



National Schools Effectiveness Study

User Guide for data

JET Education Services

14 September 2014

Acknowledgements

The National Schools Effectiveness Study (NSES) was designed and managed by JET Education Services, and subsidised by the Royal Netherlands Embassy and JET. Aneesha Mayet led the fieldwork. The final research report has been published as a book, entitled “Creating Effective Schools”, written under the editorship of Nick Taylor, Servaas van der Berg, and Thabo Mabogoane. It is requested that scholars publishing using the NSES data acknowledge the role of JET and the Royal Netherlands Embassy.

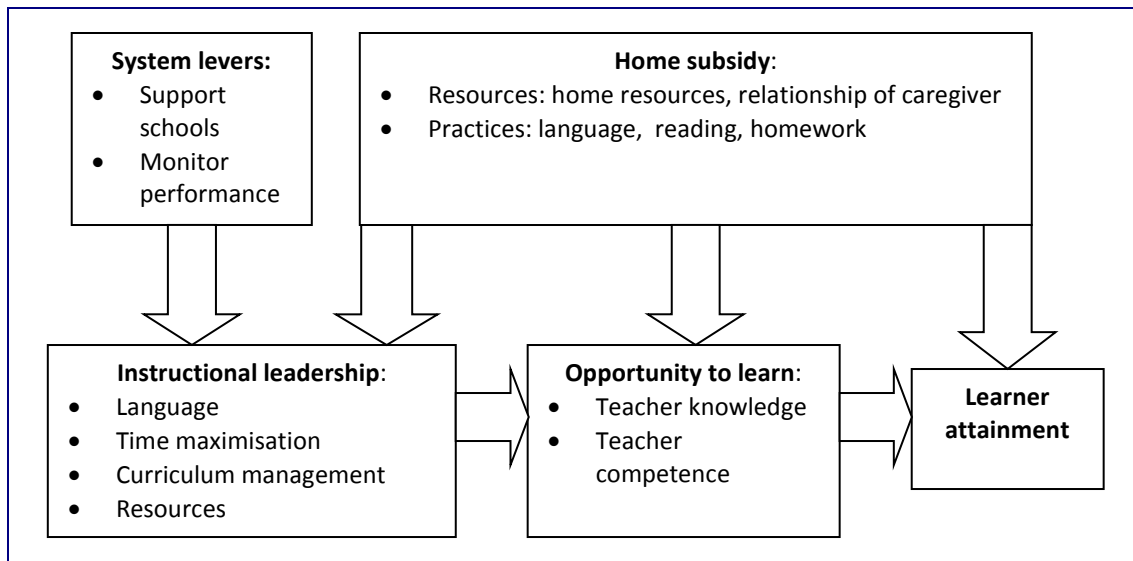
Overview of the NSES research design

The NSES is a research study aimed at identifying lessons for policy and practice for government, principals, teachers, and parents. National policy lessons arising from any research study are most powerful if they can be shown to apply to the whole school population, and in order to address this consideration a nationally representative sample of 268 schools was drawn for the study. All provinces were included in the sample except Gauteng, which was excluded when it was discovered that provincial tests were being written at the same time as the first round of NSES data collection.

The NSES followed a cohort of children for 3 years, commencing with Grade 3 in 2007 and ending with Grade 5 in 2009. Around 16 000 children participated in each year of data gathering, within which a cohort of 8 383 was successfully tracked over all three years. Fieldwork in 2007 was conducted in the month of October, while in 2008 and 2009 fieldwork was conducted during August.

The study covered a wide range of outcomes and was intended to generate information on what is important in shaping student performance in the context of South African education. Figure 1 summarises the National School Effectiveness Study research model. This model is based on a number of assumptions about how schools work in affecting learning. The model assumes that at the end of the learning chain, a set of outcome measures represent learner progress. Attainment scores in reading and mathematics are the dependent variable. Other measures are defined as intervening independent variables which, with varying degrees of independence and interdependence, impact on learner attainment.

