EDIS
Economic Development
Intelligence System
DATA · MODELS · BETTER DECISIONS

Quality education today, a more prosperous Africa tomorrow

Education Modelling

Data · Models · Better Learning Outcomes

Evidence-based modelling for equitable, quality and inclusive education systems across Africa

Integrated Education Data
Household surveys, EMIS, administrative data

Advanced Modelling
CGE, microsimulation, spatial analysis

AI-Powered Insights
Scenarios, forecasts, policy options

Decision Support Tools
For governments, partners and planners

Capacity Building
Training, manuals, academic support

EDUCATION FUELS INCLUSIVE GROWTH

COURSE	EDU 501 — Education Sector Modelling
SCHOOL	05 Sector Planning — Education
DOCUMENT	Course Technical Manual
ESTATE	EDIS Education Sector Decision Intelligence System (IntegratedEducationSystemModel)
SPECIFICATION	EDIS Education Modelling Technical Manual, EDIS, 2026
CATALOGUE	72 model cards in 8 families
ESTATE SIZE	90 Julia modules, 11,183 lines
EVIDENCE	Uganda 2016 DHS (UG2016DHS); Kenya 2022 DHS (KE2022DHS); Uganda LSMS-ISA National Panel Survey 2019/20 (UNPS)
SPATIAL	Uganda school register (EMIS), via the EDIS spatial data infrastructure; EDIS Africa education GIS database (AEGDB)
HOURS	55
VERSION	1.0
BUILT	2026-08-27
EXTRACT DIGEST	540f7c8b5d9e3695

DATA AND SOURCING STATEMENT

This manual teaches an education-sector model estate and the judgement required to use it. Every number in it was produced by running the estate's own specifications against real household surveys, a real school register and real geography, and every one of them carries its denominator, its interval and the name of the file it came from.

Four of the catalogue's families are estimated here on microdata. One is not, because the data that would estimate it do not exist on any platform this course can reach, and the manual says so in the chapter rather than in a footnote.

Contents

Preface — what this manual is, and what it is not

This is a manual for an implemented system. Every chapter that follows describes something that exists at ‘D:/EDIP/models/EducationModelling‘: 90 Julia modules across 30 layers, 11,183 lines, with 6 committed runs, a versioned contract to a computable general equilibrium model, and a documented set of things it cannot yet do.

Three things this manual does that a specification does not

- **It estimates.** The specification defines 72 models. This manual runs a representative set of them against real microdata — two national household surveys and a household panel with a consumption aggregate and a labour module — and reports what came out, including where the result was uncomfortable.
- **It reconciles.** Four sources describe primary education in the same country and they disagree. Chapter 26 puts the four numbers on one page and explains each difference, because a planner who cannot say why two figures differ cannot use either.
- **It refuses.** Ten of the seventy-two models cannot be estimated from anything available, and this manual says so in the chapter that specifies them. A model presented as estimated when it was not is the single most damaging thing an analyst can put in front of a minister.

How to read a number in this manual

Label	What it means	What you may do with it
Estimated	Computed here from microdata, with a design-based interval, an unweighted denominator and a disclosure decision.	Quote it, with all three companions.
Register	Read from an administrative source — the school register, the budget, a facility list.	Quote it as an administrative count, with its coverage. It is not a survey estimate and it has no sampling interval.
Projected	Output of the system engine under stated parameters.	Quote it with the parameters and a sensitivity range. It is not a measurement.
Specified	A model the catalogue defines that this course does not estimate.	Teach it. Do not present its output, because there is none.

THE RULE THIS MANUAL IS BUILT ON
Every number has a model behind it and a denominator under it. Two of the seven automatic fails in this course are denominator failures, and one of them was found in a shipped audit of this very estate — the arithmetic was right and the label named the wrong denominator. That is the failure mode hardest to catch and easiest to commit.

What the course does not claim

- It does not distribute microdata. Every survey used here is held under the terms of the body that collected it, and what ships with this course is the estimate, never the record.
- It does not validate against published national statistics. Where a figure here differs from a national report, this manual says the two differ and why they might; it does not assert which is right.
- It does not execute the CGE coupling. The contract is real and tested; the solve runs on a host where the macro model's dependencies load, and this is not that host.
- It does not teach a second CGE, a second survey-processing stack or a second GIS. It teaches the ones the estate has.

Provenance

Artefact	Produced by	From
‘education_tables.json‘	‘education_extract.py‘	the estate: manual, source, docs, runs, config, data platforms

Artefact	Produced by	From
‘education_estimates.json‘	‘education_analysis.py‘	the microdata, the school register and the education GIS database
38 figures	‘education_figures.py‘	the two files above
This manual and every deck	‘build_edu501.py‘	the two files above and the figures

Extract digest ‘540f7c8b5d9e3695‘. Re-running the chain regenerates every number in the course, so the course cannot drift from the estate.

How this manual is organised

Part	Chapters	What it covers
I — The estate and the evidence	1–4	What the system models, the specification it implements, the data it runs on, and how to estimate anything from a complex survey
II — The catalogue, family by family	5–12	All 72 model cards in 8 families, with the estimates this course produced for each
III — The system engine	13–17	Population, grade flow, supply and teachers, costing, and the scenario engine
IV — Geography and optimisation	18–21	The spatial infrastructure, accessibility, the nine mathematical programmes, and the capital decision
V — Integration	22–26	The CGE exchange, poverty microsimulation, the MTEF, cross-sector linkage, and reconciliation
VI — Practice and governance	27–32	Worked examples, cases, interpretation rules, the laboratories, common errors, and disclosure
Appendices	A–I	Glossary, the full card register, the mathematical specification, notation, the module index, commands, the estimate register, references, and the limitations register

Part I — The estate and the evidence
What the system models, the specification it implements, the data it runs on, and how a number gets an interval.

1. What this system models

An education ministry asks four kinds of question, and they need four different kinds of model. How many children are in school and who is not — that is a survey question. How many teachers and classrooms the next decade needs — that is a projection. Where the next school should go — that is a mathematical programme over a geography. What any of it does to the economy — that is a handover to a macro model. The estate this manual describes carries all four, and the first thing to learn is which layer answers which question.

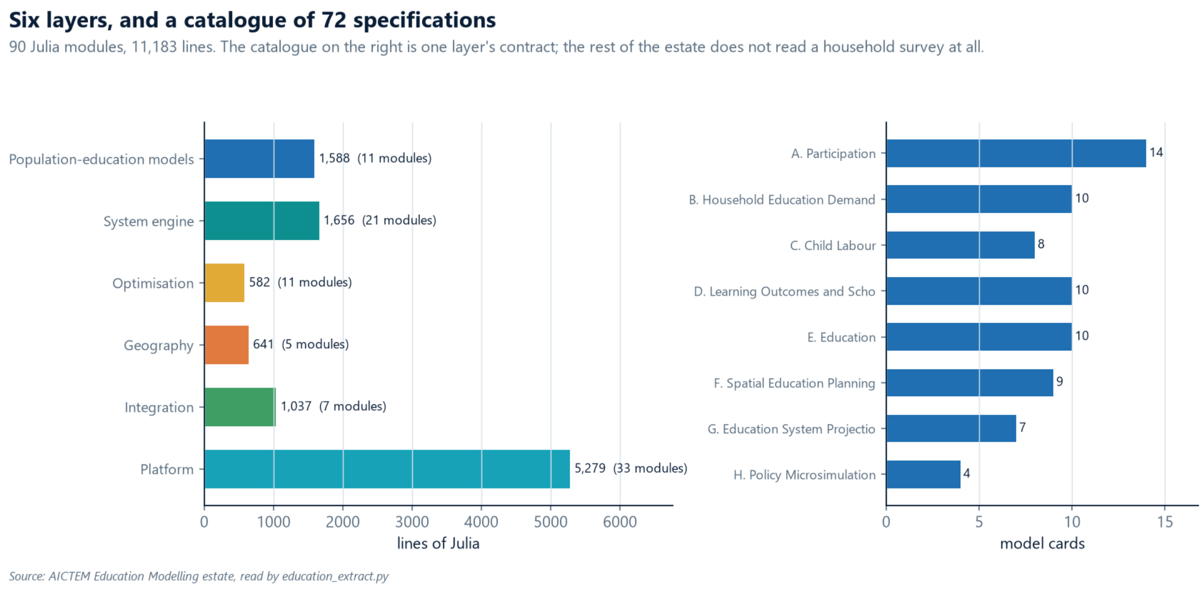


Figure 1. The estate by layer: where each kind of number is produced. — EDIS Education Modelling estate, module line counts

