

Dataset Profile

<i>USAID Education Strategy Goal(s)</i>	Goal 1
<i>Country</i>	Ethiopia
<i>Activity</i>	Improving Quality of Primary Education Program (IQPEP)
<i>Dataset</i>	PUF Data_Ethiopia_IQPEP_Endline
<i>Assessment type</i>	EGRA
<i>Dataset Phase</i>	Endline

Acknowledgement: This document is made possible by the support of the American People through the United States Agency for International Development (USAID). The contents of this document are the sole responsibility of Optimal Solutions Group, LLC, and do not necessarily reflect the views of USAID or the United States Government.

Source: The information included in this document was derived from data and reports generated by external partners as part of a USAID contract(s). Portions of this document may include verbatim references from those documents.

For more information on the SART project or SART Ed, please contact the SART Helpdesk.

By Email: info@SARTDataCollection.org

By Phone: 1-844-AID-SART (+001-844-243-7278) or +001-301-289-7398

Table of Contents

1. Overview
2. Key variables
3. Complex Survey Specifications
4. Public Use Files (PUF) Dataset Codebook
5. Data Quality Assessment
6. Limitations
7. In-bound data file(s)

Appendix A: Glossary

Appendix B: Guidelines on Using Complex Survey Specifications

Overview

Calculating frequencies, modes, and means without survey-based software applications

Using Stata for analysis of data with complex survey specifications

1. Overview

This section provides an overview of the Public Use Files (PUF) dataset.

<i>USAID Education Strategy Goal(s)</i>	Goal 1
<i>Country</i>	Ethiopia
<i>Activity</i>	Improving Quality of Primary Education Program (IQPEP)
<i>Intervention</i>	Unaffiliated
<i>Cohort</i>	Unaffiliated
<i>Grade level(s)</i>	2, 3
<i>Age(s)</i>	1- more than 30
<i>Dataset type</i>	EGRA
<i>Assessment language(s)</i>	Afan Oromo, Amharic, Sidamigna, Somaligna, Tigrinya
<i>Phase</i>	Endline
<i>Start date</i>	05/01/2014
<i>End date</i>	05/31/2014
<i>Number of observations</i>	9406

2. Key variables

This section lists the availability of key variables in the dataset

Variable name	Variable definition	Availability in dataset
<i>country</i>	Country name	Yes
<i>activity</i>	Name of activity within the country?	Yes
<i>treat_phase</i>	In which phase of the study is this treatment school student?	Yes
<i>grade</i>	What is the student's grade level?	Yes
<i>female</i>	Is the student female?	Yes
<i>language</i>	Language the test was conducted in	Yes
<i>year</i>	In what year was the student tested?	Yes
<i>treatment</i>	What level of treatment is the school receiving?	Yes
<i>clpm</i>	Correct Letters Per Minute	Yes
<i>cwpm</i>	Correct Familiar Words Per Minute	Yes
<i>cnonwpm</i>	Correct Invented Words Per Minute	Yes
<i>orf</i>	Oral Reading Fluency	Yes
<i>read_comp_score</i>	Total correct Reading Comprehension questions	Yes
<i>eq_orf</i>	Equated Oral Reading Fluency	No
<i>eq_read_comp_score</i>	Equated Total correct Reading Comprehension questions	No
<i>clspm</i>	Correct Letter Sounds Per Minute	No
<i>csspm</i>	Correct Syllable Sounds Per Minute	No

3. Complex survey specifications

The survey specifications are outlined below. See Appendix B for detailed instructions on how to construct and use these weights. For detailed information about the activity sample design, see Section 2 in the Activity Profile.

<i>Complex survey specifications</i>	No
--------------------------------------	----

4. Public Use Files (PUF) Dataset Codebook

This section provides the codebook of the PUF dataset. In addition to standard codebook elements, this section lists the availability of each variable in the PUF dataset, and highlights whether the variable was constructed by SART.

