

UWEZO

SAMPLE DESIGN FOR 2011 TANZANIA LEARNING ASSESSMEN SURVEY

**TANZANIA
NOVEMBER, 2010**

This document outlines the recommendations for the sample design of the 2011 Tanzania Learning Assessment Survey and the corresponding selection procedures. The design produces estimates for district/ male/female/Tanzania Mainland domains.

Table of Contents

Section 1: Introduction	3
1.1 Background	3
Section 2: Definitions of Terms	3
2.1 A household can be defined:	3
2.2 Head of Household:.....	4
2.3 Household or a Dwelling Unit	4
Section 3: Sample Design	5
3.1 Objectives of the Design	5
3.2 Overall Sampling Design	5
3.3 Target Population	6
3.4 Sampling Frame	6
3.5 Analytic Objectives.....	6
3.6 Stratification	7
3.7 Sample Allocation	7
3.8 Sample size	8
3.9 Sample selection	12
Section 4. Estimation procedure.....	12
4.1 Inclusion Probability	12
4.2 Weighting.....	13
Section 5: Sample size calculation	14
<i>Appendix</i>	15
References	16
Table 1: Geographical Sampling Errors for Selected Education Indicators	7
Table 2: Distribution of Selected EAs and Households by District and Region.....	9

SAMPLE DESIGN

Section 1: Introduction

1.1 Background

The Tanzania Learning Assessment Survey was conducted in Tanzania in May 2010 by UWEZO. The main objectives of this survey were to provide estimates on the quality of learning in schools by assessing the *basic literacy* and **numeracy among children aged 5 – 16 years** and other educational indicators associated with child learning in Tanzania. Results of round two TLAS survey will enable better analysis of the *basic literacy* and numeracy among children aged 5 – 16 years and other educational indicators in 133 districts of Tanzania Mainland (*Appendix 1*).

This document outlines the recommendations for the sample design for the 2011 UWEZO - Learning Assessment Survey (2011 TLAS) and the corresponding selection procedures. Uwezo, meaning “capability in Kiswahili, is a four years initiative to assess the quality of learning in schools by assessing the basic literacy and numeracy among children aged 5 – 16 years in Tanzania. The first assessment involved 42,033 children who were housed in 22,800 households with an average of 2 children interviewed per household. The 2011 TLAS is designed to give estimates for district/male/female/Tanzania Mainland domains. Some of the districts are purely rural (e.g Chato) while others are purely urban (e.g. Kinondoni), therefore Rural/Urban domain will only be feasible at national level.

Section 2 of this report provides the definitions of terms while Section 3 outlines the detail review of the sample design, sample selection, Estimation procedures and sample size calculation.

Section 2: Definitions of Terms

2.1 A household can be defined:

- (a) **A household** is a basic measure in this survey, (only normal members of household to be concerned). A household may be one-person or multi-person-household.
- (b) **One person household** is a person who lives alone in whole or part of a housing unit and has independent consumption.

- (c) **Multi-person household** is a group of two or more persons who occupy the whole or part of a housing unit and share their consumption. Usual households of this type contain husband, wife and children. Other relatives, borders, visitors and their persons are included as members of the household if they pool their resources, share their consumption and have been living with the household for at least two weeks.

Therefore, following the above given definitions i.e. item a) to c), it can be stated that a household is formed when the members of the household share the consumption by pooling their resources together¹.

2.2 Head of Household:

Head of household will mainly be used as a reference point, i.e. a person recognized as such by the rest of the members of the household to be the head. Often s/he is responsible for the financial support and welfare of the household members.

2.3 Household or a Dwelling Unit

This is all the living space occupied by one household regardless of the physical arrangement of facilities available. It may be one room occupied by lodgers or it may be one, two or more housing units occupied by an extended family or household.

¹ Adapted from 1991, 2000/01 and 2007 Household Budget Survey

Section 3: Sample Design

3.1 Objectives of the Design

The 2011 UWEZO - TLAS is designed to allow reliable estimation of most education variables for the following domains of interest:

- i. Tanzania Mainland as a whole
- ii. Urban and rural areas (only at national level)
- iii. Gender (male/female)
- iv. District level
- v. Regional level

The primary objective of the 2011 TLAS is to provide estimates with acceptable precision for selected key education indicators such as literacy status on the above domain of interest). For instance the maximum required sample size to cover all literate statuses of “0.313”² and a relative error margin of 8%, the sample size is about 1012 persons per domain, which holds for each district, urban/rural, male and female domains. Objectively, this design is an equal allocation of clusters to allow for comparison across districts.