1.1 Six layers

Layer	Modules	Lines	What it owns
Population-education models	11	1,588	The catalogue's estimating equations: discrete choice, expenditure, attainment, equity and microsimulation
System engine	21	1,656	Cohort-component population, grade flow, supply and capacity, teacher stock-flow, costing and the run engine
Optimisation	11	582	Nine programmes: school location, consolidation, teacher deployment, transport, budget, portfolio, transition, multi-objective policy and diagnostics
Geography	5	641	Spatial data, a routable road graph, catchments and climate risk
Integration	7	1,037	Six bridges: the CGE exchange, the CDCGE bridge, MTEF, the budget of the economy, the spatial infrastructure and cross-sector
Platform	33	5,279	Configuration, loaders and harmonisers, the job service, the application and the reporting layer
Total	90	11,183	

Only the first layer reads a household survey. That matters more than it sounds: it means only the first layer produces a confidence interval that comes from a sampling design. Everything above it is deterministic given its parameters, and its uncertainty is a sensitivity range, not a standard error. Confusing the two is how a projection acquires a false air of precision.

1.2 What each layer can and cannot tell you

Layer	Answers	Cannot answer
Population-education models	Who is in school, who pays what, who is excluded, what schooling is worth in the labour market	How many classrooms will be needed in 2035, or where to build them
System engine	How many pupils, teachers, classrooms and shillings the demography implies under stated parameters	Whether any individual child will enrol, or why
Optimisation	Given an objective and a budget, which sites, which deployments, which programmes	Which objective is the right one — that is a political question and the model will optimise whichever you supply
Geography	How far a settlement is from a school, which catchments overlap, which schools are exposed to flooding	Travel time, unless the routed graph is used; the shipped surface is straight-line
Integration	What the macro model receives, what a policy does to poverty, what the fiscal envelope allows	A general-equilibrium result — the coupling is a handover, not a solve

1.3 What has been verified, and by whom

ONE ENVIRONMENT CAVEAT YOU WILL MEET ON THE FIRST MORNING
A machine-level Application Control policy on this teaching host intermittently blocks freshly compiled, unsigned Julia package-cache libraries. Every command in this course therefore passes '--pkgimages=no'. It is not optional and it is the most common first-hour failure.

1.4 The shape of the rest of this manual

Part II works through the catalogue family by family, and it is the longest part because it is where the estimates live. Parts III to V work up the stack: the engine that turns a population into a requirement, the geography that turns a requirement into a location, and the bridges that hand the result to the rest of government. Part VI is practice — the worked examples, the cases, the rules and the errors.

2. The specification: seventy-two models in eight families

The estate implements the *EDIS Education Modelling Technical Manual, EDIS, 2026*. That document defines 72 models, each with the same eight fields, grouped into eight families that this manual takes as its own chapter structure. The families are not an afterthought: they are the units in which an education ministry actually organises work, and a model card that sits in the wrong family will be asked the wrong question.

2.1 The eight families

	Family	Ch.	Models	Cards	This course estimates
A	Participation, Progression and Completion	5	1–14	14	Models 1, 4, 6 and 14 on two surveys
B	Household Education Demand, Costs and Affordability	6	15–24	10	Models 15, 17, 18, 20, 21 on the household panel
C	Child Labour, Social Inclusion and Education Inequality	7	25–32	8	Models 29, 30 and 36 on the survey
D	Learning Outcomes and School Quality	8	33–42	10	**None** — see chapter 8
E	Education, Employment, Earnings and Skills	9	43–52	10	Models 43, 44, 45, 46, 47, 48 and 51 on the panel

	Family	Ch.	Models	Cards	This course estimates
F	Spatial Education Planning and Optimization	10	53–61	9	Models 53, 55, 56 on the register and the GIS database
G	Education System Projection, Costing and Public Finance	11	62–68	7	Models 62, 63, 64 and 67 from the committed run
H	Policy Microsimulation, Human Capital and CGE Integration	12	69–72	4	Models 69 and 72 on the survey and the committed run

2.2 What is on a card

Every card carries the same eight fields. Model 1 is the smallest one and is worth reading in full before any of the others, because the fields mean the same thing everywhere.

Field	Model 1 — School enrolment probability
Purpose	Estimate the probability that a school-age child is currently enrolled and identify household, individual and access constraints.
Dependent variable	Binary current enrolment status.
Core specification	Pr(Enroll_i=1)=F(β0 + X_iβ + H_iγ + A_iδ + μ_r)
Key inputs	age and age squared; sex; household consumption/poverty; parental education; household size and dependency ratio; urban/rural and region; distance or travel time to school; education costs and school type
Estimation method	Survey-weighted logit/probit; multilevel logit when cluster or district random effects are required.
Diagnostics	ROC/AUC, Brier score, calibration plot, marginal effects, design-based standard errors, multicollinearity and sensitivity to age band.
Primary outputs	Predicted enrolment probabilities, marginal effects, enrolment-risk maps and policy elasticities.
Integration	Feeds education demand projections, OOSC model, cash-transfer simulation and spatial targeting.

THE FIELD THAT DECIDES WHETHER YOU CAN RUN THE MODEL

Key inputs. A card runs when every variable it names is present in the dataset you have. It does not run when one is missing, and the correct response is to say which one — not to substitute the nearest available proxy and carry on. Family D is ten cards whose dependent variable is absent from every platform this course reaches, and chapter 8 is what saying so looks like.

2.3 The estimator families the catalogue uses

Estimation or simulation method	Cards
Logit/probit	3
Survey-weighted logit/probit	2
OLS/quantile regression	2
Two-part model	2
Logit/multilevel logit	2
Multilevel regression	2
Binary response model with age-cohort interactions	1
Fractional logit for proportions	1
Logit or gradient boosting for prediction	1
Survey-weighted logit plus small-area/spatial prediction whe	1
Binary/multilevel response model	1
Discrete-time hazard or Cox model	1
Multinomial logit or competing-risk survival model	1

Estimation or simulation method	Cards
Logit/probit with cohort controls	1
Logit/probit or nested choice model	1
Ordered logit/probit	1
OLS, Poisson/negative binomial or hurdle/count model dependi	1
Multinomial or nested logit	1

The concentration matters. Most of the catalogue is discrete choice of one kind or another, which is why chapter 4 spends its time on weighted logits and marginal effects rather than on a survey of methods. Get the binary case right and most of the catalogue follows.

2.4 How the families feed each other

Family	Feeds
A. Participation, Progression and Completion	Demand projection, the out-of-school model, cash-transfer microsimulation and spatial targeting.
B. Household Education Demand, Costs and Affordability	Fee-abolition simulation, bursary design, the poverty link and the financing envelope.
C. Child Labour, Social Inclusion and Education Inequality	Targeting rules, equity floors in the optimisation layer, and the deprivation index.
D. Learning Outcomes and School Quality	Learning-adjusted years of schooling, teacher policy and the infrastructure investment case.
E. Education, Employment, Earnings and Skills	The CGE skilled-labour mapping, rate-of-return analysis and TVET planning.
F. Spatial Education Planning and Optimization	The capital programme, teacher deployment and the MTEF capital line.
G. Education System Projection, Costing and Public Finance	The MTEF envelope, the budget-allocation programme and the financing gap.
H. Policy Microsimulation, Human Capital and CGE Integration	CDCGE, the poverty microsimulation and the cross-sector handover.

3. The data the estate runs on

The estate can reach four kinds of data, and the difference between them governs almost everything about how their outputs may be used. A household survey gives you people with sampling weights and a design. An administrative register gives you institutions with a coverage problem. A geographic layer gives you infrastructure. An indicator database gives you national aggregates someone else computed.