Variable name	Variable Definition	Response category labels	Variable type	Notes	Availability in dataset	Constructed by SART
activity	Name of activity within the country?		String		Yes	Yes
age	How old is the student?		integer		Yes	No
class_size	How many students in the student's class?		integer		No	Not Reported
clpm	Correct Letters Per Minute		float		Yes	Yes
clspm	Correct Letter Sounds Per Minute		float		No	Not Reported
cnonwpm	Correct Invented Words Per Minute		float		Yes	Yes
cohort	Cohort within intervention		String		No	Not Reported
cot_attempted	Number of Concepts of Text questions attempted.		integer	[Range: 0-10]	No	Not Reported
cot_attempted_pct	Percentage of attempted questions correct.		float	[Range: 0-1]	No	Not Reported
cot_score	Total correct Concepts of Text questions.		integer		No	Not Reported
country	Country name		String		Yes	Yes
csspm	Correct Syllable Sounds Per Minute		float		No	Not Reported

Variable name	Variable Definition	Response category labels	Variable type	Notes	Availability in dataset	Constructed by SART
cwpm	Correct Familiar Words Per Minute		float		Yes	Yes
date	On what date was the student tested?		datetime	Date format: MMDDYYYY	No	Not Reported
district	Third largest geographic breakdown, which may reflect school districts		String		No	Not Reported
encrypted_school_code	Encrypted code to identify school's within country.		String		Yes	Yes
eq_oral_read_attempted	Equated Number of Oral Reading Word questions attempted.		float		No	Not Reported
eq_oral_read_attempted_pcmt	Equated Percentage of attempted questions correct.		float		No	Not Reported
eq_oral_read_score	Equated Total correct Oral Reading Word questions.		float		No	Not Reported
eq_oral_read_score_pcmt	Equated Percentage of Oral Reading Word questions correct.		float		No	Not Reported
eq_orf	Equated Oral Reading Fluency		float		No	Not Reported
eq_read_comp_attempted	Equated Number of reading comprehension questions attempted		float		No	Not Reported
eq_read_comp_attempted_pcmt	Equated percentage of reading comprehension questions attempted		float		No	Not Reported
eq_read_comp_score	Equated Total correct Reading comprehension questions		float		No	Not Reported

Variable name	Variable Definition	Response category labels	Variable type	Notes	Availability in dataset	Constructed by SART
eq_read_comp_score_pcnt	Equated percentage of correct Reading comprehension questions		float		No	Not Reported
fam_word_attempted	Number of Familiar Word questions attempted.		integer	[Range: 0-50]	No	Not Reported
fam_word_attempted_pcnt	Percentage of attempted questions correct.		float	[Range: 0-1]	No	Not Reported
fam_word_score	Total correct Familiar Word questions.		integer	[Range: 0-50]	Yes	Yes
fam_word_score_pcnt	Percentage of Familiar Word questions correct.		float	[Range: 0-1]	No	Not Reported
fam_word_time_remain	Time remaining when finished answering questions?		integer	Number of seconds remaining [Range: 0-60]	Yes	No
female	Is the student female?	Male; Female	String		Yes	No
final_weight	The final sampling weight (e.g., probability weight)		float		Yes	No
fpc1	Finite population correction for stage 1		float		No	Not Reported
fpc2	Finite population correction for stage 2		float		No	Not Reported
fpc3	Finite population correction for stage 3		float		No	Not Reported
fpc4	Finite population correction for stage 4		float		No	Not Reported

Variable name	Variable Definition	Response category labels	Variable type	Notes	Availability in dataset	Constructed by SART
fpc5	Finite population correction for stage 5		float		No	Not Reported
grade	What is the student's grade level?	First; Second; Third; Fourth; Fifth; Sixth; Seventh; Eighth	integer		Yes	No
intervention	Intervention within activity		String		No	Not Reported
invent_word_attempted	Number of Invented Word questions attempted.		integer	[Range: 0-50]	No	Not Reported
invent_word_attempted_pcmt	Percentage of attempted questions correct.		float	[Range: 0-1]	No	Not Reported
invent_word_score	Total correct Invented Word questions.		integer	[Range: 0-50]	Yes	Yes
invent_word_score_pcmt	Percentage of Invented Word questions correct.		float	[Range: 0-1]	No	Not Reported
invent_word_time_remain	Time remaining when finished answering questions?		String	Number of seconds remaining [Range: 0-60]	Yes	No
language	Language the test was conducted in		String		Yes	No
letter_attempted	Number of Letter Name questions attempted.		String	[Range: 0-100]	No	Not Reported
letter_attempted_pcmt	Percentage of attempted questions correct.		float	[Range: 0-1]	No	Not Reported
letter_score	Total correct Letter Name questions.		integer	[Range: 0-100]	Yes	Yes