3.2 Overall Sampling Design

The design for the 2011 TLAS is a Probability Proportion to Size sample of households within clusters/EAs from each district of the Mainland Tanzania. First, the 2002 Population and Housing frame was stratified into 133 districts, which are defined as the cross-stratification of the following variables:

- Urban/Rural
- Number of persons by Village/EA
- Gender (Male, Female) by Village/EA
- Number of households by Village/EA

Once the frame has been stratified, the number of districts, EAs, households and personnel in each district were determined. The identified domain groups were used to ensure the sample

2 2007-08 HIV/AIDS and Malaria Indicator Survey

allocation and total sample size meets the precision requirements. Obtaining the required sample size in the field would require 100 percent response. Since this will not be the case, the total response rate was inflated to account for the non-response and the design effect.

3.3 Target Population

The target population to be covered by the 2011 TLAS is defined as the universe of the population aged 5 - 16 in Mainland Tanzania. Basing on the results of round one survey, the second round expects to enumerate about 159,600 children aged 5 – 16 for the whole Mainland Tanzania (about 1,200 children per district).

3.4 Sampling Frame

Administratively, Tanzania currently has 21 administrative functioning regions in the mainland. In turn, each administrative region is subdivided into districts and each district into wards. In addition to these administrative units, during the 2002 Population and Housing Census (PHC), each ward was further subdivided into census enumeration areas (EAs). Each EA is either totally urban or rural. The EAs has census information on the number of households and population by gender. The available demarcated cartographic maps for each EA in the 2002 Population and Housing Census made the census the optimum sampling frame for the 2011 TLAS. The frame for the first stage of selection was a list of Enumeration Areas containing private and institutional households. Using this distribution of households, all institutional households from all institutional EAs were excluded. The frame for the second stage of selection was the updated list of households from the listing exercise from each selected EA.

3.5 Analytic Objectives

UWEZO TEN/MET is interested in using the data collected from the 2011 TLAS to measure differences across specific domains. Specifically, the domains are:

- Tanzania Mainland
- District level
- Regional level

- Urban and rural areas (national only)
- Gender (Male, Female)

The precision targets were obtained from 2007-08 Tanzania HIV/AIDS and Malaria Indicator Survey as summarized in Table 1.

Table 1: Geographical Sampling Errors for Selected Education Indicators

Estimate	Indicatory	Urban Sample		Rural Sample		Mainland Sample	
		Women	Men	Women	Men	Women	Men
R (Percentage)	No Education	0.092	0.049	0.255	0.142	0.213	0.241
	Secondary and above	0.192	0.313	0.046	0.077	0.084	0.12
Standard Error	No Education	0.011	0.009	0.011	0.008	0.01	0.007
	Secondary and above	0.016	0.021	0.004	0.006	0.006	0.008
Relative Error	No Education	0.12	0.18	0.04	0.06	0.05	0.03
	Secondary and above	0.08	0.07	0.09	0.08	0.07	0.07

Source: 2007-08 HIV/AIDS and Malaria Indicator Survey

3.6 Stratification

In the 2002 PHC frame, EAs were grouped by districts. Therefore such stratification by urban and rural areas is also reflected in the 2011 TLAS sample. In addition, the sample is also reflected in round two domains as mentioned in sub-section 3.1 above. Stratification was by region and district. Enumeration areas were stratified according to:

- Region in 21 strata
- District (133)
- Urban/Rural
- Male/Female
- Number of households

3.7 Sample Allocation

The allocation of EAs was proportional to the number of households, adjusted to provide sufficient precision of estimates at the district level. To provide optimal sample sizes in each district we decided that

the minimum number of allocated EAs to each stratum should be 30.

The primary sampling unit (PSU) or “cluster/EA” for the 2011 TLAS was defined on the basis of EAs from the 2002 PHC frame, as having one (or for a few cases, more than one) EA per cluster. The number of clusters/EAs selected for each district is 30.

For this study, the target was to have completed interviews with 1,200 school children age 5-16 per district. The recommended sample of about 30 EAs per district selected (20 household per EA and 600 households per district) will provide about 159,600 interviews of school children age 5-16 nationally with the acceptable precisions. Table 2 below shows the distribution of selected EAs and households.

