
GOVERNMENT NOTICE


DEPARTMENT OF BASIC EDUCATION

No. 484

6 June 2011

SC009: STANDARD FOR DATA CODING

I, Angelina Matsie "Angie" Motshekga, Minister of Basic Education, after consultation with the Council of Education Ministers and in terms of section 3(4)(a) of the National Education Policy Act, 1996 (Act No. 27 of 1996), as read with section 59 of the South African Schools Act, 1996 (Act No. 84 of 1996) amend the Education Information Policy as read with paragraph 62 of the Education Information Policy published in Government Gazette No. 26710 of 27 August 2004, amended 6 August 2010, gazette 33426, hereby publish the standard SC008: Data Quality Standard for Surveys as set out in the schedule.



MRS ANGIE MOTSHEKGA, MP
MINISTER

DATE: 25 February 2011

SCHEDULE

SOUTH AFRICAN EDUCATION INFORMATION STANDARDS

SC009

Standard for Data Coding

December 2009



basic education

Department:
Basic Education
REPUBLIC OF SOUTH AFRICA

Education Management Information Systems (EMIS)
Department of Basic Education
Private Bag X895
0001 PRETORIA

Points of Contact

Mr Enos Sello
Telephone: 012 357 3670
Email: sello.en@dbe.gov.za

Or

Ms Khabo Skosana
Telephone: 012 357 3667
skosana.k@dbe.gov.za

Table of Contents

1. Purpose of the standard.....	4
2. Scope and Applicability of the standard	4
3. Standard for data coding	4
3.1. Standard for defining data codes	4
3.3. Standard for the coding of invalid, null and zero values during the data capture process.....	5
4. Standard for the core tables used in EMIS survey data sets.....	7
4.1. Standard for the General Information table of an institution	7
4.2. Standard for the master enrolment tables of each survey data set.....	10
4.2.1. Standard for the learner/student master enrolment table	10
4.2.2. Standard for the staff master enrolment table	10
4.3. Standard for the Core Information tables of each annual survey data set.....	11
5. Standard Code Lists.....	11
6. APPENDIX: CODE LISTS MOST COMMONLY USED IN EMIS SURVEYS AND DATA SETS	

1. Purpose of the standard

The purpose of this Standard is to specify the rules for the development of data codes to be used in the design of data capture tools and instruments for the capturing and consolidation of education data sets. Standard data code lists allow data values from different education data sets to be mapped into a common set of values, in order to ensure consistency of data exchange and reporting across systems. The use of standardised data codes and standard data tables improves the quality and the interoperability of data by increasing data compatibility between data sets and systems.

2. Scope and applicability of this Standard

This Standard applies to the development of data codes, so as to be used to enumerate list values in the design of any data capture instrument, whether paper-based or electronic, for an education data set. The Standard also specifies the key tables to be used in education survey data sets. The appendix is published quarterly and includes a list of the approved data codes that are used to enumerate the list values used in education data sets.

3. Standard for data coding

3.1. Standard for defining data codes

Data codes are used for enumerating the values of classifications that are used in education data sets. Data codes are used extensively in the EMIS surveys and in all electronic education systems. The following rules must be applied when defining new data code lists:

- a) The sequence of codes must be irrelevant.
- b) The codes must not be overloaded with meaning and the only intelligence that may be embedded in a code is error detection and possibly error correction.
- c) Codes may never be reused and cannot be changed.
- d) Codes must be hidden from the user in all electronic systems. The user must only be presented with the descriptions of the codes from which to select an option.

- e) Where codes need to be used on hard-copy survey forms to save space, all codes must be fully enumerated on all the survey forms.
- f) New and temporary codes may only be released by the Education Information Officer of the national Department of Basic Education.

3.2. Standard for the capturing of data codes

The following rules apply to the capturing of data codes into electronic systems:

- a) Each data field that requires a coded answer must be restricted to the authorised code values only.
- b) The electronic data capture tool must not allow code values outside of the required range to be captured.
- c) Yes and no (true/false) data fields must be encoded and must not use 0 as a default answer for either 'Yes' or 'No' (true/false).