Variable name	Variable Definition	Response category labels	Variable type	Notes	Availability in dataset	Constructed by SART
letter_score_pcmt	Percentage of Letter Name questions correct.		float	[Range: 0-1]	No	Not Reported
letter_sound_attempted	Number of Letter Sound questions attempted.		integer	[Range: 0-100]	No	Not Reported
letter_sound_attempted_pcmt	Percentage of attempted questions correct.		float	[Range: 0-1]	No	Not Reported
letter_sound_score	Total correct Letter Sound questions.		integer	[Range: 0-100]	No	Not Reported
letter_sound_score_pcmt	Percentage of Letter Sound questions correct.		float	[Range: 0-1]	No	Not Reported
letter_sound_time_remain	Time remaining when finished answering questions?		integer	Number of seconds remaining [Range: 0-60]	No	Not Reported
letter_time_remain	Time remaining when finished answering questions?		integer	Number of seconds remaining [Range: 0-60]	Yes	No
level	For non-traditionally aged students, what grade level are they learning at?	First; Second; Third; Fourth; Fifth; Sixth; Seventh; Eighth	integer		No	Not Reported
list_comp_attempted	Number of Listening Comprehension questions attempted.		integer	[Range: 0-10]	No	Not Reported
list_comp_attempted_pcmt	Percentage of attempted listening comprehension questions correct.		float	[Range: 0-1]	No	Not Reported
list_comp_score	Total correct Listening Comprehension questions. Percentage of Listening Comprehension questions correct.		integer	[Range: 0-10]	Yes	Yes
list_comp_score_pcmt	Percentage of Listening Comprehension questions correct.		float	[Range: 0-1]	Yes	Yes

Variable name	Variable Definition	Response category labels	Variable type	Notes	Availability in dataset	Constructed by SART
month	In what month was the student tested?	January; February; March; April; May; June; July; August; September; October; November; December	String		Yes	Yes
multigrade	Is the student's class a multiple-grade classroom?	No; Yes	String		No	Not Reported
oral_read_attempted	Number of Oral Reading Word questions attempted.		integer	[Range: 0-50]	No	Not Reported
oral_read_attempted_pcnt	Percentage of attempted questions correct.		float	[Range: 0-1]	No	Not Reported
oral_read_score	Total correct Oral Reading Word questions.		integer	[Range: 0-150]	Yes	Yes
oral_read_score_pcnt	Percentage of Oral Reading Word questions correct.		float	[Range: 0-1]	No	Not Reported
oral_read_time_remain	Time remaining when finished answering questions?		integer	Number of seconds remaining [Range: 0-60]	Yes	No
orf	Oral Reading Fluency		float		Yes	Yes
pa_diff_init_sound_attempted	Number of PA: Different Initial Sound questions attempted.		integer	[Range: 0-10]	No	Not Reported
pa_diff_init_sound_attempted_pcnt	Percentage of attempted PA: Different Initial Sound questions correct.		float	[Range: 0-1]	No	Not Reported
pa_diff_init_sound_score	Total correct PA: Different Initial Sound questions.		integer	[Range: 0-10]	No	Not Reported
pa_diff_init_sound_score_pcnt	Percentage of PA: Different Initial Sound questions correct.		float	[Range: 0-1]	No	Not Reported

Variable name	Variable Definition	Response category labels	Variable type	Notes	Availability in dataset	Constructed by SART
pa_init_sound_attempted	Number of PA: Initial Sound questions attempted.		integer	[Range: 0-10]	No	Not Reported
pa_init_sound_attempted_pcnt	Percentage of attempted PA: Initial Sound questions correct.		float	[Range: 0-1]	No	Not Reported
pa_init_sound_score	Total correct PA: Initial Sound questions.		integer	[Range: 0-10]	Yes	Yes
pa_init_sound_score_pcnt	Percentage of PA: Initial Sound questions correct.		float	[Range: 0-1]	Yes	Yes
pa_num_sound_attempted	Number of PA: Number of Sounds questions attempted.		integer	[Range: 0-10]	No	Not Reported
pa_num_sound_attempted_pcnt	Percentage of attempted PA: Number of Sounds questions correct.		float	[Range: 0-1]	No	Not Reported
pa_num_sound_score	Total correct PA: Number of Sounds questions.		integer	[Range: 0-10]	No	Not Reported
pa_num_sound_score_pcnt	Percentage of PA: Number of Sounds questions correct.		float	[Range: 0-1]	No	Not Reported
pa_phon_sound_attempted	Number of PA: Phonemes in Words questions attempted.		integer	[Range: 0-10]	No	Not Reported
pa_phon_sound_attempted_pcnt	Percentage of attempted PA: Phonemes in Words questions correct.		float	[Range: 0-1]	No	Not Reported
pa_phon_sound_score	Total correct PA: Phonemes in Words questions.		integer	[Range: 0-10]	No	Not Reported
pa_phon_sound_score_pcnt	Percentage of PA: Phonemes in Words questions correct.		float	[Range: 0-1]	No	Not Reported