Figure 1: National School Effectiveness Study Research Model



Learner performance, the main outcome variable, was assessed by means of literacy and mathematics tests which were administered in English (in all schools) to the learners. It is common in large scale studies (such as TIMSS, SACMEQ or PIRLS) to collect data on educational activities by means of survey questionnaires. Such methods do not always provide the most valid kind of data, given the well-known tendency for principals and teachers to reflect their practices in a favourable light. Thus, the NSES school-level and classroom data was collected by means of interviews and direct observations using structured instruments and fieldworkers experienced in the work of schools. Due to budget limitations the NSES did not undertake classroom observations. This is a limitation of the study, given the importance of teaching quality to learner performance. However, we did assess teacher practices through an analysis of planning and assessment records, and undertook a detailed analysis of pupil writing in both maths and literacy by looking at all the exercise books in each subject of the best student in each class. We also administered a very short test in their respective subjects to maths and language teachers.

The survey instruments

The project administered the same test to these learners each year and also collected data on home background, learner characteristics, teacher knowledge, learner writing, school resources, and curriculum management practices in a sample of 268 schools in eight of the nine provinces (excluding Gauteng).

In 2007, literacy and mathematics assessments were administered to participating grade 3 learners. In addition, each learner completed a background questionnaire, and a

questionnaire was administered to school principals. In 2008 and 2009, children in grades 4 and 5 respectively were again tested in literacy and mathematics. In addition to learner background questionnaires and school principal questionnaires, in 2008 and 2009 teachers of grade 4 and 5 respectively, completed questionnaires. The teacher questionnaires contained teacher characteristics, self-reported information on classroom and other teacher practices and short subject knowledge tests. In addition, fieldworkers conducted a document review of, *inter alia*, learner workbooks to measure aspects of curriculum coverage and classroom practice. Consequently, the following survey instruments were administered and are provided in this public release:

Table 1. Survey Instruments administered in the NSES

Instrument	Year
Literacy and Numeracy Test Instruments	Same instruments administered in 2007, 2008 and 2009
Learner questionnaire	Administered in 2007, 2008 and 2009
Principal questionnaires	Administered in 2007, 2008 and 2009. Each year's instrument was somewhat different.
Teacher questionnaire	Administered in 2008 to grade 4 teachers and in 2009 to grade 5 teachers. It included short Literacy and Mathematics tests. The same tests were administered in 2008 and 2009.

The datasets provided

Four datasets are provided in Stata format. The first, labelled "RNE 2007-08-09" contains 8383 rows (one per individual student) for all the students that were tracked for all three years. The data is in wide format, such that each individual student will have variables corresponding to different years, e.g. homelanguage_2007 and homelanguage_2008. The unique identifier for learners who are part of this cohort is a combination of the school id and a number for the learner within a school. For example, learner 268121 is learner number 121 in school number 261. A similar naming

convention was used to identify maths and English teachers who were interviewed in 2008 and 2009. Teacher 268m001 is maths teacher number 1 in school 268. Teacher 44e001 is English teacher number 1 in school 44.

The other three datasets contain each year's full cross-section. The dataset labelled "RNE2007learner-principal" contains 16 503 grade 3 students (one row per student) in the year 2007. The dataset labelled "RNE2008learner-principal-teacher" contains 15 698 grade 4 students (one row per student) in the year 2008. The dataset labelled "RNE2009learner-principal-teacher" contains 14 396 grade 5 students (one row per student) in the year 2009. Each of these three waves contains the 8383 students who participated in all three waves, as well as other students who either participated in one or two of the waves.

