, or could inform, policymaking for education in South Africa are the master list of schools, SNAP data, the annual schools survey, and performance data.

---

* RESEP, Department of Economics, Stellenbosch University

1   Van Wyk, C., 'An overview of education data in South Africa: An inventory approach', Stellenbosch Working Paper Series No. WP19/2015, www.ekon.sun.ac.za/wpapers/2015/wp192015

2. Master list of schools

This list is a record of all schools in the country. It uniquely identifies each school with a school identifier, generally called the 'EMIS number'. The list uses key fields such as the school's examination number, which is used to link school examination data to the administrative data. It also uses other basic data fields that could provide answers to questions about school socioeconomic status, learner enrolment, teachers and learner-teacher ratios.

Schools are divided into five categories (quintiles) based on the socioeconomic status of the community in which the school is situated. Quintile 1 schools are the poorest and quintile 5 schools the most affluent. Analyses comparing schools' performance often use school quintiles as control measures for socioeconomic status, to take into account the effect of, for example, poor infrastructure, shortage of materials and deprived home backgrounds on school performance.

The master list of schools is publicly accessible on the DBE website.[2] It is a useful dataset for education planners and researchers and is even widely used in the private sector by those who regularly deal with schools. It is also used to match school data across years and link it with other datasets such as examination data, using the unique school identifier.

3. SNAP data

The Annual SNAP survey contains data recorded on the 10th school day of the year. It is an important source of information for three purposes: the allocation of funds per learner based on the national norms and standards for school funding, the allocation of teachers to schools, and the annual publication of education data. Data from the SNAP are publicly available for all learners per school by grade for the years 1997 to 2014.[3] Some of the flow-through trends that can be observed in the South African education system are shown in Figure 1, based on the SNAP data for enrolment in public schools from 2009 to 2013. Although the graph is a series of cross-sections (it does not track the same cohort of learners, but only gives the number of learners in each grade for each year), it gives a good overall picture of trends and patterns in the entire education cycle.
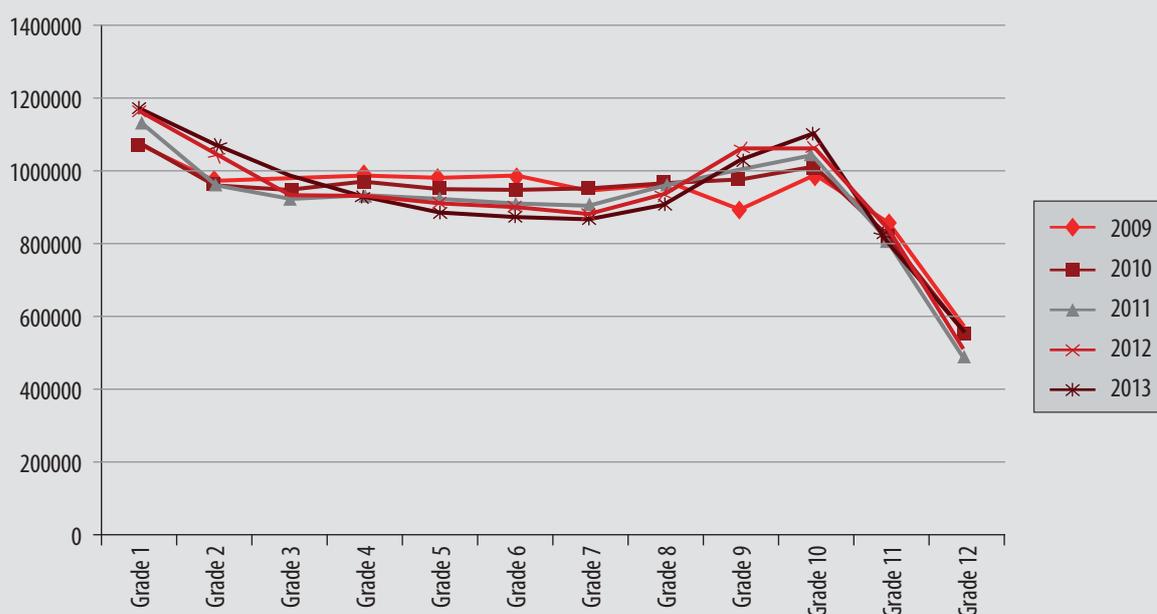
**FIGURE 1: Enrolment in public schools by grade and year, 2009–2013**

4. Annual schools survey (ASS)

This is a comprehensive survey of all public and independent schools in South Africa. It was designed to provide comparable information on public and independent schools and trend data over time. The ASS is completed by all schools in the country on a specific day, usually in March. These data are a useful resource for determining the proportion of over-age learners, enrolment, repetition and dropout rates by gender and province. For example, calculations using ASS data show that of every 100 learners who start grade one, roughly 50 will drop out before grade 12 (most of this happens in grades 10 and 11), 40 will pass the NSC examination, and only 14 will qualify for university.