Variable name	Variable Definition	Response category labels	Variable type	Notes	Availability in dataset	Constructed by SART
read_comp_attempted	Number of Reading Comprehension questions attempted.		integer	[Range: 0-10]	No	Not Reported
read_comp_attempted_pcnt	Percentage of attempted questions correct.		float	[Range: 0-1]	No	Not Reported
read_comp_score	Total correct Reading Comprehension questions.		integer	[Range: 0-10]	Yes	Yes
read_comp_score_pcnt	Percentage of Reading Comprehension questions correct.		float	[Range: 0-1]	Yes	Yes
region	The largest geographic breakdown of the activity		String		Yes	No
school_type	What type of school does the student attend?	Public; Private; Religious_Mission; Community; Combined; Regular; Youth; Public_Islamiyya; Monrovia; Mont_Barclay; Grant-Aided	String		Yes	No
state	The second largest geographic breakdown of the activity		String		No	Not Reported
strata1	Stratum identifiers for stage 1		String		No	Not Reported
strata2	Stratum identifiers for stage 2		String		No	Not Reported
strata3	Stratum identifiers for stage 3		String		No	Not Reported
strata4	Stratum identifiers for stage 4		String		No	Not Reported
strata5	Stratum identifiers for stage 5		String		No	Not Reported

Variable name	Variable Definition	Response category labels	Variable type	Notes	Availability in dataset	Constructed by SART
student_id	Unique student ID number.		String	Must be unique	Yes	No
su1	Sampling units for stage 1		float		No	Not Reported
su2	Sampling units for stage 2		float		No	Not Reported
su3	Sampling units for stage 3		float		No	Not Reported
su4	Sampling units for stage 4		float		No	Not Reported
su5	Sampling units for stage 5		float		No	Not Reported
syllable_sound_attempted	Number of Syllable Sound questions attempted.		integer	[Range: 0-100]	No	Not Reported
syllable_sound_attempted_pcmt	Percentage of attempted Syllable Sound questions correct.		float	[Range: 0-1]	No	Not Reported
syllable_sound_score	Total correct Syllable Sound questions.		integer	[Range: 0-100]	No	Not Reported
syllable_sound_score_pcmt	Percentage of Syllable Sound questions correct.		float	[Range: 0-1]	No	Not Reported
syllable_sound_time_remain	Time remaining when Syllable Sound questions finished answering?		integer	Number of seconds remaining [Range: 0-60]	No	Not Reported
treat_phase	In which phase of the study is this treatment school student?	Baseline; Midline 1; Midline 2; Midline 3; Midline 4; Endline	String		Yes	Yes

Variable name	Variable Definition	Response category labels	Variable type	Notes	Availability in dataset	Constructed by SART
treatment	What level of treatment is the school receiving?	Control; Partial Treatment; Full Treatment	String		Yes	Yes
unt_oral_read_attempted	Number of Untimed Oral Reading Word questions attempted.		integer	[Range: 0-150]	No	Not Reported
unt_oral_read_attempted_pct	Percentage of attempted untimed oral reading words correct.		float	[Range: 0-1]	No	Not Reported
unt_oral_read_score	untimed oral reading		integer	[Range: 0-150]	No	Not Reported
unt_oral_read_score_pct	Percentage of Untimed Oral Reading Word questions correct.		float	[Range: 0-1]	No	Not Reported
unt_oral_read_time_remain	Time remaining when finished answering questions?		integer	[Range: 0-60]	No	Not Reported
unt_orf	Untimed Oral Reading Fluency		float		No	Not Reported
unt_read_comp_attempted	Number of Untimed Reading Comprehension questions attempted.		integer	[Range: 0-10]	No	Not Reported
unt_read_comp_attempted_pct	Percentage of attempted Untimed Reading Comprehension questions correct.		float	[Range: 0-1]	No	Not Reported
unt_read_comp_score	Total correct Untimed Reading Comprehension questions.		integer	[Range: 0-10]	No	Not Reported
unt_read_comp_score_pct	Percentage of Untimed Reading Comprehension questions correct.		float	[Range: 0-1]	No	Not Reported
urban	Is the school in an urban area?	Rural; Urban	String		No	Not Reported

