NSC School-level Datasets: Short User Guide

May 2026

1. Overview

These school-level National Senior Certificate (NSC) datasets were derived by the DBE Directorate: Research Coordination, Monitoring and Evaluation (RCME) from the November NSC "343" reports received from the Directorate for National Examinations. The data used for this release covers the years 2008 to 2025.

2. Source data and coverage

The source dataset contains full-time candidates appearing in the November NSC examinations. Supplementary examinations in earlier years and June examinations in more recent years are excluded. Users should note that, during the Multiple Exam Opportunities (MEO) period, some candidates wrote subjects across more than one examination opportunity. This affects the interpretation of incomplete result statuses and means that the total numbers of candidates in these datasets might differ slightly from published "Total Wrote" figures in NSC reports.

Note that in 2019 some changes were made to the education district classifications, meaning that the district variable in the datasets from 2008 to 2018 is somewhat different from the district variable in the datasets from 2019 to 2025. The data contains separate variables for this: a district for pre 2019 and a district for post 2019 variables.

3. Available school-level files

File pattern	Unit of observation	Main purpose
NSC centre overall YYYY.dta	One row per year × school × gender	Overall NSC result-status counts by school and gender.
NSC centre subject YYYY.dta	One row per year × school × gender × subject	Subject-level mark distributions by school, gender and subject.

In the file names, YYYY refers to the examination year, for example NSC centre overall 2016.dta.

4. Overall school-gender dataset

The overall dataset is created by first keeping one record per candidate within a given year, then collapsing to school-by-gender level. This means each candidate contributes once to the total candidate count and to the relevant pass-category count.

Variable	Description
year	Year of the NSC examination.
emis	Unique school identifier. This is intended to be consistent across years.
gender	Candidate gender: 0 = Female; 1 = Male. Schools with candidates of only one gender will have one row.
province	Province code: 1 EC, 2 FS, 3 GP, 4 KZ, 5 LP, 6 MP, 7 NW, 8 NC, 9 WC.
quintile	Official school poverty quintile classification. Missing for some schools, especially independent schools.
sector	School sector: 1 = independent; 2 = public.
district	Old education district classification, used for 2008 to 2018 files.

district19	New education district classification, used for 2019 to 2025 files.
n_candidates	Number of candidates in the school-gender group.
n_pass_bachelor	Number of candidates achieving a Bachelor pass.
n_pass_diploma	Number of candidates achieving a Diploma pass.
n_pass_hc	Number of candidates achieving a Higher Certificate pass.
n_pass_nsc	Number of candidates achieving a National Senior Certificate pass.
n_fail	Number of candidates who did not achieve the NSC.
n_pass_lsen	Number of candidates achieving an NSC pass for learners with special educational needs.
n_incomplete	Number of candidates with missing result status in the November dataset.
n_any_pass	Number of candidates with any pass outcome: Bachelor, Diploma, Higher Certificate, NSC or LSEN pass.

5. Subject school-gender dataset

The subject dataset, where included, retains candidate-subject records before collapsing. A candidate who wrote seven subjects can therefore contribute to seven subject-specific aggregates. The unit of observation is year × EMIS school × gender × subjectcode.

Variable	Description
year, emis, gender, province, quintile, sector	Same definitions as in the overall school-gender dataset.
district_pre2019/ district_post2019	Old district variable for 2008 to 2018; new district variable for 2019 to 2025.
subjectcode	NSC subject code. For example: 4 = Accounting, 24 = English First Additional Language, 25 = English Home Language, 53 = Life Orientation, 56 = Mathematical Literacy, 57 = Mathematics, 63 = Physical Sciences.
n_candidates_subject	Number of candidates with a non-missing final percentage in that subject-school-gender group.
n_mark_0_9 to n_mark_90_100	Number of candidates in each ten-percentage-point mark band: 0–9%, 10–19%, …, 90–100%.
n_mark_missing	Number of subject records with missing final percentage, if this variable is included in the file.
mean_finalperc	Mean final percentage for the subject-school-gender group, calculated over non-missing marks.

6. Important interpretation notes

- For some exam centres, there was no valid EMIS number in the source data. In these cases, the “centreno” was imputed into the “emis” field. In these cases, it may not be possible to merge with other datasets (such as the Masterlist of Schools) on “emis”.
- The count variables are counts, not percentages. Users can calculate rates by dividing by n_candidates or n_candidates_subject as appropriate.
- For overall pass rates, use n_any_pass / n_candidates or the relevant pass-category count divided by n_candidates.
- For subject mark distributions, the denominator should normally be n_candidates_subject. If missing marks are relevant to the analysis, include n_mark_missing in the denominator explicitly.
- The old and new district classifications are not directly comparable. Use district_pre2019 for 2008–2018 and district_post2019 for 2019–2025.
- The school-level files do not contain individual candidate identifiers, race, age or subject-level individual records.
- Because data are aggregated by gender, very small cells may still occur in some schools or subjects. Users should exercise care when reporting results for very small groups.

7. Example Stata code

Calculate Bachelor pass rate in the overall dataset:

```
use "NSC centre overall 2025.dta", clear
gen pct_bachelor = 100 * n_pass_bachelor / n_candidates
```

Calculate the percentage of Mathematics candidates achieving 50% or above:

```
use "NSC centre subject 2025.dta", clear
keep if subjectcode == 57
gen n_50plus = n_mark_50_59 + n_mark_60_69 + n_mark_70_79 + n_mark_80_89 + n_mark_90_100
gen pct_50plus = 100 * n_50plus / n_candidates_subject
```

Append annual overall files:

```
clear
save "NSC centre overall all years.dta", emptyok replace

forvalues y = 2008/2025 {
    use "NSC centre overall `y'.dta", clear
    append using "NSC centre overall all years.dta"
    save "NSC centre overall all years.dta", replace
}

use "NSC centre overall all years.dta", clear
```