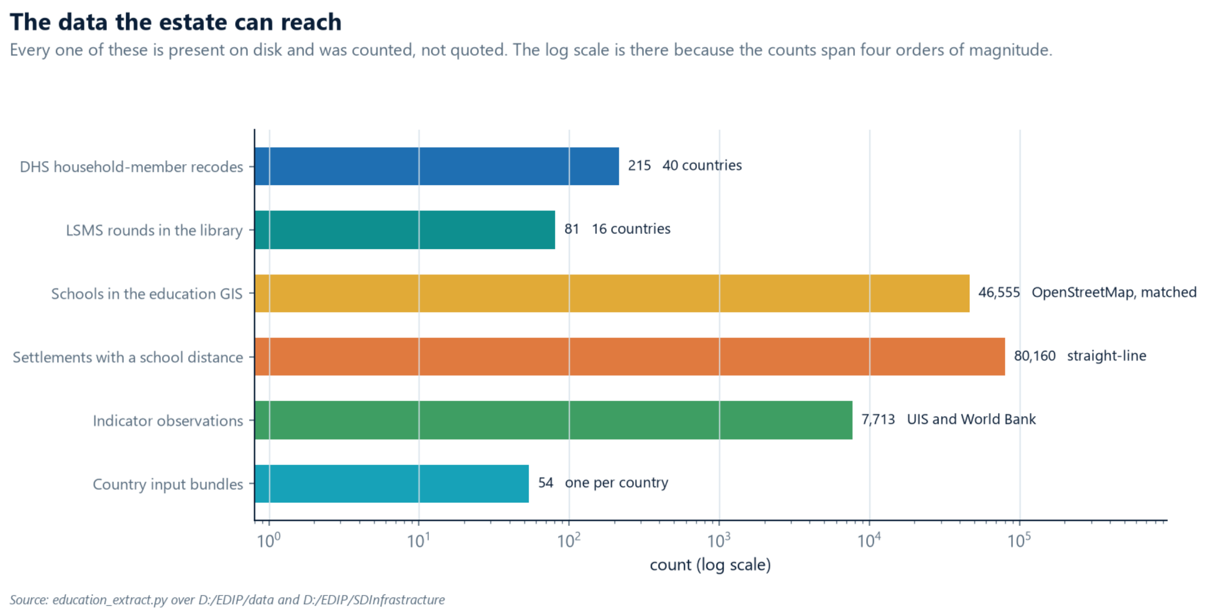


Figure 2. What the estate can reach, and what it has ingested. — Data usage audit; DHS inventory;

LSMS library; education GIS database

3.1 Household surveys

Source	Holds	Education content	Access
Demographic and Health Surveys	215 household-member recodes across 40 countries	Attendance during the school year, level and grade attended, years of schooling, highest level, plus wealth quintile, region and residence	Restricted, registered-user terms
Living Standards Measurement Study	81 rounds across 16 countries	An education module with attendance, attainment and itemised expenditure; a labour module with earnings, hours and formality; a consumption aggregate with a poverty line	Restricted, World Bank microdata terms
Continent-scale harmonisation	Coverage for 55 countries; education variables mapped for 37	Attainment and expenditure variables catalogued	Catalogued only — no microdata downloaded (EG-16)

The two surveys this course estimates on are chosen for contrast, not convenience. The health-and-demography survey carries a large sample, a clean design and a thin education module. The household panel carries a small sample, a consumption aggregate, a labour module and an itemised education-expenditure block. Family A can only be done well on the first; families B and E can only be done at all on the second.

3.2 What the education recode actually says

The household-member recode's attendance variable is defined with three codes: no; currently attending; attended at some time during the current school year. Both surveys used in this manual record only the first and the third. The code that would carry *currently attending* never appears.

Code	Label	Uganda 2016 DHS (UG2016DHS)	Kenya 2022 DHS (KE2022DHS)
0	No	61,210	95,707
1	Currently attending	0	0
2	Attended at some time during the current school year	29,957	60,864

EG-03 — THE LABEL YOU MUST CARRY
Every participation figure in this manual is therefore *attended during the current school year*, not *currently attending*. The difference is a child who left in March and was surveyed in September. Reporting the first under the second's name overstates current participation by an unknown amount, and the recode gives you no way to bound it.

3.3 The school register

An administrative register is not a sample and has no sampling error. It has a coverage problem instead, and the register this course uses has two.

Level	Records
Primary	16,884
Secondary	3,594
Post-primary	102
Other	174
Non-formal	439

Coverage problem	Size	What it does to your analysis
Schools recording no pupils	4,126 of 16,884 primary schools	The register's enrolment total is far below the run's, and any ratio computed over all records is a ratio over a denominator that includes empty rows
Schools without a usable coordinate	64 % carry a coordinate inside the country; 53 % of pupil-bearing schools reach a county	Any map is drawn on a non-random subset. The pupil-teacher ratio on the mappable schools is 48.2 against 49.8 on the full register

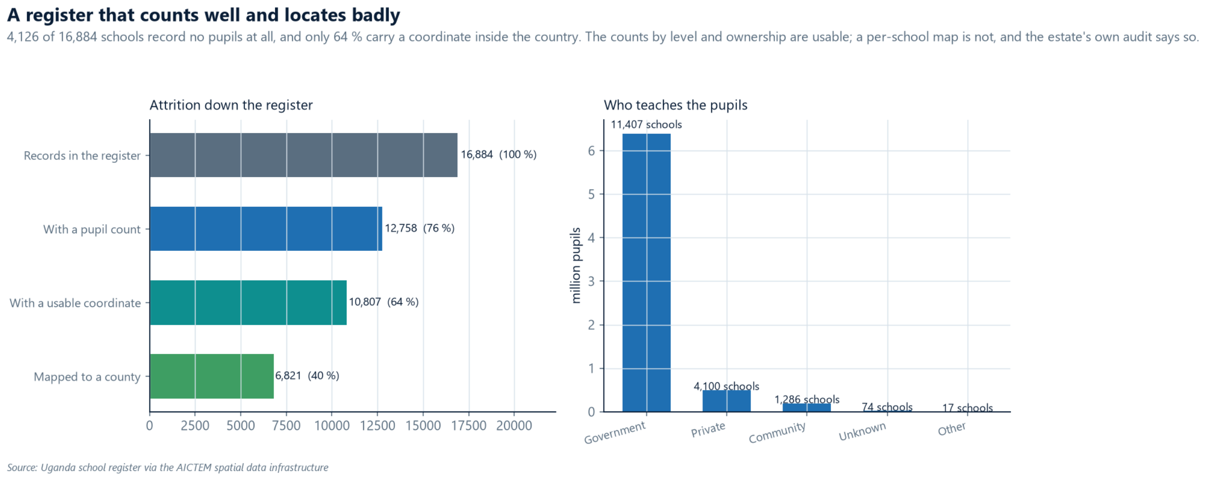


Figure 3. What the school register does and does not support. — Uganda school register (EMIS)

3.4 The geographic layers

Layer	Size	What it supports
Matched school points	46,555 across five countries	School density, catchments, siting
Settlements	81,746	Demand origins for accessibility
Settlement-to-school distances	80,160, straight-line	Coverage curves and county medians
Administrative boundaries	geoBoundaries admin-0 to admin-2	The geography every map and every join is drawn on
Spatial data infrastructure	10 sector layer groups	School registers by level, population grids, derived travel surfaces

EG-13 AND EG-14 — TWO THINGS EVERY ACCESSIBILITY FIGURE MUST SAY
The distances in this manual are straight-line. They ignore roads, rivers and terrain, and they always flatter access. And the settlement register's population field sums to 5,570,285 across 80,160 settlements, well below the national population, so a population-weighted coverage share is a share of the settlement population that carries a value — never of the country.

3.6 The canonical schema, and how a survey reaches it

Every loader in the estate harmonises to one set of variable names. A model card names canonical variables; a mapping file declares how a particular survey's own names translate into them. That indirection is what lets the same specification run on two surveys with entirely different codebooks — and it is where a great deal of quiet error can enter, so the mapping is a declared artefact rather than code.

What the mapping declares	Why it is declared rather than inferred
Which source variable becomes which canonical one	A guess based on a variable label is how ‘hv121‘ becomes 'currently enrolled'. The mapping forces someone to write it down.
Which fields are numeric and which are categorical	Reading a value label as a number, or a code as a category, produces output that looks fine and means nothing.