3.3. Standard for the coding of invalid, null and zero values during the data capturing process

3.3.1. The Standard makes provision for a distinction between a zero value, an incorrect value and a null value (no response) in the data sets, based on the following rules:

No answer

- a) If the user or respondent fails to enter a value into an answer cell (for all data types) in a flat structure, then an empty string or null value must be captured for that record or data row, so as to indicate a null response to the question.
- b) If the user or respondent fails to enter a value into an answer cell in a normalised structure (such as a grid), then no record or row must be captured.

Zero answer

- c) If the user or respondent enters a zero (0) value into a numeric answer cell in a flat structure, then a zero (0) must be captured for that record or data row.

- d) If the user or respondent enters a zero into an answer cell in a normalised structure (such as a grid), then the record or row must be captured with a zero (0) answer.

Invalid answer

- e) If the user or respondent enters a zero (0) value into a text or date answer cell in a flat structure, then a zero (0) must be captured for that record or data row, indicating an invalid response.
- f) If the user or respondent enters a code value that falls outside the code range for a range answer cell, then a zero (0) must be captured, indicating an out-of-range error or invalid response.
- g) Answers, such as 'Not applicable' and 'Other' must be included in the code range if they constitute a valid answer.

Not applicable

- h) If the option, 'Not applicable', is a valid response, then this should be coded as a list value and should form part of the valid code range.

Other

- i) If the option, 'Other', is a valid response, then this should be coded as a list value and should form part of the valid code range.

3.3.2. The following symbols may also be used to indicate the difference between a zero value and incorrect/missing data in text answer fields:

- n: Quantity nil/zero
- m: Data missing (not answered)
- a: Not applicable
- x: Invalid response

3.4. Standard for the naming of data tables and fields in a data set

The following conventions must be used in the data sets:

- a) Names of fields and tables should be descriptive of the entity or attribute.
- b) Any table containing list code values must start with the prefix '*lst_*'.
- c) Any table containing data values must start with the prefix '*tbl_*'.
- d) Multiple occurrences of the same field must inherit an occurrence number, .e.g. *CellNumber_1*, *CellNumber_2*.

4. Standard for the core tables used in EMIS survey data sets

The requirements for conducting EMIS surveys were set out in the *Government Gazette*, No. 29757 of April 2007.

The requirements listed below, specify the standard for the core tables to be included in each of the surveys specified in the above-mentioned *Government Gazette*.

4.1. Standard for the general information table of an institution

The same core fields must be used for the general information table of every survey data set, in order to ensure consistency across the different data sets in the identification and description of educational institutions.

4.1.1. The minimum core fields, which must be used in each survey for the identification of the institution, are the following:

Field	Data type	Description
National EMIS number	Numeric (9 digits)	National EMIS number of the institution.
Name of institution	Text (200)	The official name of the institution.
Province code	Numeric (1 digit)	The official numeric code for the province. (This is a code value that is linked to the list values for the allowed provincial code values.)
District/Region	Text (100)	The official name of the district or region.

4.1.2. The physical address fields for the institution must be included in all annual survey data sets and must always be in the following format on all forms where they are used:

Field	Data type	Description
Street number	Text(10)	Physical address: Full street number, including prefixes (e.g. A20), suffixes (e.g. 33C) and ranges (e.g. 18 – 20).
Street name	Text (100)	Physical address: Street name, including type (e.g. Road, Avenue); modifier (e.g. Extension); and directional (e.g. North, East), but excluding number. The draft South African Address Standard, SANS 1883, caters for seven non-postal types of addresses, viz. building address, farm address, informal address, intersection address, landmark address, site address and street address. This field may be used to carry the relevant data for these other types of addresses.
Town/City	Text (100)	Physical address: Name of town or city.
Country	Text (100)	Physical address: Country.
Postal code	Text (20)	Physical address: Postal or zip code.

