

SDI Education datasets: Anonymization Protocol

Objectives

In order to release the SDI Education dataset as a Public Use File, it is necessary to ensure the privacy of its participants. To this end, and with the intention of avoiding the re-identification of the schools, teachers and pupils, the following measures are taken:

1. Deletion of variables that contain confidential information or lead to re-identification.
2. Deletion of value labels that contain confidential information or lead to re-identification.
3. Recoding of variables that could lead to re-identification of the observations into broader group categories.
4. Trimming and/or censoring of specific unique values and outliers that might allow re-identification.

Results

The resulting datasets keep the usefulness of the data intact while greatly protecting the privacy of the respondents and reducing the identification risk. Most harmonized variables (>80 %) are unaffected by the anonymization process. In very few instances, some variables are relabeled to avoid the publication of the exact questions/answers of confidential assessment instruments.

All identifiers are being cleaned (deletion of value labels). Some variables are being recoded into categorical variables to ensure confidentiality.

1. Disclosure risk and confidentiality protection

Microdata often contain confidential or sensitive information, which makes release of these datasets in their original form impossible. Release of the data could reveal this confidential information and lead to a breach of privacy of the respondents. This has ethical and, in many cases, legal objections. Furthermore, when confidentiality is not guaranteed, current and potential future respondents are less likely to be willing to respond in future surveys.

The aim of this process is to create a Public Use File (PUF), which is a dataset that is freely accessible to the greater public taking in consideration the identified issue. This PUF must also minimize as much as possible unnecessary disturbances to the original to preserve the usability and quality as much as possible when possible.

Risk in the Statistical Disclosure Control context is the probability or likelihood that disclosure by an (hypothetical) intruder of a record occurs. Disclosure can be **identity disclosure**, when the identity of an individual or entity in the dataset is correctly revealed, or **attribute disclosure**, when the intruder gains new (confidential) information from the dataset. Identity disclosure can imply attribute disclosure. The risk is dependent on several factors, amongst others the frequency of **keys** (i.e. combinations of values of key variables), sample size and sampling weights as well as the availability of external information to intruders to use for re-identification. The disclosure scenarios for a particular dataset describe these parameters and the way an intruder can use a dataset to gain new information.

The acceptable level of risk depends on the release type, e.g. scientific use file, public use file, or other ways of release and the sensitivity of the data. This dataset was prepared to be released as **PUF and hence needs a higher level of protection**. Also, the potential harm caused by disclosure should be considered when determining the acceptable risk level.

In similar microdata releases with, for instance, business survey data, the geographical level is highly reduced, large companies are suppressed and the level of detail in the data is reduced to protect the records. Generally, the period between the survey and data release is also specified, e.g. 1 year. In case of the SDI education survey, the period between the survey and the data release introduces already uncertainty into several variables, such as number of pupils. It should be noted that a complete elimination of disclosure risk is not possible.

2. SDI Education data set

The SDI Education dataset consists of a series of country-year surveys, each containing information of schools, teachers and pupils. The main concern for re-identification and confidentiality are the teachers and pupils. However, since the data are hierarchical, i.e. teachers and pupils belong to schools, the re-identification of a school might lead to the re-identification of a teacher or pupil too.

The datasets consist of the following:

- Set 1: School level data
 - Set 1a: School management/finances (Module 3)¹
- Set 2: Teacher level
- Set 3: Child level
- Set 4: Time on task

These datasets contain sensitive and confidential variables, especially on the level of the teacher but also on the school level.

3. Actions taken

The Anonymization process is done with the aid of the statistical software Stata. All anonymization steps are reproducible with the Stata script for each of the detailed datasets. The process starts from the harmonized dataset: before the start of the anonymization process, any final data quality corrections are made.

Each anonymization script covers the following steps:

1. Identification of ready-to-release variables.
2. Identification and removal/anonymization of variables due to the sensitivity of its data. Identification and removal/anonymization of variables due to the distribution of the data that could lead to high risks of re-identification.
3. Identification and recoding of variables into categories to deal with outliers (top recoding), sample unique and continuous variables whose values represent high risk of reidentification (special unique).