What the mapping declares	Why it is declared rather than inferred
The weight, and its divisor	Survey weights are often stored scaled by a million. A missing divisor moves every weighted total by six orders of magnitude and no diagnostic catches it.
Value-code recodes, one by one	So that a code with no mapping becomes missing and is flagged, rather than silently taking a default.
What is unavailable	The most valuable field. It records, for each canonical variable this survey cannot supply, why — so a downstream model fails loudly instead of running on a proxy.

THE RULE THE HARMONISER ENFORCES
Nothing is imputed. A code with no declared mapping becomes missing and is flagged; a canonical variable the survey does not carry is declared unavailable with a note saying where it would live if it existed. Both are visible to whatever runs next. The alternative — a default that quietly fills the gap — is how a model comes to be estimated on something nobody chose.

3.6.1 What the household-member mapping declares unavailable

Canonical variable	Why the recode cannot supply it
Consumption	Not in the household-member recode. The wealth quintile is a relative asset index, not money, and cannot be converted to one.
Expenditure	Not collected in the core recodes at all.
Poverty	There is no monetary poverty line in this instrument. The wealth quintile is a proxy for position, not for poverty.
Wages and occupation	In the individual women's and men's recodes, not the household roster this course reads.
Literacy	Likewise.
Parental education	Requires linking a child's record to a parent's line in the roster, which is possible and is not done here.

That table is why families B and E in this course are estimated on the household panel rather than on the health-and-demography survey. It is not a judgement about which survey is better; it is a statement about which questions each instrument asked.

3.6.2 What the household-panel mapping declares

The panel's mapping is longer because the panel carries more. It names the education section, the roster, the identification block and the consumption aggregate; the person and household keys; and — item by item — every component of education expenditure that chapter 6 sums: fees, registration, examination, boarding, uniforms, books, transport, day care and other, plus the respondent's own reported total.

WHY THE ITEM LIST MATTERS
Because summing the items and taking the reported total are two different quantities, and they do not agree. This course computes both and reports the comparison rather than choosing; one expenditure item turns out to be recorded in units that are not shillings, and it is flagged in the shipped data rather than summed silently. A mapping that listed only a single total would have hidden that.

3.7 Harmonising a school register

A school register is harmonised differently from a survey, because the problem is not translation but identity: the same school appears in more than one source under more than one name, and deciding whether two records are one school is a modelling decision.

Stage	What it does	What it must keep
Ingest	Reads each source's records as they are	Every record, including ones that will be rejected
Validate	Applies the source's declared rules	The rule that rejected each rejected record
Match	Proposes candidate pairs across sources	The candidates considered, not only the ones accepted

Stage	What it does	What it must keep
Resolve	Decides which candidates are the same school, and records a match confidence	The confidence, so a downstream user can filter on it
Publish	Writes the deduplicated register	A pointer from each published school back to its contributing source records

The estate's education GIS platform keeps all five. Its rejected-records table and its download-attempts table are the two a sceptical reader should look at first: a pipeline that discards its failures cannot be audited, and this one does not discard them.

3.8 What harmonisation cannot fix

- **A variable the survey did not ask.** No mapping creates data. The unavailable declarations are the honest form of this, and family D is the largest example in the course.
- **A code the survey never used.** The attendance variable's middle code is declared by the recode and absent from both surveys; harmonisation faithfully carries that absence through, which is exactly what it should do.
- **A definition that differs between instruments.** Two surveys can both have a variable called 'attending' and mean different things by it. Harmonisation aligns names, not concepts, and chapter 26 is what happens when that is forgotten.
- **A coordinate that is wrong.** The school register's coordinates pass every format check and place half the schools outside the country. Validation catches malformed data; it cannot catch data that is well formed and false.

3.9 The order to read a new dataset in

	Step	What you are looking for
1	The codebook, before the data	What each variable was asked as, and its full code list
2	The code list against the data	Which declared codes actually appear. This is where the attendance finding came from
3	The design variables	Strata, clusters, weight — and whether any stratum has one cluster
4	The special codes	97, 98, 99 and their relatives. They are not values and they leave the denominator
5	The universe of each module	Who was asked. A skip pattern is not missingness
6	The unavailable list	What this instrument cannot answer, so you stop looking

3.10 What a learner may and may not hold

- **No microdata leaves the platform it lives on.** Every estimate in this course was produced on the machine that holds the data, and what ships is the estimate.
- **No survey cluster coordinate is used or published**, in any form. School coordinates are infrastructure and are mapped freely; cluster coordinates stand for households and are not.
- **No cell below the disclosure threshold is reported.** The threshold in this course is 25 unweighted observations, and every estimate table carries the decision explicitly.

4. Estimation under a complex survey design

Every survey this course touches is stratified, clustered and unequally weighted. Those three facts change the answer in three different ways, and only one of them is fixed by applying the weights.

4.1 What the weights do, and what they do not

Design feature	Effect on the point estimate	Effect on the variance
Unequal selection probability	Corrected by the sampling weight	The weights themselves add variance
Stratification	None	Reduces it
Clustering	None	Increases it, often by a factor of two to thirteen

A weighted proportion with an unweighted standard error is therefore not half right. It is right in the middle and wrong at the edges, which is exactly where a decision gets made.

4.2 The ratio estimator and its linearised variance

Almost every education indicator is a ratio: attenders over children, spenders over households, pupils over teachers. The estimator is the weighted ratio, and its variance comes from linearising around the estimate.

E-0-01

$$\hat R = \frac{\sum_h \sum_i \sum_j w_{hij} y_{hij}}{\sum_h \sum_i \sum_j w_{hij} x_{hij}}$$

$$z_{hij} = w_{hij}\,(y_{hij} - \hat R\, x_{hij})$$

$$\widehat{\mathrm{Var}}(\hat R) = \frac{1}{\hat X^{2}}\sum_h \frac{n_h}{n_h - 1}\sum_i (z_{hi\cdot} - \bar z_{h\cdot})^{2}$$

design-based variance of a ratio

(1)

The inner sum is over primary sampling units within a stratum, not over people. That is the whole of the clustering correction: people inside a cluster resemble each other, so the cluster is the unit that carries independent information.

4.3 Singleton strata

A stratum with one sampling unit contributes no within-stratum variance and the formula divides by zero. Three responses exist: drop the stratum, which biases the point estimate; collapse it into a neighbour, which understates the variance by an unstated amount; or treat it as a certainty unit contributing nothing, which understates the variance by a bounded and declared amount. This manual takes the third and records how often it fires.

Survey	Strata	Clusters	Persons	Singleton strata
Uganda 2016 DHS (UG2016DHS)	34	696	91,167	none
Kenya 2022 DHS (KE2022DHS)	92	1,691	156,571	none

Neither survey has one. That is worth checking rather than assuming: it is the check that fails silently.

4.4 The design effect, and what it costs

E-0-02

$$\mathrm{deff} = \frac{\widehat{\mathrm{Var}}_{\text{design}}(\hat p)}{\hat p (1-\hat p)/n}, \qquad n_{\text{eff}} = \frac{n}{\mathrm{deff}}$$

design effect and effective sample size

(2)

Indicator	Estimate	95 % interval	deff	n	n effective
Primary net attendance	78.8 %	77.9–79.7 %	2.77	20,573	7,437
Secondary net attendance	17.8 %	16.6–19.1 %	3.41	12,331	3,616
Any attendance, 6–18	83.1 %	82.3–83.9 %	3.86	32,904	8,527
Out of school, primary age	10.5 %	9.6–11.3 %	4.08	20,573	5,039

Indicator	Estimate	95 % interval	deff	n	n effective
Over-age for grade	62.0 %	60.9–63.1 %	3.03	23,295	7,679
No education, 15+	13.3 %	12.6–14.0 %	4.96	45,775	9,222