4.1.3. The postal address fields for the institution must be included in all annual survey data sets and must always be in the following format on all the forms where they are being used:

Field	Data type	Description
Postal elements	Text (200)	SANS 1883 and the South African Post Office have five postal address types, four of which are relevant here: box/bag address, site address, street address and village address (the other is Post Restante). This field contains the one or two lines that appear above the Post Office name: <ul style="list-style-type: none"> • Box or bag number. • Mail delivery agent identifier (e.g. PostNet suite number; bag number). • Address number; section identifier or cadastral reference. • Building unit identifier and/or street number; street identifier. • SAPO village house number; SAPO village name.
Post Office	Text (100)	Postal address: Name of Post Office
Country	Text (100)	Postal address: Country
Mailing postal code	Text (100)	Postal address: Postal code

4.1.4. The contact data fields for the institution must be included in all annual survey data sets and must always be in the following format on all forms where they are being used:

Field	Data type	Description
Telephone number	Text (20)	The full 10-digit number, including the area code for local numbers. The field must also allow for party line numbers and international numbers.
Facsimile number	Text (20)	The full number, including the area code. This field must also allow for fax-to-cellphone numbers and international numbers.
Cell number	Text (20)	The full 13-digit cellphone number. This field must also allow for international numbers.
E-mail address	Text (150)	The full e-mail address.

Multiple occurrences of the same field must inherit an occurrence number, e.g. *CellNumber_1*, *CellNumber_2*.

4.2. Standard for the master enrolment tables of each survey data set

Each survey data set must contain master enrolment tables that are used to control the totals of all other survey tables within the data set. The two main master tables are the learner/student-enrolment master table and the staff-count master table.

4.2.1. Standard for the learner/student-enrolment master table

This table must capture the numbers of all learners/students enrolled at the institution, according to grade and gender, using the following specifications:

Field	Data type	Description
National EMIS number	Numeric (9 digits)	National EMIS number of the institution.
Data year	Numeric (4 digits)	The relevant survey year.
Grade code	Numeric (2 digits)	The official numeric code for the relevant grade. (This is a code value that is linked to the list values for the allowed grades or level values.)
Gender code	Numeric (1 digit)	The official numeric code for the relevant gender. (This is a code value that is linked to the list values for the allowed gender values.)
Quantity	Numeric	This is the answer or value entered into the cell that represents the number of learners, based on the variables of grade and gender.

4.2.2. Standard for the staff-count master table

This table must capture the numbers of all members of staff employed at the institution, using the following specifications:

Field	Data type	Description
National EMIS number	Numeric (9 digits)	National EMIS number of the institution.
Data year	Numeric (4 digits)	The relevant survey year.
Remuneration code	Number (1 digit)	The official numeric code for the relevant remuneration type. (This is a code value that is linked to the list values for the allowed remuneration values, such as state or SGB employees.)
Staff type code	Number (2 digits)	This is a code value that is linked to the list

		values for the allowed staff type values, such as educator, support staff, administrative staff.
Duration code	Number (2 digits)	This is a code value that is linked to the list values for the allowed duration values, such as full-time or part-time.
Nature of appointment code	Number (2 digits)	This is a code value that is linked to the list values for the allowed nature of appointment values, such as permanent, temporary, substitute.
Gender code	Number (1 digit)	This is a code value that is linked to the list values for the allowed gender values.
Quantity	Number	This is the answer or value entered into the cell, representing the number of staff members in the category, based on the above variables.

4.3. Standard for the core information tables of each annual survey data set

- a) Each annual survey data set must contain the following core data tables for data collection:
- i) Learner/Student data table, according to age, grade and gender.
 - ii) Learner/Student data table, according to population group, grade and gender.
 - iii) Learner/Student data table, according to home language, grade and gender.
 - iv) Learner/Student data table, according to language of instruction, grade and gender.
- b) The totals in these core tables must balance with the totals in the master enrolment table.
- c) Each annual survey data set must contain a table that collects the number of learners/students per grade, and the gender of learners who experience barriers to learning, per barrier description.