3.1 Ready to release variables.

There is a subset of variables that do not imply considerable risks of disclosure. These variables were identified and revised. We proceeded to select them considering:

- General information: Country, survey year, urban, etc. This information is preserved and there is no risk associated with the release.
- Randomized id/keys: region, district, province, etc. This information is shared considering it doesn't contain descriptive information. In general terms, it preserves the variability of the data but doesn't provide an associated value label.

¹ The publication of new (or updated) PUF datasets includes the Education SDI's module 3 which compiles information of the school's management and financial activities (e.g. budget, expenditure, fundraising, monitoring and evaluation, etc.). This module has been published separately from the school level dataset as it was not harmonized to a common standard. Given distinct educational (and administrative) structures, module 3 is uniquely designed according to the country's interests and context, creating constraints for a harmonization process across SDIs. It is important to note that, despite this dataset not being harmonized, it is still subject to a thorough anonymization process like the rest of PUF datasets.

- Specific information: a subset of infrastructure variables, a subset of assessment variables, other school and teacher characteristics, etc. These variables gathered through SDI survey represent a high risk of breach of confidentiality. We analyze each of these and their distributions to check if unique values allow for reidentification.

3.2 Identification of variables to delete

- Sensitive information

The SDI surveys gather data that includes sensitive information: names, financial information and specific descriptions, among others, that should not be available to public.

To avoid its disclosure, some of them are directly removed from the database and others are just transformed to a “Confidential” or “.c” value, allowing only the knowledge of their existence (to be considered its release upon request under strict protocols) and the identification of missing values.

There are other variables that do not represent a risk of disclosure but include important information that could damage future waves of surveys. These variables contain detailed descriptions of the assessments and their correct answers. All of them are being removed, keeping only general information on the type of question, its result/score and the label associated to interpret the latter.

- Variables with distributions that represent risks of reidentification

We identify variables with distributions that represent high risks of reidentification and for which recoding and other perturbative methods cannot account for the mentioned risks. These variables contain sample unique (a unique combination of values for the selected categorical key variables in the dataset and is at high risk of re-identification) and are represented mainly by descriptions/details of “Other” categories. It is not possible to share the information contained in any of the respective variables due to the specificity of the answers.

Table A and B shows the deleted and anonymized variables.

Table A. Overview of removed variables by dataset

Deleted Variable	Description
<u>School level dataset</u>	
gps_lat, gps_lon, gps_miss_ind, gps_lat2, gps_lon2	GPS variables
hfid, _GIS_merge	Variables that come from internal analysis and checks done by SDI team
<u>Teacher level dataset</u>	
All variables ending in “_mark”	Information on teacher’s specific answer
All variables ending in “_correction”	Information on correct answer

lit_4_denom

Variables that come from internal analysis of SDI team

Child level dataset

lang_nomiss-total_theta_se

Variables that come from internal analysis and checks done by SDI team

Note: a. Additional variables deleted for the 2016 Madagascar's Education SDI.

Table B. Overview of anonymized variables

Anonymized Variable	Description
<u>School level database</u>	
refusal_reason1	If refused, reason for refusal (1)
fv_enum_name1	First Visit Enumerator (s) 1 Name
sv_enum_name1	Second Visit Enumerator (s) 1 Name
region_name	Region Name
district_name	District name
town_code ^b	Town code
fokontany_code ^b	Fokontany code
pedag_admin_zone_c ^b	Pedagogical Administration Zone Code
ward_code	Ward
ward_name	Ward name
village_street_code	Village/street
village_street_name	Village/street name
school_name	School Name
school_code	School Code (Other ID)
emis_code	EMIS Code
fv_date_day	Date of 1st Visit (day)
fv_date_month	Date of 1st visit (month)
fv_date_year	Date of 1st visit (year)
sv_date_day	Date of 2nd Visit (day)
sv_date_month	Date of 2nd visit (month)
sv_date_year	Date of 2nd visit (year)
team_lead_date_day	Team leader date (day)
team_lead_date_month	Team leader date (month)
team_lead_date_year	Team leader date (year)
super_date_day	Supervisor date (day)
super_date_month	Supervisor date (month)
super_date_year	Supervisor date (year)
fv_arrival_time_hr	At what time did you arrive at the school? (hour)
fv_arrival_time_min	At what time did you arrive at the school? (min)
fv_end_time_hr	At what time did you leave at the school? (hour)
fv_end_time_min	At what time did you leave at the school? (min)
sv_arrival_time_hr	At what time did you arrive at the school? (hour) (second visit)