Table 1. Uganda 2016 DHS (UG2016DHS), design-based

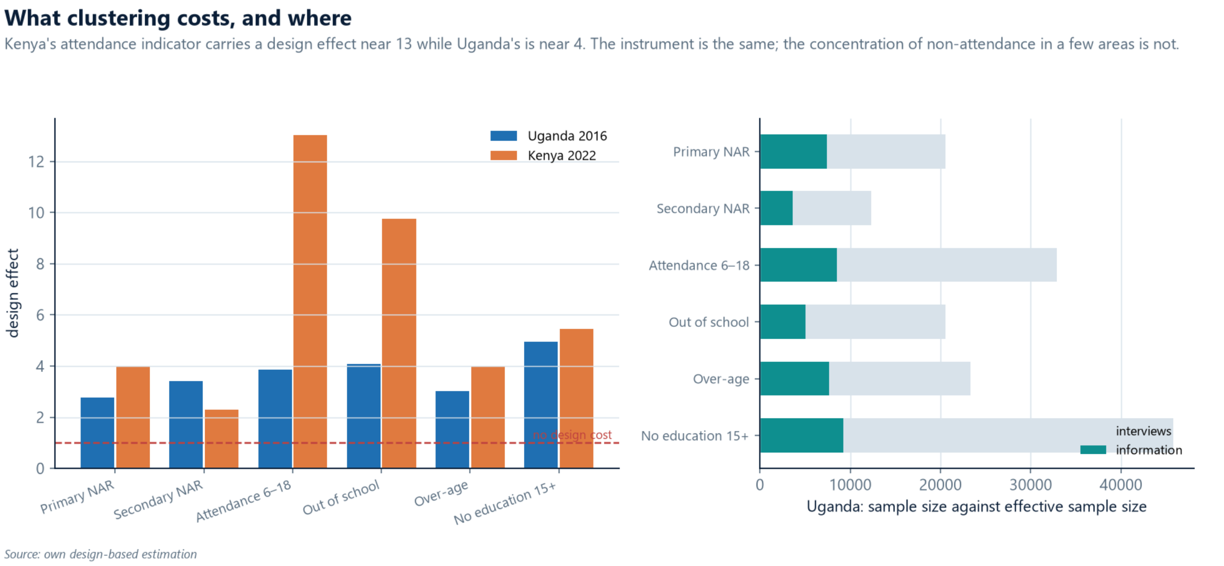


Figure 4. The design effect for the same indicator differs by country because the geography of the outcome differs. — Own estimation, Uganda 2016 DHS and Kenya 2022 DHS

THE SAME INDICATOR, TWO DESIGN EFFECTS
Attendance among 6–18 year olds carries a design effect near 13 in one survey and near 4 in the other. The instrument is the same and the estimator is the same. What differs is the geography of the outcome: where non-attendance is concentrated in a few arid districts, the clusters carry far less independent information. A design effect is a fact about the country as much as about the design.

4.5 Regression under the design

A weighted logit gets the coefficients right and the standard errors wrong, for the same reason a weighted proportion does. The correction is a sandwich whose middle is summed over clusters within strata rather than over observations.

E-0-03

$$\hat V = \mathcal{I}^{-1}\left[\sum_h \frac{n_h}{n_h-1}\sum_i (u_{hi} - \bar u_h)(u_{hi} - \bar u_h)^{\top}\right]\mathcal{I}^{-1}, \qquad u_{hi} = \sum_j w_{hij}\,(y_{hij} - p_{hij})\,x_{hij}$$

design-based sandwich
(3)

Fitted on ages 6–17 in Uganda 2016 DHS (UG2016DHS) — n = 29,538, McFadden R^2 = 0.181 — the correction is not cosmetic.

Term	β	SE, model-based	SE, design-based	Inflation	AME
age	1.6748	0.0445	0.0482	×1.08	+16.90 pp
age^2	-0.0781	0.0020	0.0021	×1.08	-0.79 pp
female	-0.0637	0.0404	0.0436	×1.08	-0.64 pp
urban_d	-0.0084	0.0630	0.1078	×1.71	-0.08 pp
wealth: Highest	0.1506	0.0930	0.1032	×1.11	+1.52 pp
wealth: Lowest	-0.9897	0.0719	0.0911	×1.27	-9.98 pp
wealth: Middle	-0.1691	0.0734	0.0837	×1.14	-1.71 pp
wealth: Second	-0.4553	0.0704	0.0850	×1.21	-4.59 pp

Term	β	SE, model-based	SE, design-based	Inflation	AME

Table 2. Weighted logit of attendance. Region fixed effects fitted and omitted from the table

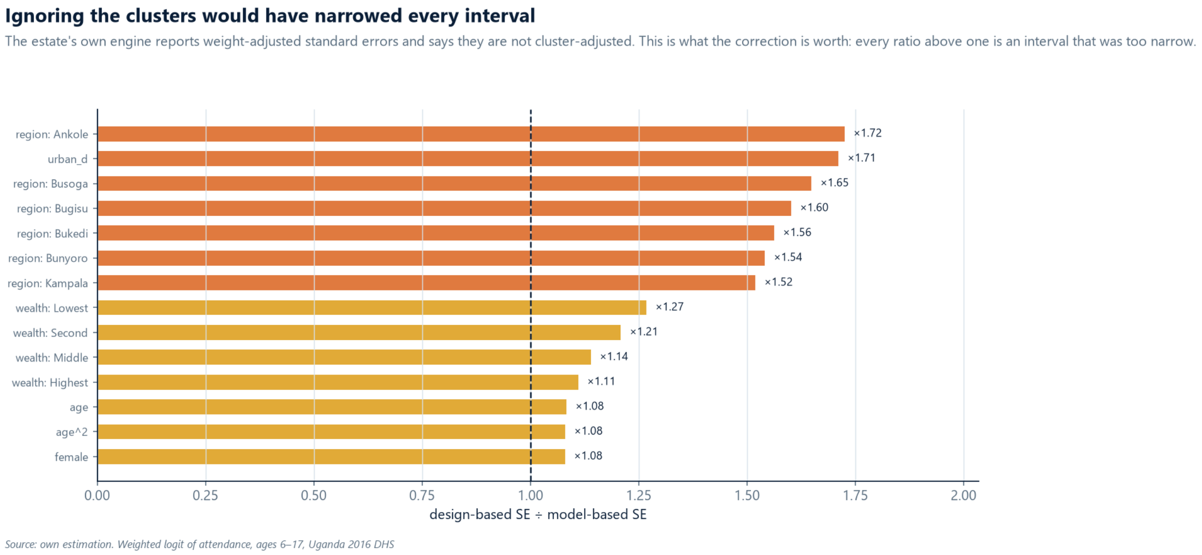


Figure 5. What a design-based standard error costs, coefficient by coefficient. — Uganda 2016 DHS, own estimation

The estate's own estimation engine documents that its standard errors are weight-adjusted and not cluster-adjusted, with design-based clustering listed as a follow-up. This is what that follow-up is worth: every ratio above one in the figure is an interval that was too narrow, and the largest of them is $\times 3.15$.

4.6 The three scales a coefficient lives on

A logit coefficient is on the log-odds scale, which no policy-maker has ever acted on. Two conversions matter.

E-0-04

$$\mathrm{OR} = e^{\beta}, \qquad p_1 = \frac{o_0 e^{\beta}}{1 + o_0 e^{\beta}}, \qquad o_0 = \frac{p_0}{1-p_0}$$

$$\mathrm{AME} = \frac{1}{\sum_i w_i}\sum_i w_i\,p_i (1-p_i)\,\beta$$

odds ratio, risk and average marginal effect

(4)