Under this allocation, a minimum of about 600 households completed interviews is expected in each district of the 133 districts. The distribution of 30 EAs by district was selected by PPS sampling while household sample distribution is by systematic random sampling, hence the sampling design for the study is an equal probability selection method sample (EPSEM).

3.8 Sample size

The sample size for 2011 LAS was 79,800 households. The number of selected EAs was 3,990. Within these EAs 20 Households were selected. This sample size was determined according to precision of estimates (from previous related household study, a design effect of 1.82 and an expected non-response rate of around 5%. With the planned sample size and expected non-response rate, we expected to achieve an effective sample size of 600 households per district.

A review of sample allocation is given in Table 2. This sample distribution should provide precision of estimations for different type of assessments (district/urban/rural and male/female) at the district/regional/national level. However, the tabulation of survey results for urban/rural domain within district/region is not recommended.

Table 2: Distribution of Selected EAs and Households by District and Region

Region Name	District Code	District Name	Selected EAs	Number of Households Per EA	Number of Households Per District
Dodoma	01	Kondoa	30	20	600
Dodoma	02	Mpwapwa	30	20	600
Dodoma	03	Kongwa	30	20	600
Dodoma	04	Dodoma Rural	30	20	600
Dodoma	05	Dodoma Urban	30	20	600
Dodoma	06	Bahi (Splited from Dodoma Rural)	30	20	600
Dodoma	07	Chamwino (Splited from Dodoma Rural)	30	20	600
Dodoma Total Districts	7		210	20	4,200
Arusha	01	Monduli	30	20	600
Arusha	02	Arumeru	30	20	600
Arusha	03	Arusha	30	20	600
Arusha	04	Karatu	30	20	600
Arusha	05	Ngorongoro	30	20	600
Arusha	06	Longido (Splited from Monduli)	30	20	600
Arusha	07	Meru (Splited from Arumeru)	30	20	600
Arusha	08	Arusha (R) (from Arumeru)	30	20	600
Arusha Total Districts	8		240	20	4,800
Kilimanjaro	01	Rombo	30	20	600
Kilimanjaro	02	Mwanga	30	20	600
Kilimanjaro	03	Same	30	20	600
Kilimanjaro	04	Moshi Rural	30	20	600
Kilimanjaro	05	Hai	30	20	600
Kilimanjaro	06	Moshi Urban	30	20	600
Kilimanjaro	07	Siha (Splited from Hai)	30	20	600
Kilimanjaro Total Districts	7		210	20	4,200
Tanga	01	Lushoto	30	20	600
Tanga	02	Korogwe	30	20	600
Tanga	03	Muheza	30	20	600
Tanga	04	Tanga	30	20	600
Tanga	05	Pangani	30	20	600
Tanga	06	Handeni	30	20	600
Tanga	07	Kilindi	30	20	600
Tanga	08	Mkinga (Splited from Muheza)	30	20	600
Tanga Total Districts	8		240	20	4,800
Morogoro	01	Kilosa	30	20	600
Morogoro	02	Morogoro	30	20	600
Morogoro	03	Kilombero	30	20	600
Morogoro	04	Ulanga	30	20	600
Morogoro	05	Morogoro Urban	30	20	600
Morogoro	06	Mvomero	30	20	600
Morogoro Total Districts	6		180	20	3,600
Pwani	01	Bagamoyo	30	20	600
Pwani	02	Kibaha R - Kibaha	30	20	600
Pwani	03	Kisarawe	30	20	600
Pwani	04	Mkuranga	30	20	600
Pwani	05	Rufiji	30	20	600