Variable name	Variable Definition	Response category labels	Variable type	Notes	Availability in dataset	Constructed by SART
vocab_word_attempted	Number of Vocabulary Word questions attempted.		integer	[Range: 0-20]	No	Not Reported
vocab_word_score	Total correct Vocabulary Word questions.		integer		No	Not Reported
wt1	Weight for stage 1		float		No	Not Reported
wt2	Weight for stage 2		float		No	Not Reported
wt3	Weight for stage 3		float		No	Not Reported
wt4	Weight for stage 4		float		No	Not Reported
wt5	Weight for stage 5		float		No	Not Reported
year	In what year was the student tested?		integer		Yes	Yes

5. Data Quality Assessment

This section summarizes the quality assessment of the PUF dataset.

	Answer
Validity. Was there a sufficient evidence that the outcome measure(s) were collected in the same manner for the intervention or treatment and comparison groups?	Yes
Validity. Was test equating performed (if applicable)?	Not reported
Reliability. Was inter-rater reliability procedures conducted?	Yes

6. Limitations

This section summarizes all documented anomalies or limitations of the PUF dataset.

In the original datasets received from the implementing partner, observations for the time remaining variable included both missing (i.e., null) and non-null values.

In the endline dataset, Optimal constructed the following sub-task score variables by imputing the missing values in the time remaining variable to zero:

oral reading fluency (orf);
correct letters per minute (clpm);
correct words per minute (cwpm); and
correct non-sense words per minute (cnonwpm).

Based on guidance from USAID/W, Optimal imputed the missing values to zero in order to align the sub-task score variables in the endline dataset to the mean sub-task scores presented in the evaluation report.

Should you wish to construct these sub-task score variables without imputing the missing values for the time remaining variables, please review the variable definition of each sub-task score in Section 4 of the Public Use Files (PUF) Dataset Codebook of the data profile.

Separately, there is data mismatch for the sub-task scores for the Benshangul Gumuz region. For this region, the mean scores reported in the evaluation report could not be replicated using the data present in the endline dataset.

7. In-bound data file(s)

This section highlights specific structural components of the original in-bound data file(s) that were used to construct the PUF dataset.

<i>Title</i>	Control and IQPEP Data Merged for second time size of 9406_Labels.dta
<i>Status</i>	Active
<i>Codebook received with in-bound data file(s)</i>	Yes
<i>Language of the in-bound data file(s)</i>	English
<i>File format of the in-bound data file(s)</i>	dta
<i>Structure of the in-bound data file(s)</i>	Cross-sectional
<i>Shape of the in-bound data file(s)</i>	Long
<i>Implementing partner</i>	FHI 360
<i>Implementing partner point-of-contact</i>	Charles Gale (FHI360)

Appendix A: Glossary

Appending files	When two or more files are joined together vertically, i.e., adding additional rows (observations) to a fixed number of columns (variables) in the data table.
Cohort	A group of students who are part of a specific intervention but are distinct from another group (e.g., different time of entry into the activity). The cohort distinction is generally made by the activity implementer and/or evaluator.
Finite population correction (fpc)	The finite population correction (FPC) factor is used to adjust the standard error of a sample mean when sampling is done without replacement and the sample size is at least 5 percent of the population. The FPC factor always has a value of less than one
Implementing partner	Organization involved in facilitating a USAID funded activity
In-bound file	File received from a third party organization including USAID/W, Missions, and implementing partners
Intervention	Students within an activity who are part of the same reading implementation methodology
Public Use files (PUF)	Standardized assessment data file
Merging files	When two or more files are joined together horizontally using a single id variable, i.e., adding more columns (variables) fixed number of rows (observations) in the data table
Phase	Period/cycle of data collection (e.g., baseline, midline, or endline)
POC	Point of contact
Variable status: Collected	A variable that was collected during the assessment testing
Variable status: Constructed by SART	A variable that was constructed by SART using other variables collected during assessment testing
Variable status: Missing	A variable that was not provided to SART but there is evidence that it was collected during assessment testing.
Variable status: Not collected	A variable that was not collected as it was not part of assessment testing



USAID
FROM THE AMERICAN PEOPLE