sv_arrival_time_min	At what time did you arrive at the school? (min) (second visit)
sv_end_time_hr	At what time did you leave at the school? (hour) (second visit)
sv_end_time_min	At what time did you leave at the school? (min) (second visit)
fv_date_comb	Date of First Visit (combined)
sv_date_comb	Date of Second Visit (combined)
team_lead_date_comb	Team leader date (combined)
super_date_comb	Supervisor date (combined)
fv_arrival_time	At what time did you arrive at the school: First Visit
fv_end_time	At what time did you leave the school: First Visit
sv_arrival_time	At what time did you arrive at the school: Second Visit
sv_end_time	At what time did you leave the school: Second Visit
county_name	County name
controller_name	Controller Name
village_name	Village name
control_date	Control Date
town_name	Town name
fokontany_name	Fokontany name
pedag_admin_zone_n	Pedagogical Administration Zone Name
fv_enum_name2	First Visit Enumerator (s) 2 Name
sv_enum_name2	Second Visit Enumerator (s) 2 Name
refusal_reason2	If refused, reason for refusal (2)
subcounty_name	Sub-County/Division/Town Council
parish_name	Parish Code
operator_id	Data entry operator ID
state_school_code	State School Code
private_school_code	Private School Code
knec_code_new	School KNEC Code (new)
knec_code_old	School KNEC Code (old)
tsc_code_public	School TSC Code (public)
moe_code	School MoE Code
division_name	Division name
IDEN_name	IDEN name
m1_resp_name	M1 What is your name?
m1_resp_phone	M1 Please can we have your Mobile Phone number
m1_sc_ownership	M1 What's this school's ownership type
m1_sc_ownership_other ^b	M1 Other Ownership
m1_resp_post_other ^b	M1 Which Position do you occupy in this facility? (Other)
m1_satelite_num	M1 If your school is a central or community school, how many satellites does it
m1_satelite_id	M1 If it is a satellite school, enter the identifier of its central school
m1_toilet_type5	M1 Toilet covered pit

m1_toilet_type6
 m1_toilet_type_other^b
 m1_toilet_type_other1
 m1_toilet_type_other2
 m1_water_type_other
 m1_road_type_other
 m1_means_transport_other^b
 m1_satellite_dist
 m1_comments1
 m1_qao_feed_discuss
 m1_school_type_other
 m1_qao_last_visit^b
 m1_school_open[n]^b
 m1_school_close[n]^b
 m2_comments1
 m2_comments2
 m2_comments3
 m2_n_comp_rooms
 m4_late_enum
 m4_teachname
 m4_comments1
 m4_comments2
 m4_comments3
 m4_comments4
 m4c_teach_callname
 m4c_teach_callname_n
 m4c_teach_projector
 m4c_teach_screen
 m4c_teach_ed_software
 m4c_teach_videogame
 m4c_teach_video
 m4d_start_teach
 m4d_post_other

M1 Toilet uncovered pit
 M1 Other types of latrine
 M1 Other – toilet
 M1 Other – latrine
 M1 Other Source of water
 M1 Other types of road
 M1 other means of transport
 M1 For satellite schools, how far is the school from Central and satellite? (km)
 M1 Comments 1
 M1 If Yes, did you discuss the information with:
 M1 What's the school type (Other)
 M1 When was the last visit of the official government quality assurance officer (day/month/year)
 M1 For each term *n* last year, when did the school open (day/month/year)
 M1 For each term *n* last year, when did the school close (day/month/year)
 M2A Comments 1
 M2B Comments 1
 M2B Comments 2
 M2 How many computer rooms?
 M4 How many minutes late were you? (If not late, write 0).
 M4 Teacher name
 M4 Comments: 1
 M4 Comments: 2
 M4 Comments: 3
 M4 Comments: 4
 M4 Did the teacher call pupils by name while teaching?
 M4 How many pupils did the teacher call by name?
 M4 Did the teacher use a video projector?
 M4 Did the teacher use a projector screen?
 M4 Did the teacher use an educational software or CD?
 M4 Did the teacher use a video game?
 M4 Did the teacher use a video?
 M4 What year did you begin teaching?
 M4 Other Position Specify