The ASS data enable us to answer questions like 'Where in the system is the highest dropout and repetition?' Figure 2, based on the ASS, shows patterns of enrolment by grade and age. In the lower grades most learners are in the grade appropriate for their age. But in the higher grades, because of repetition of grades, there is a much wider age range, even though some of those most over age may already have dropped out. The graph indicates an over-age (repeater) problem in the higher grades.
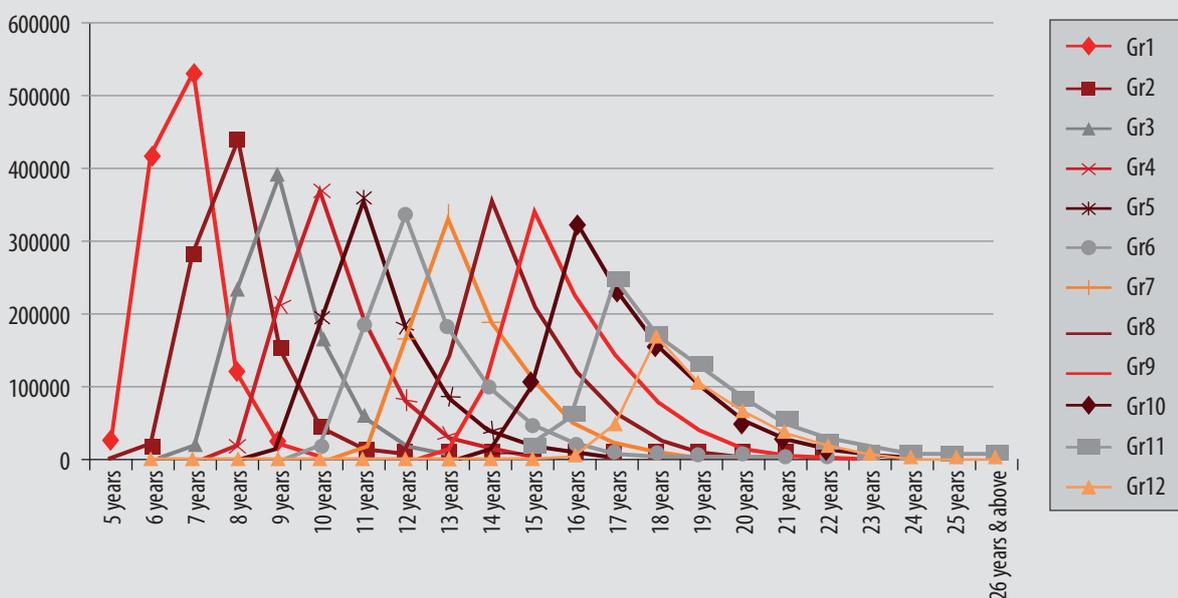
FIGURE 2: Enrolment by grade and age in 2011

Enrolment-driven data management is central to the South African government's efforts to redress inequalities. The enrolment-driven nature of the education system is reflected in the provision of teachers for school posts and in funding allocations, for example those based on the national norms and standards for school funding. Accurate and up-to-date data are a prerequisite for efficient and equitable distribution of resources.

5. Performance data

School outcomes are often used to determine the quality and efficiency of the education system. Standardised test scores are the best way to measure school performance, particularly if compared with international scores. In South Africa, grade 12 matric results used to be the only nationally standardised exam available. Since 2011 the annual national assessment (ANA), a test in literacy and numeracy for grades 1 to 6 and grade 9, has also become available. Comprehensive learning outcomes data now exist, including performance data in South Africa such as the national senior certificate (NSC) and the annual national assessment (ANA), and learning outcomes data including international test data such as SACMEQ, PIRLS and TIMMS. These five types of performance data are described below.

National Senior Certificate (NSC)

The NSC results, recorded since 1994, are captured per subject for every learner who writes the NSC exam. The body of research on student achievement and school performance in South Africa using the NSC data is growing and is of particular importance for policy formulation. The Schools Report is a publication by the DBE on the performance of individual schools in the NSC for the past three years (2012 to 2014). This available and accessible dataset is extremely useful and could increase the use of the NSC results, enabling analysts to determine patterns and trends in school performance to inform policymaking.

Annual National Assessments (ANAs)

The ANAs are standardised national assessments for languages and mathematics in the senior phase (grades 7 to 9) and the **intermediate phase (grades 4** to 6) and in literacy and numeracy in the foundation phase (grades 1 to 3). The question papers and marking memoranda (exemplars to guide the markers) are supplied by the DBE. Schools conduct the tests themselves and also mark and moderate them internally. This has raised some questions about the quality of such data. Discussions about external moderation and oversight are currently underway, and some rounds of the ANA have incorporated a 'verification' component using an external service provider to re-mark a sample of scripts and oversee the test administration in a subsample of schools.