Pwani	06	Mafia	30	20	600
Pwani	07	Kibaha U - Kibaha	30	20	600
Pwani Total Districts	7		210	20	4,200
Dar es Salaam	01	Kinondoni	30	20	600
Dar es salaam	02	Ilala	30	20	600
Dar es Salaam	03	Temeke	30	20	600
Dar es Salaam Total Districts	3		90	20	1,800
Lindi	01	Kilwa	30	20	600
Lindi	02	Lindi Rural	30	20	600
Lindi	03	Nachingwea	30	20	600
Lindi	04	Liwale	30	20	600
Lindi	05	Ruangwa	30	20	600
Lindi	06	Lindi Urban	30	20	600
Lindi Total Districts	6		180	20	3,600
Mtwara	01	Mtwara Rural	30	20	600
Mtwara	02	Newala	30	20	600
Mtwara	03	Masasi	30	20	600
Mtwara	04	Tandahimba	30	20	600
Mtwara	05	Mtwara Urban	30	20	600
Mtwara	06	Nanyumbu (Splited from Masasi)	30	20	600
Mtwara Total Districts	6		180	20	3,600
Ruvuma	01	Tunduru	30	20	600
Ruvuma	02	Songea Rural	30	20	600
Ruvuma	03	Mbinga	30	20	600
Ruvuma	04	Songea Urban	30	20	600
Ruvuma	05	Namtumbo	30	20	600
Ruvuma Total Districts	5		150	20	3,000
Iringa	01	Iringa Rural	30	20	600
Iringa	02	Mufindi	30	20	600
Iringa	03	Makete	30	20	600
Iringa	04	Njombe	30	20	600
Iringa	05	Ludewa	30	20	600
Iringa	06	Iringa Urban	30	20	600
Iringa	07	Kilolo	30	20	600
Iringa	08	Njombe Mjini	30	20	600
Iringa Total Districts	8		240	20	4,800
Mbeya	01	Chunya	30	20	600
Mbeya	02	Mbeya (R)	30	20	600
Mbeya	03	Kyela	30	20	600
Mbeya	04	Rungwe	30	20	600
Mbeya	05	Ileje	30	20	600
Mbeya	06	Mbozi	30	20	600
Mbeya	07	Mbarali	30	20	600
Mbeya	08	Mbeya Urban	30	20	600
Mbeya Total Districts	8		240	20	4,800
Singida	01	Iramba	30	20	600
Singida	02	Singida Rural	30	20	600
Singida	03	Manyoni	30	20	600

Singida	04	Singida Urban	30	20	600
Singida Total Districts	4		120	20	2,400
Tabora	01	Nzega	30	20	600
Tabora	02	Igunga	30	20	600
Tabora	03	Uyui	30	20	600
Tabora	04	Urambo	30	20	600
Tabora	05	Sikonge	30	20	600
Tabora	06	Tabora Urban	30	20	600
Tabora Total Districts	6		180	20	3,600
Rukwa	01	Mpanda	30	20	600
Rukwa	02	Sumbawanga Rural	30	20	600
Rukwa	03	Nkasi	30	20	600
Rukwa	04	Sumbawanga Urban	30	20	600
Rukwa	05	Mpanda Mjini (From Mpanda)	30	20	600
Rukwa Total Districts	5		150	20	3,000
Kigoma	01	Kibondo	30	20	600
Kigoma	02	Kasulu	30	20	600
Kigoma	03	Kigoma Rural	30	20	600
Kigoma	04	Kigoma Urban	30	20	600
Kigoma Total Districts	4		120	20	2,400
Shinyanga	01	Bariadi	30	20	600
Shinyanga	02	Maswa	30	20	600
Shinyanga	03	Shinyanga Rural	30	20	600
Shinyanga	04	Kahama	30	20	600
Shinyanga	05	Bukombe	30	20	600
Shinyanga	06	Meatu	30	20	600
Shinyanga	07	Shinyanga Urban	30	20	600
Shinyanga	08	Kishapu	30	20	600
Shinyanga Total Districts	8		240	20	4,800
Kagera	01	Karagwe	30	20	600
Kagera	02	Bukoba Rural	30	20	600
Kagera	03	Muleba	30	20	600
Kagera	04	Biharamulo	30	20	600
Kagera	05	Ngara	30	20	600
Kagera	06	Bukoba Urban	30	20	600
Kagera	07	Chato (Splited from Biharamulo)	30	20	600
Kagera	08	Misenyi - Bukoba Rural	30	20	600
Kagera Total Districts	8		240	20	4,800
Mwanza	01	Ukerewe	30	20	600
Mwanza	02	Magu	30	20	600
Mwanza	03	Nyamagana	30	20	600
Mwanza	04	Kwimba	30	20	600
Mwanza	05	Sengerema	30	20	600
Mwanza	06	Geita	30	20	600
Mwanza	07	Missungwi	30	20	600
Mwanza	08	Ilemela	30	20	600
Mwanza Total Districts	8		240	20	4,800
Mara	01	Tarime	30	20	600
Mara	02	Serengeti	30	20	600
Mara	03	Musoma Rural	30	20	600
Mara	04	Bunda	30	20	600

Mara	05	Musoma Urban	30	20	600
Mara	06	Rorya (Splited from Tarime)	30	20	600
Mara Total Districts	6		180	20	3,600
Manyara	01	Babati	30	20	600
Manyara	02	Hanang	30	20	600
Manyara	03	Mbulu	30	20	600
Manyara	04	Simanjiro	30	20	600
Manyara	05	Kiteto	30	20	600
Manyara Total Districts	5		150	20	3,000
Grand Total Nationally	133		3,990		79,800

3.9 Sample selection

A two stage stratified sample was used. The first stage units were EAs, the second stage units were households. Before enumeration for the main survey, all selected EAs will be updated through listing.