5. Standard code lists

The following standards apply to the use of common code lists for education data sets:

- a) Only the national Department of Basic Education may create new standard code lists for Basic Education.

- b) Any user of education data, who requires a new code to be generated, must apply to the relevant section of the national Department of Basic Education to generate the code.
- c) All code lists must be approved by the Education Information Officer of the National Department of Basic Education, before they may be released for use.
- d) All code lists in use by the Department of Basic Education must be lodged with the Education Information Officer.
- e) All code lists can be obtained from the Education Information Officer.
- f) The Education Information Officer will release quarterly updates of the code lists as per the attached Appendix. The quarterly code lists will be published on the Department of Basic Education's website.
- g) Each code list will contain the date of release and the version number of the code list.
- h) The layout of all published code list tables must be as follows:

Field name	Description
Code	The value of the code.
Description	A description of the code.

- i) The name of the code list table must start with the prefix '*lst_*' and indicate the code type, such as '*lst_grades*' (code list table enumerating the school grades).

6. APPENDIX: CODE LISTS MOST COMMONLY USED IN EMIS SURVEYS AND DATA SETS

Release date: December 2009

Version 1.0

Note: There are several code lists defined outside of the EMIS surveys and data sets, which are in use in education systems. Some of these code lists are very extensive and to repeat them in this Appendix would be redundant and would increase the risk of conflicting standards. An example of such a code list is the subject code list for FET subjects, which is defined by the Curriculum Unit of the Department of Basic Education, but is obtainable from the Education Information Officer.

The following code lists are defined by EMIS for use in education data sets. The reference code links the code list table to the description of the entity, as defined in the SC006 *Dictionary of Education Concepts and Terms*.

Reserved code numbers

The following code numbers are reserved and may be used with any code list, if so required:

- Code Number 94 is reserved for the option, 'Not applicable', and may be included in any code list, where required, by the particular data collection.
- Code Number 95 is reserved for the option, 'None', and may be included in any code list, where required, by the particular data collection.
- Code Number 96 is reserved for the option, 'Other', and may be included in any code list, where required, by the particular data collection.
- Code Number 99 is reserved for the option, 'Unknown', and may be included in any code list, where required, by the particular data collection.

Alphabetical index of code list tables

Table	Ref. code
Academic barriers to learning	10038
Accommodation for ECD centres	10100
Address types	10007
Age	10019
Age enrolment	10019
Age groups for ECD learners	10019
Age groups for mortality statistics for educators	10019
Age groups for mortality statistics for learners	10019
Assessment body	10030
Assistive devices for learners with special educational needs	10200
Availability of assistive devices	10200
Barriers to learning	10038
Combined school	10064
Disabilities of educators	10276
Duration of appointment	10098
Education band	10146/10140/10160
Education subsector	10386
Educator qualification type	10312
Ex-Department of Education	10122
Extra-curricular activities	10125
Funding type	10139
Gender	10144

General Education and Training	10146
Grade	10149
Intervention school	10186
Languages	10193
Language level	10190
Language of Learning and Teaching	10191
Language type	10165/10191/10284
Learner transfer categories	10198/9
Legal status of an institution	10205
Mortality	10229
National EMIS number	10235
Nature of appointment	10242
NQF levels	10239
Orphan	10257
Ownership of building	10420
Ownership of institution	10262
Ownership of land	10263
Parent/learner relationship	10266
Personnel categories	10274
Population group	10279
Practical status of institution	10283
Professional non-teaching staff	10250
Progression status of learners	Legacy Codes
Provinces	Definition not required
Qualification category	10311
Qualification type	10312
Remuneration source	10322
Reporting status of institution	10325
Rural/Urban	10331/10412
School level	10342
School phases	10343
School specialisation	10346
School term	10347
Section 21 functions	10353
SNE school specialisation	10374
Specialised facility	10377
Staff type	10379
Stakeholders	10380
Support staff	10392
Teaching level	10398
True/False	Definition not required
Yes/No	Definition not required