School management/finances (M3)^b

m3saq1_autre
 m3saq2d[x]
 m3saq2e
 m3sbq2c_autre[n]
 m3sbq2d_autre[n]
 m3scq10_autre

M3 respondent's position - Other (Specify)
 Financial origin type [x]: Observation
 Other origin (Specify)
 Item [n]: Source, other
 Item [n]: Type, other
 Who is responsible for developing budget - Other (Specify)

m3scq11_autre	Who is responsible for giving budget final approval - Other (Specify)
m3scq12g_autre	Budget information was provided to - Other (Specify)
m3scq13_autre	Who is responsible for purchasing materials within the establishment? - Other (Specify)
m3scq15_autre	Who monitors the use of funds? - Other (Specify)
m3scq17g_autre	What's in the report: Other (Specify)?
m3scq18j_autre	Grievance / comments / feedback mechanism: Other (Specify)?
m3scq2c2a	Other association (Specify)
m3scq6g_autre	Field to which the most amount is allocated - Other (Specify)
m3scq8	FEFFI member: did you encounter any difficulties in developing the program? - Other (Specify)
m3scq9	Times you ask general assembly meeting to communicate the SAE - Other (Specify)
m3sdq3d_autre_[x]	Books received for children for topic [x]: Other origin (Specify)
m3sdq4d_autre	Books received for teachers for topic [x]: Other origin (Specify)
m3sdq5d_autre	Books received for children for topic [x]: Other number (Specify)
m3sdq6d_autre_[x]	Books received for teachers for topic [x]: Other number (Specify)

Teacher level dataset

m2a_name	M2a First and last names
m2a_post_other	M2a Specify - Position in the school
m2a_educ_training_other ^b	M2a Other - Highest level of educational training (certificate) attained
m2a_contract_other ^b	M2a Specify - Contract Status
m2a_comments	M2a Module comments
m2b_fv_absence_r_other ^b	M2b Specify - Reason for absence in first visit
m2b_sv_absence_r_other ^b	M2b Specify - Reason for absence in second visit?
m2b_name	M2b First and last names
m2b_born_here	M2b Were you born in this [geographical area]?
m2b_salary ^b	M2b Is it possible to know your salary? what you have received lately?
m2b_perf_area_other ^b	M2b Other- Performance areas commented on
m2b_unpaid_claims_1	M2b Do you have any other unpaid claims? (1)
m2b_unpaid_claims_2	M2b Do you have any other unpaid claims? (2)
m2b_unpaid_claims_3	M2b Do you have any other unpaid claims? (3)
m2b_unpaid_claims_4	M2b Do you have any other unpaid claims? (4)
m2b_unpaid_claims_other	M2b Other unpaid claims?
m2b_comment1	M2b Module 2B Comments: 1
m2b_civil_status	M2b What is your civil status?

m2b_spouse_employment

m2b_comment2

m6_date_day

m6_name

m6_date

m6_post_other

m6_enum_name

m6_comments

teach_a1

teach_a2

teach_b1

teach_b2

teach_c1

teach_c2

teach_d1

teach_e_i

teach_e_ii

teach_e_iii

teach_f_i

teach_f_ii

teach_f_iii

teach_d2

swot_as1

swot_as2

swot_as3

swot_aw1

swot_aw2

swot_aw3

swot_bs1

swot_bs2

swot_bs3

swot_bw1

swot_bw2

swot_bw3

swot_as4

swot_aw4

swot_bs4

M2b What is the employment status of your spouse?