Southern and Eastern African Consortium for Monitoring Educational Quality (SACMEQ)

SACMEQ is an international consortium of 15 Ministries of Education in southern and eastern Africa that work together to share expertise in teaching education planners to apply scientific methods to monitor and evaluate the conditions of schooling and the quality of education. A growing body of literature in South Africa uses SACMEQ data in empirical studies to evaluate the learner in South Africa. SACMEQ was conducted at grade 6 level in South Africa in 2000, 2007 and 2013.

Progress in International Reading and Literacy Study (PIRLS)

PIRLS is an international comparative study aimed at benchmarking reading and literacy levels across countries. It tests grade 4 children from 49 countries. This worldwide assessment and research project is designed to measure trends in children's literacy and to collect information about policy and practices related to learning to read and teaching reading. PrePIRLS is a stepping stone to participating in PIRLS. It provides a way to assess reading at the end of the primary school cycle for a range of developing countries. It uses the same concept of reading as PIRLS but sets less difficult reading tests. Fewer countries have taken part in PrePIRLS than in PIRLS, which limits the comparative dimension.

Trends in International Mathematics and Science Study (TIMSS)

TIMSS tests the mathematics and science achievement of grade 8 learners internationally and grade 8 and 9 learners in South Africa. South Africa participated in TIMSS in 1995, 1999 (grade 8 only), 2002 (grades 8 and 9) and 2011 (grade 9 only).

6. Data sharing

EMIS datasets such as the ASS or performance datasets such as the NSC examination results on their own do not provide comprehensive enough information for all aspects of education planning, monitoring and policy formulation. Combining different data sources is vital for decision making and policymaking. Relevant information must be extracted from multiple sources (EMIS, examinations results and so on) and linked, integrated or merged by using a common field across those sources.

Data integration is vital for creating longitudinal datasets. These datasets track the progress of individual learners through the education system or the development of individual schools. The learner or school is assigned a unique identification code which is kept consistent and accurate from year to year. A learner's progress may be tracked not only over time but also across schools or districts within the country.

7. Learner unit record data systems in South Africa

Some systems in South Africa keep records of individual learners. They enable us to answer questions like 'Who are the learners who dropped out of the system?' or 'What is the profile of learners who progressed without any repetition?' The following three systems produce learner unit records.

School Administration and Management System (SAMS)

SA-SAMS is an off-line (desktop) school administration and management system that has been widely distributed and piloted in all provinces. The SA-SAMS rollout plan and staff training are managed by each province. The availability of such learner unit data would in principle make it possible to analyse the flow-through patterns in terms of repetition, dropout and progression, which is not possible with aggregated datasets. The Free State has implemented the system and all data collection processes of the SNAP and ASS are conducted using this method. The method has also been successfully implemented in the Eastern Cape.

Learner Unit Record Tracking System (LURITS)

LURITS is a standardised system that assigns a unique national identifier to learners in South African public schools. It tracks the movement of learners from year to year and also from school to school and provides accurate enrolment numbers and learner profile data for planning and strategic decision making. LURITS is not yet fully operational.

Centralised Education Management Information System (CEMIS)

CEMIS is a web-enabled system used in the Western Cape, mainly as a learner registration and tracking system. It registers learners so as to track and monitor individual learners in the province, their registration, transfers between schools, examination passes, and so on.

8. Data quality problems and recommendations

Since its establishment in 1994 the DBE has been tasked with providing good, relevant data for policy, planning and administration and developing mechanisms for accountability and for monitoring and evaluating the education system. It has been helped in this task by the establishment of EMIS in 1995.

On the basis of my research I offer four policy recommendations for improving the quality of datasets and making them accessible in a form that will be user-friendly both for research and for policymaking.

1. Although the master list of schools is publicly available on the DBE website, the quality of the data is of concern. Quality should be improved by regular updating and filling in of missing values, particularly in key fields such as quintile or exam number. For example, if we look at the item 'NoFeeSchool' in the Q1 2015 Masterlist, we find that 35% of schools are classified as 'To be updated'.
2. The master list of schools should also be made available earlier in the year.
3. Accessibility of the SNAP and ASS data should be improved by making these data available in a user-friendly file format, such as a database, spreadsheet or comma delimited text file that can be downloaded.
4. The current ANA and NSC data, already accessible on the DBE website in PDF format, should be made available in a more user-friendly data file format, such as a database, spreadsheet or comma delimited text file that can be downloaded.