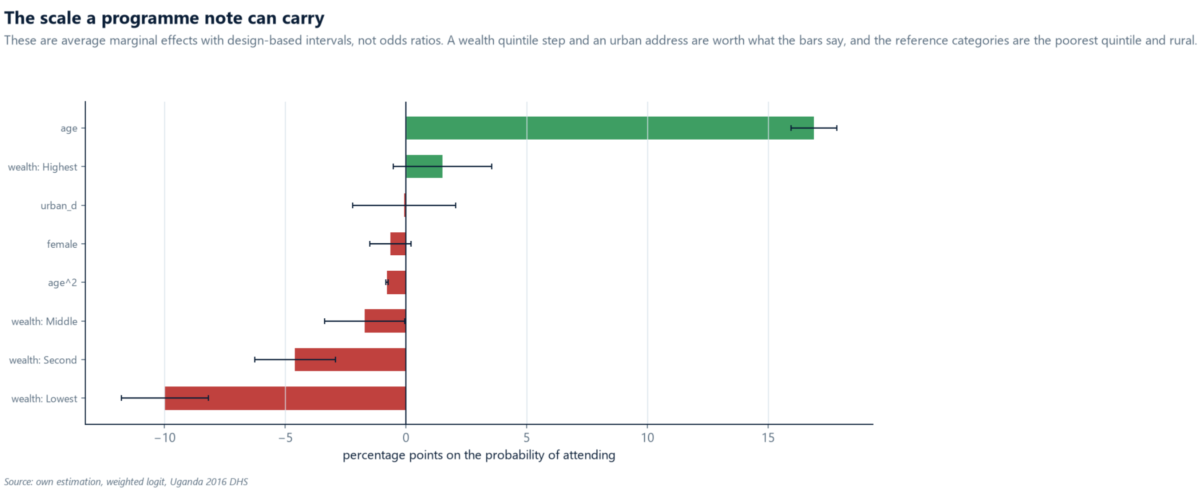


Figure 6. Average marginal effects on attendance, in percentage points. — Uganda 2016 DHS, own estimation

WHAT GOES IN THE NOTE
The average marginal effect, in percentage points, with its design-based interval. An odds ratio may accompany it. An odds ratio alone, in front of a reader who will hear a risk ratio, is an automatic fail in this course.

4.9 The estimator families the catalogue actually uses

Sections 4.1 to 4.8 cover the binary case, which is most of the catalogue. It is not all of it. Seven other families appear, each with its own specification, its own diagnostic and its own characteristic misreading. This section is the short version; the card for any given model names which family it belongs to.

Family	Where it appears	The diagnostic it owes	How it is read wrongly
Binary choice	Most of families A, C, D, F	Discrimination, calibration, marginal effects	The odds ratio reported as a risk ratio
Multinomial choice	School-type and occupation choice	Independence of irrelevant alternatives, or a specification that does not need it	Coefficients read as effects on the level rather than relative to a base category
Ordered choice	Attainment, service ladders	The parallel-lines assumption, tested rather than assumed	One odds ratio quoted as though it held at every cut-point
Count with exposure	Children ever born, repetition counts	Overdispersion, and whether the offset is right	A rate compared across units with different exposure
Duration and hazard	Dropout, school-to-work transition	Proportional hazards, and censoring handled explicitly	A hazard ratio read as a cumulative probability
Two-part	Household education expenditure	Whether the zeros are structural, and the retransformation	The conditional elasticity quoted as an unconditional one
Selection correction	Earnings	The exclusion restriction, named	A significant Mills term treated as evidence the correction worked
Decomposition	Gender and location gaps	That the identity closes	The residual named as a mechanism

4.9.1 Multinomial choice

When the alternatives are unordered and substitution between them is the object — which school type, which occupation — collapsing to a binary throws away the question. The specification is a set of coefficient vectors, one per alternative, identified only relative to a base.

E-0-05

$$\Pr(y_i = j) = \frac{\exp(X_i\beta_j)}{\sum_{k=1}^{J}\exp(X_i\beta_k)}, \qquad \beta_1 = 0$$

multinomial logit
(5)

Two things follow. Every coefficient is relative to the base category, so changing the base changes every number without changing the model. And the marginal effect of a covariate on one alternative's probability depends on all the coefficients, not just that alternative's — which is why the marginal effects, not the coefficients, are what a reader can act on.

THE ASSUMPTION THAT GETS IGNORED
Independence of irrelevant alternatives: the ratio of two alternatives' probabilities must not depend on what else is available. For school choice that is a strong claim — adding a nearby private school plausibly draws more from one option than another. The card names the assumption; the honest responses are a nested or mixed specification, or reporting the limitation.

4.9.2 Ordered choice

Attainment levels and service ladders are ordered, and treating them as unordered discards that. The proportional-odds form assumes one coefficient vector with different intercepts at each cut-point.

E-0-06

$$\Pr(y_i \le j) = F(\kappa_j - X_i\beta), \qquad \kappa_1 < \kappa_2 < \dots < \kappa_{J-1}$$

ordered logit or probit

(6)

The single β is the assumption: a covariate is taken to shift the odds of being above every cut-point by the same amount. It is testable, and the honest workflow tests it rather than assuming it. Where it fails for one covariate, a partial-proportional-odds specification frees that covariate and keeps the rest; where it fails widely, a multinomial model is the fallback and the ordering is reported as unexploited.

4.9.3 Counts with an exposure offset

Children ever born, grades repeated, terms absent — counts over a window of exposure that differs across people. The offset is what makes them comparable.

E-0-07

$$\mathbb{E}[y_i \mid X_i] = \exp\!\left(X_i\beta + \ln t_i\right)$$

Poisson with exposure

(7)

The log-exposure term enters with a coefficient fixed at one, which is what turns a count model into a rate model. Without it the model compares counts across people who were at risk for different lengths of time and attributes the difference to the covariates. The diagnostic the card owes is overdispersion: if the variance exceeds the mean, the Poisson standard errors are too small and a negative binomial or a robust variance is required.

4.9.4 Duration and hazard

Dropout and the transition from school to work are timing questions, and the observation is censored: many people have not yet had the event when the survey stops.

E-0-08

$$h(t \mid X) = h_0(t)\exp(X\beta), \qquad \mathrm{HR} = e^{\beta}$$

proportional hazards

(8)

WHAT A HAZARD RATIO IS NOT

It compares the instantaneous rate of the event among those still at risk. It is not a cumulative probability, it is not a risk ratio, and it does not tell you how many more children will drop out. Converting it to something a programme can act on requires the baseline hazard, which the proportional specification deliberately leaves unspecified.

4.9.5 Two-part models

Household education expenditure has a mass at zero that is not a small value — it is a different decision. The two-part form estimates the decision and the amount separately, and chapter 6 shows why both are needed.

E-0-09

$$\Pr(E_i > 0) = F(Z_i\gamma)$$

$$\mathbb{E}[\ln E_i \mid E_i > 0] = X_i\beta$$

participation, then level

(9)

The coefficient on log consumption in the second equation is an elasticity **conditional on spending**. Quoting it as an unconditional elasticity ignores the households in the first equation that spend nothing, and those are exactly the households a bursary is for. And because the second equation is in logs, predicting a level from it requires a retransformation that the card names and that is easy to forget.

4.9.6 Selection correction

Chapter 9 works this through at length, because on this data it is the central problem rather than a technical footnote. The short form:

E-0-10

$$\ln w_i = X_i\beta + \rho\sigma\lambda(Z_i\gamma) + \varepsilon_i, \qquad \lambda(\cdot) = \frac{\phi(\cdot)}{\Phi(\cdot)}$$

two-step selection correction

(10)

Z must contain something X does not. Without that, λ is a smooth function of variables already in the wage equation and the model is identified by the curvature of the normal hazard alone. Chapter 9 runs it both ways and the difference is large, decisive-looking and meaningless.

4.9.7 Decomposition

A gap between two groups, split into the part explained by differing characteristics and the part explained by differing returns to them.

E-0-11

$$\bar y_A - \bar y_B = \underbrace{(\bar X_A - \bar X_B)\beta_B}_{\text{endowments}} + \underbrace{\bar X_B(\beta_A - \beta_B)}_{\text{coefficients}} + \underbrace{(\bar X_A - \bar X_B)(\beta_A - \beta_B)}_{\text{interaction}}$$

threefold Oaxaca-Blinder

(11)

It is an accounting identity and it closes exactly. That is the only thing it guarantees. The coefficient term is a residual with respect to the covariates included, and it moves as soon as one is added — which is the test, and which is why naming a mechanism for it is an over-claim.

4.9.8 Latent-variable construction

Empowerment indices, deprivation indices and school-quality factors are constructs, not measurements. The specification is explicit about the loadings, and the loadings are the model's statement about how much each observed item tells you about the thing you cannot see.

E-0-12

$$x_{ik} = \mu_k + \lambda_k \eta_i + \epsilon_{ik}$$

a single-factor measurement model

(12)

Two disciplines follow. Report the loadings, because an index without them is a number nobody can interrogate. And where the weights are chosen rather than estimated — as they are in the estate's deprivation index, whose weights live in a configuration file and must sum to one — report the sensitivity of the area ranking to alternative weight sets. A ranking that survives is a finding; one that does not is a statement about the weights.