1st stage: EA selection

EAs were selected systematically with Probability Proportional to Size (PPS) within each stratum (district) from the list of EAs. The measure of size of each EA was the number of households from the 2002 PHC. EAs were sorted within each district/stratum according to their geographical serial numbers. Using systematic selection on the sorted list, a high level of implicit geographical stratification and effective sample distribution was provided. A total of 30 EAs were selected from each district/stratum.

2nd stage: Households Selection from Selected EAs

The households were chosen randomly by systematic sampling method. A total of 20 households were selected from each selected EAs.

Section 4. Estimation procedure

4.1 Inclusion Probability

In order to obtain representative estimates for population, it is necessary to attach to each household a weight. Weights consist of two factors: initial weight as result of sampling design and a correction factor for non-response.

The initial weight for each household is equal to the inverse of inclusion probability (this

inclusion probability is a product of inclusion probabilities from the two stages). The Primary Sampling Units (PSUs) were EAs from the 2002 PHC while the Secondary Sampling Units (SSU) were selection of 20 households from each selected 30 EAs in a district.

Total household inclusion probability is:

$$P_{hi} = \frac{b_{hi}}{B_{hi}} \times \frac{m_{hi}}{M_{hi}}$$

Where

P_{hi} = Total inclusion probability of a household selected in i^{th} EA, in h^{th} district/stratum;

B_{hi} = Total number of households in a selected EA

$b_{hi} = 20$, is the number of selected households in the sample from the updated list of households in i^{th} EA in h^{th} district/stratum;

$m_{hi}=30$, is the number of selected EAs in the sample from the h^{th} district/stratum;

M_{hi} is the total number of EAs in the h^{th} district/stratum;

Two probability components correspond to two stages of the sample selection.

4.2 Weighting

Initial design weight is equal to inverse of the inclusion probability such that:

$$W_{hi} = \frac{B_{hi}}{b_{hi}} \times \frac{M_{hi}}{m_{hi}}$$

Where:

W_{hi} is the initial design weight for a household in in i^{th} EA, in h^{th} district/stratum

After data collection the initial weight would be adjusted for non-response using the following formula:

$$W_{hiAdj} = W_{hi} \times \frac{Z_{projec}}{\hat{Y}_{hi}}$$

W_{hiAdj} = adjusted weight for household in i^{th} EA, in h^{th} district/stratum

Z_{projec} = Projected district population for the survey year

\hat{Y}_{hi} = estimated population from the survey

In order to analyze and publish the 2011 LAS results, it is necessary to measure standard error of estimates and determine confidence intervals. There is specialized software for estimations of parameters and their errors based on complex sample data (SPSS, WesVar and STATA).

Section 5: Sample size calculation

The formula used for sample size calculation for the 2011 TLAS is :

$$n = z^2 \times (p \times q) \times \frac{DEFF^2}{e^2}$$

where:

p is the prevalence value of the indicator

q is $(1-p)$

$DEFF^2$ is an average value through values for estimates in a typical THIS survey, to be about 1.8125

e is the relative error (defined as standard error of indicator over estimated indicator value) desired.

Appendix 1: Selected EAs by Ward, District and Region

References

Cochran, W.G. (1977), "*Sampling Techniques*", 3rd edition. New York: J.W. Wiley and Sons, Inc.

Kalton, G. (1983), "*Introduction to Survey Sampling*", Beverly Hills, CA: Sage Publications.

Kish, L. (1965), "*Survey Sampling*", New York: J.W. Wiley and Sons, Inc.

Lohr, Sharon L. (1999), "*Sampling: Design and Analysis*", Pacific Grove, CA: Duxbury Press.

Pfeffermann, D (1996), "The Use of Sampling Weights for Survey Data Analysis", *Statistical Methods in Medical Research*, 1996: 5: 239-261

Särndal, C-E. Swenñson, B. and Wretman, J. (1991), "*Model Assisted Survey Sampling*", New York: Springer-Verlag.