Selected data issues

1. Attrition

The NSES did not follow individual students who did not remain within their peer group (i.e. grade and school). Therefore, individuals who dropped out of school, repeated a grade, or changed schools were lost from the study. In addition, those students absent in any of the three survey waves are not included in the realised panel of 8383 students. In the South African context, a very small proportion of primary school-aged children drop out of school, so the main sources of attrition are probably absenteeism, grade repetition and school migration. However, the extent of attrition in the NSES is larger than one would expect these factors to produce. Only 8383 of the original 16 503 grade 3 students surveyed, also appear in the grade 4 and grade 5 samples. Therefore, it is likely that inaccuracies in the unique identifiers and matching process also contributed to attrition.

Given that attrition may not be random, the analyst may wish to apply some form of reweighting or other correction to address sample selection when working with the panel dataset.

2. Weights and complex sample design

The NSES sampling was stratified by province (although Gauteng did not participate). The survey followed a stratified one-stage cluster design with province as the unit of stratification. Within provinces, schools were sampled with equal probability. Within

selected schools all learners in the grade were sampled. Consequently, within provinces the sample is self-weighting. For the calculation of national statistics, the variable “learner_weight” should be used. This variable is the same for all students within each province. When calculating statistics at the school-level (i.e. when each school is counted once) the “school_weight” should be used. This weight variable is equal to the “learner_weight” multiplied by the number of learners in the grade.

When calculating standard errors it is also necessary to adjust for the fact that students are clustered within schools. We therefore recommend using STATA’s complex sampling commands, by first specifying the survey design (as shown below) and then using “svy:mean” and “svy:reg” to calculate means and to run regressions, respectively:

```
svyset idschool [pweight=learner_weight], strata(province)
```

3. String variables

Many of the variables in the data are currently in string format. One can make use of STATA’s “destring” command to convert such variables into numeric format. For example, the variable “q4422” is a string variable for the number of assessment records in mathematics in 2009. The following code will generate a new variable called “math_records_2009” in numeric format (the “force” option converts non-numeric strings such as “vs” into missing data):

```
destring q4422, gen(math_records_2009) force
```

4. Data inconsistencies

Incompatible responses across years or across instruments may exist and these have not been changed through a data cleaning process. For example, several individuals have a different gender across the years of the survey, according to the data. Where such inconsistencies occur, the analyst will need to make their own decisions.

5. Many zero scores in 2007

A large number of students have recorded scores of zero in the literacy and/or numeracy tests in 2007. For example, 487 out of the 8383 students in the realised panel have zero scores in literacy in 2007 and 523 of the 8383 students have zero scores in

numeracy. These numbers come down sharply in 2008 and 2009. The fact that many children were not being instructed in English in 2007 and yet had to write the tests in English may have contributed to this situation. However, researchers should be aware of this feature of the data given that it is possible that some other data capturing peculiarity may have contributed. One possibility is that numerous cases of non-response were coded as zero, instead of missing. Data users may prefer to recode all zero scores to missing, as a sensitivity check in their analysis.

6. Question on other children in the household

The question in the learner questionnaires about how many other children live at home is problematic because the questionnaire left no response category equal to zero other children. The minimum number of other children amongst the available response categories was 1. This variable should therefore be used with caution.

7. Missing data

In all variables in the four data sets where no distinction has been made between the answer “No” and missing data, this response has been left as system missing.

Recommended acknowledgement

We recommend that researchers using the NSES data include the following acknowledgement in their work:

Data for the National School Effectiveness Study (NSES) were collected between 2007 and 2009 on a nationally representative sample of schools in South Africa. The project was managed by JET Education Services and funded by the Royal Netherlands Embassy.

Recommended citation

The following citation should be included in all papers using the NSES data:

JET Education Services. (2010). The National School Effectiveness Study: Data for Waves 1 to 3. November 2010.

Copyright notice:
Copyright: JET Education Services; 2014

JET Division: Research and Planning
Project: RNE NSES
Funder: Royal Netherlands Embassy
JET Education Services
PO Box 178
WITS 2050 South Africa.
Tel: +27 011 403 6401

Fax: +27 011 3397844
Web site: [Http://www.jet.org.za](http://www.jet.org.za)