4.10 What every one of them owes

Whatever the family, four things travel with the estimate and the manual will not accept it without them.

	What	Why
1	The design-based interval	Because the sample is clustered and the model-based one is too narrow
2	The unweighted denominator	Because it is what the interval's width actually rests on
3	The disclosure decision	Because it travels with the cell, not with the publication
4	The scale a reader will act on	A marginal effect, a rate, a level — not a log-odds and not a bare coefficient

4.11 Disclosure

Of 648 estimate cells produced across the two surveys, 0 fall below the unweighted threshold of 25 and are marked undisclosed. They are computed, they are stored, and they are not reported. The rule is not about the estimate's quality — a cell of twelve children can have a perfectly good point estimate — it is about what a small cell reveals.

4.12 What a cross-source check does and does not establish

This manual estimates primary participation on two different surveys with two different designs, and the figures differ. That is not a validation exercise and it does not tell you which is right. Chapter 26 works through why four sources disagree about the same country, and the answer in every case is definitional, temporal or coverage — never that one of them is wrong.

Part II — The catalogue, family by family
All 72 model cards in 8 families, with the estimates this course produced from real household and school data — and, in one chapter, with none.

5. A. Participation, Progression and Completion

Specification chapter	5
Models	1–14 (14 cards)
Estimated in this course	4 on microdata
Feeds	Demand projection, the out-of-school model, cash-transfer microsimulation and spatial targeting.

5.1 What this family estimates

Whether a child is in school, whether they entered at all, how far they get, and where they leave.

5.2 The trap

READ THIS BEFORE THE MODELS
Attendance is not enrolment and neither is completion. The recode that says a child 'attended during the current school year' does not say they are attending now, and in both surveys this course uses, the code that would have meant 'currently attending' is absent from the data.

5.3 The models

#	Model	Dependent variable	Method	Status
1	School enrolment probability	Binary current enrolment status	Survey-weighted logit/ probit	**Estimated**
2	Ever-attendance / school-entry model	Binary ever-attended indicator among relevant ages	Binary response model with age-cohort interactions	Specified
3	Age-at-entry model	Age at first school entry	OLS/quantile regression	Specified
4	Attendance intensity model	Share of scheduled days attended or days absent	Fractional logit for proportions	**Estimated**
5	Chronic absenteeism model	Binary indicator above a defined absenteeism threshold	Logit or gradient boosting for prediction	Specified
6	Out-of-school children model	OOSC status	Survey-weighted logit plus small-area/ spatial prediction when census covariates are available	**Estimated**
7	Grade repetition model	Repeat indicator	Logit/probit	Partly — over-age

#	Model	Dependent variable	Method	Status
8	Promotion model	Promoted to next grade	Binary/multilevel response model	Specified
9	Dropout hazard model	Time/grade until dropout	Discrete-time hazard or Cox model	Specified
10	Competing-risks school-exit model	Exit to work, inactivity, marriage/ household work, migration or other state	Multinomial logit or competing-risk survival model	Specified
11	Primary completion model	Primary completion indicator	Logit/probit with cohort controls	Specified
12	Secondary completion model	Secondary completion indicator	Logit/probit	Specified
13	Tertiary entry model	Tertiary participation/ entry	Logit/probit or nested choice model	Specified
14	Educational attainment ordered model	Ordered categories from no schooling through tertiary	Ordered logit/probit	**Estimated**

Table 3. Family A — all 14 cards

5.4 The specifications that carry the family

Model 1. School enrolment probability

Estimate the probability that a school-age child is currently enrolled and identify household, individual and access constraints.

Model 1
Pr(Enroll_i=1)=F(β0 + X_iβ + H_iγ + A_iδ + μ_r)
Binary current enrolment status
(13)

Key inputs	age and age squared; sex; household consumption/poverty; parental education; household size and dependency ratio; urban/rural and region; distance or travel time to school; education costs and school type
Method	Survey-weighted logit/probit; multilevel logit when cluster or district random effects are required.
Diagnostics	ROC/AUC, Brier score, calibration plot, marginal effects, design-based standard errors, multicollinearity and sensitivity to age band.
Outputs	Predicted enrolment probabilities, marginal effects, enrolment-risk maps and policy elasticities.
Integration	Feeds education demand projections, OOSC model, cash-transfer simulation and spatial targeting.

Model 6. Out-of-school children model

Predict which school-age children are not participating in education and quantify geographic concentration.

Model 6
Pr(OOSC_i=1)=F(Poverty_i, Access_i, Gender_i, ParentEdu_i, Work_i, Region_i)
OOSC status
(14)

Key inputs	age; sex; poverty; disability; child work; parent education; school distance; urban/rural
Method	Survey-weighted logit plus small-area/spatial prediction when census covariates are available.
Diagnostics	Classification metrics, district calibration, uncertainty intervals.

Key inputs	age; sex; poverty; disability; child work; parent education; school distance; urban/rural
Outputs	OOSC count/rate by district and grid cell.
Integration	One of the principal targeting layers in the Education Decision System.

Model 9. Dropout hazard model

Estimate when and why students leave school before completion.

Model 9
h_i(t)=h_0(t) exp(X_iβ)
Time/grade until dropout
(15)

Key inputs	poverty; costs; distance; age-for-grade; child work; household shocks; gender; marriage risk proxies; school quality
Method	Discrete-time hazard or Cox model; survey weights where supported.
Diagnostics	Proportional-hazards test, competing-risks sensitivity, survival calibration.
Outputs	Grade-specific dropout hazards and survival curves.
Integration	Feeds cohort projections and dropout policy simulation.

Model 14. Educational attainment ordered model

Model attained education as an ordered outcome.

Model 14
Edu_i*=X_iβ+ε_i; Edu_i=j if κ_{j-1}<Edu_i*$\leq$$\kappa$_j
Ordered categories from no schooling through tertiary
(16)

Key inputs	age/cohort; sex; poverty; parents education; location; access
Method	Ordered logit/probit; generalized ordered model if proportional-odds fails.
Diagnostics	Parallel-lines test, threshold ordering, marginal effects.
Outputs	Probability of each attainment category.
Integration	Constructs human-capital distributions for household and CGE modules.

Model 2. Ever-attendance / school-entry model

Model whether children ever enter formal schooling and distinguish delayed entry from permanent exclusion.

Model 2
Pr(EverAttend_i=1)=F(X_iβ + H_iγ + Access_iδ)
Binary ever-attended indicator among relevant ages
(17)

Method. Binary response model with age-cohort interactions

Model 3. Age-at-entry model

Estimate determinants of school starting age and delayed entry.

Model 3
EntryAge_i = α + X_iβ + H_iγ + ε_i
Age at first school entry
(18)

Method. OLS/quantile regression; survival formulation when right-censoring matters

Model 4. Attendance intensity model

Explain regularity of attendance among enrolled students.

Model 4
E(AttendanceRate_i|X)=g^{-1}(X_iβ)
Share of scheduled days attended or days absent
(19)

Method. Fractional logit for proportions; Poisson/negative binomial for absence counts

Model 5. Chronic absenteeism model

Identify students with persistently high absence.

Model 5
Pr(ChronicAbsent_i=1)=F(X_iβ+S_sγ)
Binary indicator above a defined absenteeism threshold
(20)

Method. Logit or gradient boosting for prediction; multilevel model for school effects

Model 7. Grade repetition model

Estimate the probability that a learner repeats the current grade.

Model 7
Pr(Repeat_i=1)=F(Student_iβ + Household_iγ + School_iδ)
Repeat indicator
(21)

Method. Logit/probit; multilevel logistic model

Model 8. Promotion model

Estimate probability of successful grade promotion.

Model 8
Pr(Promote_i=1)=F(X_iβ + S_sγ)
Promoted to next grade
(22)

Method. Binary/multilevel response model

Model 10. Competing-risks school-exit model

Distinguish different destinations after leaving education.

Model 10
Pr(J_i=j|exit)=exp(X_iβ_j)/Σ_k exp(X_iβ_k)
Exit to work, inactivity, marriage/household work, migration or other state
(23)

Method. Multinomial logit or competing-risk survival model

Model 11. Primary completion model

Estimate determinants of completing primary education.