M2b Module 2B Comments: 2

M6 Date (day)

M6 Teacher name (Official first, middle and last names)

M6 Date (Day/Month/Year)

M6 Specify - What is your position at this school

M6 Enumerator(s) name

M6 Comments

3a) Tell the pupils what the aims of the lesson

3a) Tell the pupils what the aims of the lesson

3b) What specific learning outcomes do you want the pupils to achieve? (list 2 I

3b) What specific learning outcomes do you want the pupils to achieve? (list 2 I

3c) Write out two questions that you will ask to pupils to show that they have (...)

3c) Write out two questions that you will ask to pupils to show that they have (...)

3d) Write out a question that you will ask pupils to show that they can apply (...)

3e) What can or should the government do about road accidents?

3e) What can or should the government do about road accidents?

3e) What can or should the government do about road accidents?

3f) Why is it difficult?

3f) Why is it difficult?

3f) Why is it difficult?

3d) Write out a question that you will ask pupils to show that they can apply wh

4a. Strength1. short assessment of XXX's letter

4a. Strength2. Short assessment of XXX's letter

4a. Strength3. Short assessment of XXX's letter

4a. Weakness 1. Short assessment of XXX's letter

4a. Weakness 2. Short assessment of XXX's letter

4a. Weakness 3. Short assessment of XXX's letter

4b. Strength 1. Short assessment of YYY's letter

4b. Strength 2. Short assessment of YYY's letter

4b. Strength 3. Short assessment of YYY's letter

4b. Weakness 1. Short assessment of YYY's letter

4b. Weakness 2. Short assessment of YYY's letter

4b. Weakness 3. Short assessment of YYY's letter

4a. Strength4. Short assessment of XXX's letter

4a. Weakness 3. Short assessment of XXX's letter

4b. Strength 4. Short assessment of YYY's letter

swot_bw4
eval_a
eval_c1
eval_c2
eval_c3
c_name
c_age
c_gender
c_post
c_contract
c_fulltime
c_educ_level
c_educ_training
Id

4b. Weakness 4. Short assessment of YYY's letter
5a) What is the class average for English?
5c) Comment on these three learners (1)
5c) Comment on these three learners (2)
5c) Comment on these three learners (3)
Name (Consolidated)
Age (Consolidated)
Gender (Consolidated)
Position (Consolidated)
Contract (Consolidated)
Full-time/Part-time (Consolidated)
Education Level (Consolidated)
Education Training (Consolidated)
ID IRT

Child level

emis_code
m5s0_date_d
m5s0_date_m
m5s0_date_y
m5s0_teacher_n
m5s0_1st_enumerator_n
m5s0_1st_enumerator_c
m5s0_2nd_enumerator_n
m5s0_2nd_enumerator_c
m5s0_region_c
m5s0_region_n
m5s0_district_c
m5s0_district_n
m5s0_ward_c
m5s0_ward_n
m5s0_village_c
m5s0_village_n
m5s0_town_c^b
m5s0_fokontany_c^b
m5s0_school_n
m5s0_school_c
m5s0_teamleader_result
m5s0_supervisor_result
m5s0_enumerator_c
m5s0_county_n
m5s0_county_c
m5s0_division_n
m5s0_division_c

School EMIS Code
M5 Date (day)
M5 Date (month)
M5 Date (year)
M5 Teacher name
M5 First Visit Enumerator (name)
M5 First Visit Enumerator (code)
M5 Second Visit Enumerator (name)
M5 Second Visit Enumerator (code)
M5 Region Code
M5 Region Name
M5 District code
M5 District name
M5 Ward code
M5 Ward name
M5 Village/street code
M5 Village/street name
M5 Commune (town)_code
M5 Fokontany_code
M5 School name
M5 School code
M5 Team leader result
M5 Supervisor result
M5 Enumerator Name
M5 County name
M5 County code
M5 Division name
M5 Division code