Reference code index of code list tables

Reference code	Table
10007	Address types
10019	Age
10019	Age enrolment
10019	Age groups for ECD learners
10019	Age groups for mortality statistics for educators
10019	Age groups for mortality statistics for learners
10030	Assessment body
10038	Academic barriers to learning
10038	Barriers to learning
10064	Combined school
10098	Duration of appointment
10100	Accommodation for ECD centres
10122	Ex-Department of Education
10125	Extra-curricular activities
10139	Funding type
10140	Further Education and Training
10144	Gender
10146	General Education and Training
10149	Grade
10186	Intervention school
10190	Language level
10191	Language of Learning and Teaching
10193	Languages
10198/9	Learner transfer categories
10200	Assistive devices for learners with special educational needs
10200	Availability of assistive devices
10205	Legal status of an institution
10229	Mortality
10235	National EMIS number
10239	NQF levels
10242	Nature of appointment
10250	Professional non-teaching staff
10257	Orphan
10262	Ownership of institution
10263	Ownership of land
10266	Parent/learner relationship
10274	Personnel categories
10276	Disabilities of educators
10279	Population group
10283	Practical status of institution

10311	Qualification category
10312	Qualification type
10322	Remuneration source
10325	Reporting status of institution
10342	School level
10343	School phases
10346	School specialisation
10347	School term
10353	Section 21 functions
10374	SNE school specialisation
10377	Specialised facility
10379	Staff type
10380	Stakeholders
10386	Education subsector
10392	Support staff
10398	Teaching level
10420	Ownership of building

CODE LIST TABLES

6.1. Academic barriers to learning

Ref. Code: 10038

Table: Ist_LearningBarriers_Academic

Academic difficulties that create learning barriers for mainstream or SNE learners.

Code	Description
1	Reading Difficulties
2	Numeric Difficulties
3	Language Difficulties

6.2. Accommodation for Early Childhood Development centres

Ref. Code: 10100

Table: Ist_AccommodationType

Type of accommodation used by an ECD centre.

Code	Description
1	School
2	Church
3	Community Hall or Centre
4	House
96	Other

6.3. Address types*Ref. Code: 10007***Table: Ist_AddressType**

Describes the type of address.

Code	Description
1	Physical Address
2	Postal Address
3	E-mail Address

6.4. Age*Ref. Code: 10019***Table: Ist_Age**

The age of a learner in years as at the last day of December in the year of the survey.

Code	Description
1	0-3
2	4 Years
3	5 Years
4	6 Years
5	7 Years
6	8 Years
7	9 Years
8	10 Years
9	11 Years
10	12 Years
11	13 Years
12	14 Years
13	15 Years
14	16 Years
15	17 Years
16	18 Years
17	19 Years
18	20 Years
19	21 Years
20	22 Years
21	23 Years
22	24 Years
23	25 Years
24	26 Years and Above
96	Other

6.5. Age enrolment

Ref. Code: 10019

Table: Ist_Age_Grade1Enrolment

The learner age of enrolment in Grade 1, as at the first day of January in the year of the survey.

Code	Description
1	Younger than 5 Years
2	5 Years
3	6 Years
4	7 Years
5	8 Years
6	9 Years
7	10 Years
8	11 Years
9	12 Years
10	12+ Years

6.6. Age groups used for ECD learners

Ref. Code: 10019

Table: Ist_ECD_AgeGroups

Different age groupings used in ECD surveys.

Code	Description
1	0 - 1 Years
2	2 - 3 Years
3	4 - 5 Years
4	5 and Below
5	6 Years
6	7 and Above

6.7. Age groups for mortality statistics for educators

Ref. Code 10019

Table: Ist_MortalityAgeEducators

Age groups used for educator mortality statistics.

Code	Description
1	20-24
2	25-29
3	30-34
4	35-39
5	40-44
6	45-49
7	50-54
8	55 and Older