m5s0_parish_n	M5 Parish name
m5s0_parish_c	M5 Parish code
m5s0_date_dmy	M5 Date (Day/Month/Year)
m5s0_province_n	M5 Province name
m5s0_IDEN_n	M5 IDEN Name
m5s1_first_name	M5 Pupil's First name
m5s1_age	M5 Age
m5s1_lang1_teacher_thisyr_n	M5 Name of your language 1 teacher this year
m5s1_math_teacher_thisyr_n	M5 Name of your Math teacher this year
m5s1_inschool_lastyr	M5 Were you in this school last year
m5s1_inschool_lastyr_oth ^b	M5 In which class were you last in last year
m5s1_lang1_teacher_lastyr_n	M5 Name of your language 1 teacher last year
m5s1_math_teacher_lastyr_n	M5 Name of your Math teacher last year
m5s1_lang1_teacher_lastyr_c ^b	M5 Code from module 2A
m5s1_math_teacher_lastyr_c	M5 Code of your Math teacher last year
m5s1_agreetoparticipate	M5 Do you agree to participate in this exercise
m5s1_time_starttest_hr	M5 Time started the test (HR)
m5s1_time_starttest_min	M5 Time started the test (MN)
m5s1_time_endtest_hr	M5 Time ended the test (HR)
m5s1_time_endtest_min	M5 Time ended the test (MN)
m5s1_name	M5 Pupil's name (first and last)
m5s1_lang2_teacher_lastyr_n	M5 Name of your language 2 teacher last year
m5s1_master_lastyr_n	M5 Name of your Master last year
m5s1_master_thisyr_n	M5 Name of your Master this year? (name)
m5s1_school_n	M5 School name
m5s1_teacher_CE2_n	M5 Name of teacher (CE2)
m5s1_teacher_CE1_n	M5 Name of teacher (CE1)
m5q12a	M5 District name
m5q13a	M5 County name
agree_participate	Do you agree to participate in this exercise
uniqueid	

Time on task

date_day	Date (Day)
date_month	Date (Month)
date_year	Date (Year)
date_comb	Date (Day/Month/Year)
observation_end	Time observed (Constructed)
comments	Module 4a Time on Task comments

Note: b. Additional variables anonymized for the 2016 Madagascar's Education SDI.

3.3 Recoding of variables

Some variables represent a risk of reidentification because of its composition and distribution. In order to share its contents, it is necessary to recode them in categories. The recoded variables cover both continuous (e.g. age, number of employees, etc.) and specific categorical variables in which certain categories are too scarce (e.g. position in the establishment, etc.). A detailed list of the recoded variables and the changes applied is listed in Table C.

The recoding consists in trimming tails of distribution or top recoding (e.g. year the school opened: every school opened before 1950 is grouped), transforming continuous values into ranges (e.g. using decades instead of years), and/or broadening categories to group possible answers (e.g. "Owner/Director/Head teacher/Principal/Deputy head teacher" into one value). Lastly, rather than grouping the whole distribution of values into categories, few variables' values are censored or anonymized when their specificity/uniqueness (e.g. excessively large values, rare values, etc.) might allow re-identification.

As part of the recoding process, we ensure that all identifiers have no value labels that could contain specific information that leads to the reidentification of schools, teachers or pupils.

Table C. Overview of recoded variables

Recoded Variable	Description	Recode
<u>School level</u>		
m1_resp_post	M1 Which Position do you occupy in this facility?	1-3 Owner/Director/Head teacher/Principal/Deputy head teacher as grouped in one category; others are kept the same
m1_school_type	M1 What's the school type	Suppress 3 schools with different classification to missing
m1_school_year	M1 When did this school begin operating	Years transformed into decades [1950-2010], all years below 1950 are grouped
m1_travelttime_govt	M1 Approximate traveling time from school to local govt education office (in min	Grouped in 30 min intervals, [0-360+], all above 360 are grouped
m1_days_in_session	M1 What was the actual number of days during which school was in session in the	Grouped in 10 days intervals, starting less than [100-250+], all below 100 are grouped and above 250 into another
m1_resp_age	M1 What is your age?	Transformed into decades [30s-60+], all below 30 are grouped and above 60 into another
m1_road_distance	M1 What is the distance between the school and the asphalt road?	Transformed in 3 categories: less than 100, 100-1000, more than 1000. Values like 99, 999, are considered missing values
m1_delegation_dist	M1 What is the distance in kilometers between school and delegation?	Transformed in 2 categories: 10 or less, more than 10
m1_water_time_dist	M1 Avg time to go and come back to main source of water (including avg waiting t	Transformed in 3 categories: less than 15, 16-60, 61+
m1_travelttime_govt_h	M1 Approximate traveling time from school to local govt education office (Hours)	Transformed in 3 categories: 0, 1-3, 4 or more
m1_travelttime_govt_m	M1 Approximate traveling time from school to local govt education office (Minute	Transformed in 4 categories: 0-15,16-30, 31-45, 46-60
m1_smt_meet ^c	M1 How many times did the SMT meet last year?	11 or more meetings were grouped together
m1_sc_meet_lastyr ^c	M1 How many times did the SC/BoD meet in the last year?	10 or more meetings were grouped together
m1_qao_n_visits ^c	M1 How many times did the school receive a visit from an official quality assurance officer?	20 or more visits were grouped together
m1_meeting_days ^c	M1 General (union) meetings	10 or more days were grouped together

m1_strike_days ^c	M1 Stikes (days)	10 or more days were grouped together
m1_happy_days ^c	M1 Happy events (marriage, baptisms, births, ritual receptions and Malagasy cults)	10 or more days were grouped together
	M1 Holidays (national holiday, religious holidays, etc.)	10 or more days were grouped together
m1_sad_days ^c	M1 Sad events (deaths, etc.)	10 or more days were grouped together
m2_n_teachers	M2 How many teachers work in this school?	Transformed in 5 categories: 0-10, 11-20, 21-30, 31-40, 41+
m2_n_non_teachers	M2 How many non-teaching staff work in this school?	Transformed into a dummy: 0, 1 or more
m4d_age	M4 Age	Transformed into decades [30s-60+], all below 30 are grouped and above 60 into another

School
management/finances
(M3)^c

m3saq1	M3 Respondent's position in the school	All observations from category " <i>autre membre de bureau</i> " were recoded as " <i>autres (a preciser)</i> ".
--------	--	--

Teacher level

m2a_post	M2a Position in the school	1-3 Owner/Director/Head teacher/Principal/Deputy head teacher as grouped in one category; others are kept the same
m2a_age	M2a Age	Transformed into decades [30s-60+], all below 30 are grouped and above 60 into another
m2a_educ_training ^c	M2a Highest level of teacher training (certificate) attained	Observations from category " <i>Bachelor of Education</i> " and " <i>CFEN or equivalent</i> " were recoded as " <i>Other - Specify</i> ".
m2a_educ_training_yr ^c	M2a Year education certificate obtained	Years transformed into decades [1980-2010], all years below 1980 are grouped together
m2a_start_teach ^c	M2a What year did you start teaching?	Years transformed into decades [1980-2010], all years below 1980 are grouped together
m2a_start_teach_this ^c	M2a What year did you start teaching at this school?	Years transformed into decades [1980-2010], all years below 1980 are grouped together

m2b_post	M2b Position in the school	1-3 Owner/Director/Head teacher/Principal/Deputy head teacher as grouped in one category; others are kept the same
m2b_start_teach	M2b Since what year have you been teaching?	Years transformed into decades [1980-2010], all years below 1980 are grouped together
m2b_age	M2b Age	Transformed into decades [30s-60+], all below 30 are grouped and above 60 into another
m2b_start_teach_this	M2b Since what year do you teach in this school?	Transformed into 2 categories: before 2010, 2010-2019
m6_age	M6 Age	Transformed into decades [30s-60+], all below 30 are grouped and above 60 into another
m6_years_teach	M6 Number of years teaching	less than 10, 10-19,20-29, 30+
m6_post	M6 What is your position at this school	1-3 Owner/Director/Head teacher/Principal/Deputy head teacher as grouped in one category; others are kept the same
c_age	Age (Consolidated)	Transformed into decades [30s-60+], all below 30 are grouped and above 60 into another
c_post	Position (Consolidated)	1-3 Owner/Director/Head teacher/Principal/Deputy head teacher as grouped in one category; others are kept the same

Note: c. Additional variables recoded for the 2016 Madagascar's Education SDI.