Model 11
Pr(CompletePrimary_i=1)=F(X_iβ)
Primary completion indicator
(24)

Method. Logit/probit with cohort controls

Model 12. Secondary completion model

Estimate determinants of lower/upper secondary completion.

Model 12
Pr(CompleteSecondary_i=1)=F(X_iβ)
Secondary completion indicator
(25)

Method. Logit/probit; sequential models by education stage

Model 13. Tertiary entry model

Estimate transition into university/college after secondary school.

Model 13
Pr(Tertiary_i=1)=F(X_iβ + Price_iγ + Access_iδ)
Tertiary participation/entry
(26)

Method. Logit/probit or nested choice model

5.5 What this course estimated

Model 1 and model 4 are the same estimating equation on two different universes, and model 6 is their complement. All three were run on both surveys with a design-based variance.

Indicator	Model	Uganda 2016 DHS (UG2016DHS)	Kenya 2022 DHS (KE2022DHS)
Primary net attendance	1, 4	78.8 % (77.9–79.7 %)	78.2 % (77.2–79.1 %)
Secondary net attendance	1, 4	17.8 % (16.6–19.1 %)	38.7 % (37.8–39.6 %)
Gross primary attendance ratio	4	114.4 % (112.9–116.0 %)	123.8 % (122.1–125.4 %)
Any attendance, ages 6–18	1	83.1 % (82.3–83.9 %)	93.2 % (92.4–94.0 %)
Out of school, primary age	6	10.5 % (9.6–11.3 %)	5.8 % (5.0–6.6 %)
Over-age for grade	3, 7	62.0 % (60.9–63.1 %)	50.6 % (49.5–51.6 %)
Years of schooling, 20–24	14	8.0 (7.8–8.1)	10.7 (10.6–10.9)
No education at all, 15+	2, 14	13.3 % (12.6–14.0 %)	8.3 % (7.9–8.8 %)

Table 4. Design-based estimates with 95 % intervals

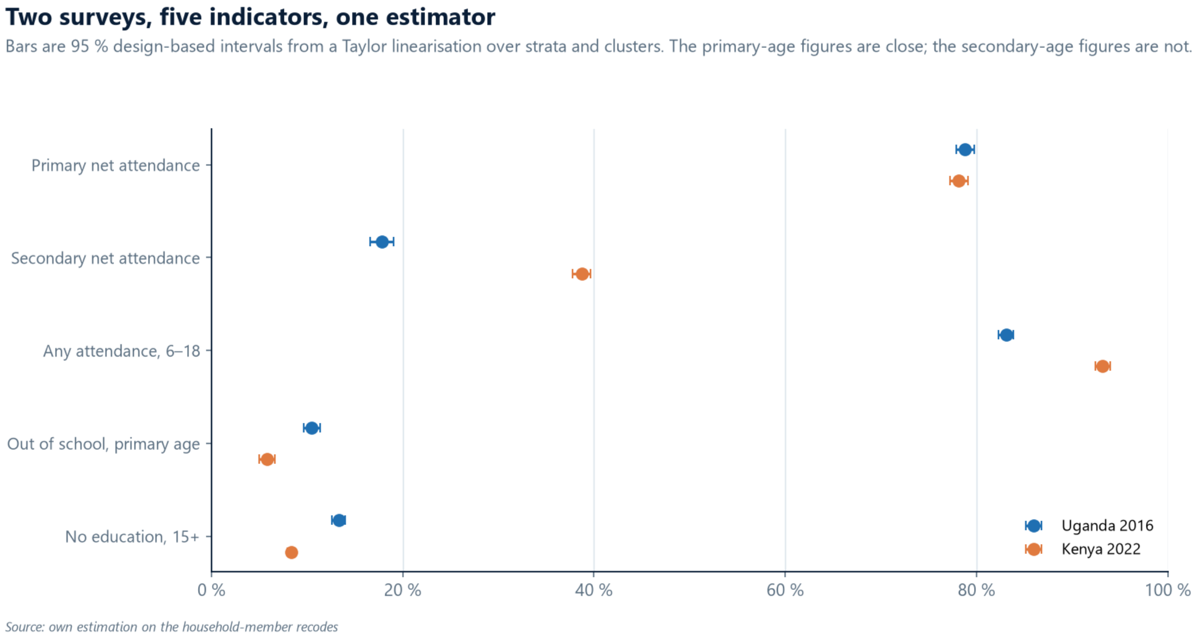


Figure 7. Participation by age band and country, with design-based intervals. — Uganda 2016 DHS and Kenya 2022 DHS, own estimation

5.5.1 The gross ratio is not a proportion

The gross primary attendance ratio is 1.144 — above one, because it counts primary attenders of every age over the primary-age population. It is a ratio of two different universes and it is not bounded by one, which is why it has no design effect in the table: the simple-random-sample comparison that produces a design effect assumes a proportion.

The gap between the gross ratio and the net rate is the over-age problem, and it is large.

5.5.2 Where the risk sits

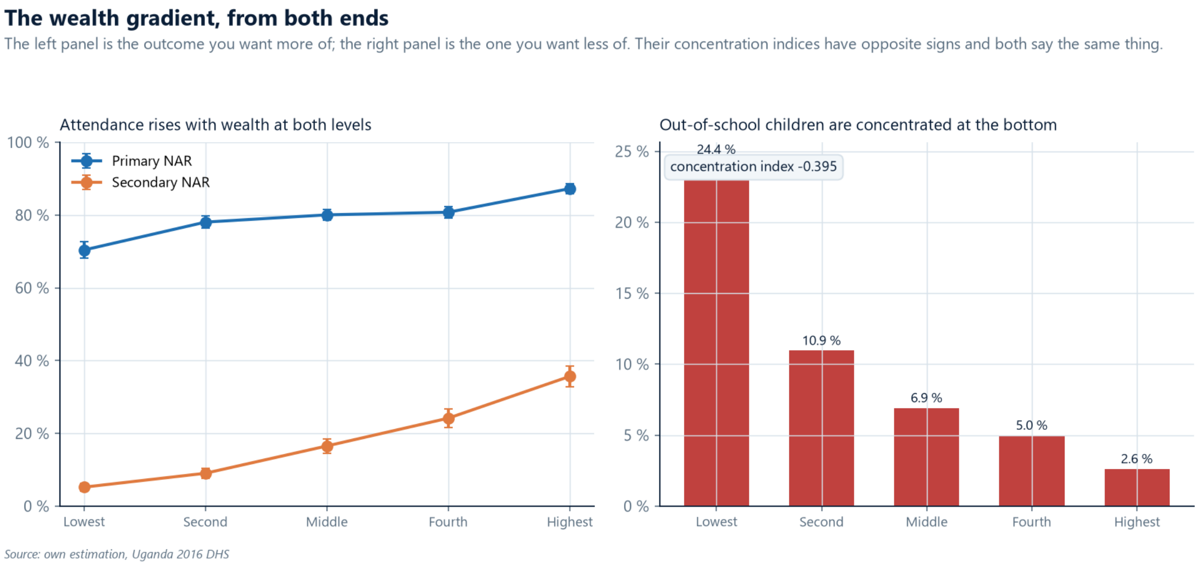


Figure 8. Participation by wealth quintile: the gradient, and where it bites. — Uganda 2016 DHS, own estimation

Wealth quintile	Primary NAR	Secondary NAR	Out of school	n
Lowest	70.4 %	5.2 %	24.4 %	5,259
Second	78.0 %	9.0 %	10.9 %	4,470
Middle	80.0 %	16.5 %	6.9 %	4,269

Wealth quintile	Primary NAR	Secondary NAR	Out of school	n
Fourth	80.8 %	24.2 %	5.0 %	3,757
Highest	87.3 %	35.6 %	2.6 %	2,818

The concentration index of out-of-school status is -0.3949 — strongly negative, meaning exclusion is concentrated among the poor. The concentration index of attendance itself is 0.0463, mildly positive. Those two numbers say the same thing and their signs are opposite, because one measures a good and the other a bad. Reading the sign without the outcome is how a concentration index gets reported backwards.

5.5.3 Region

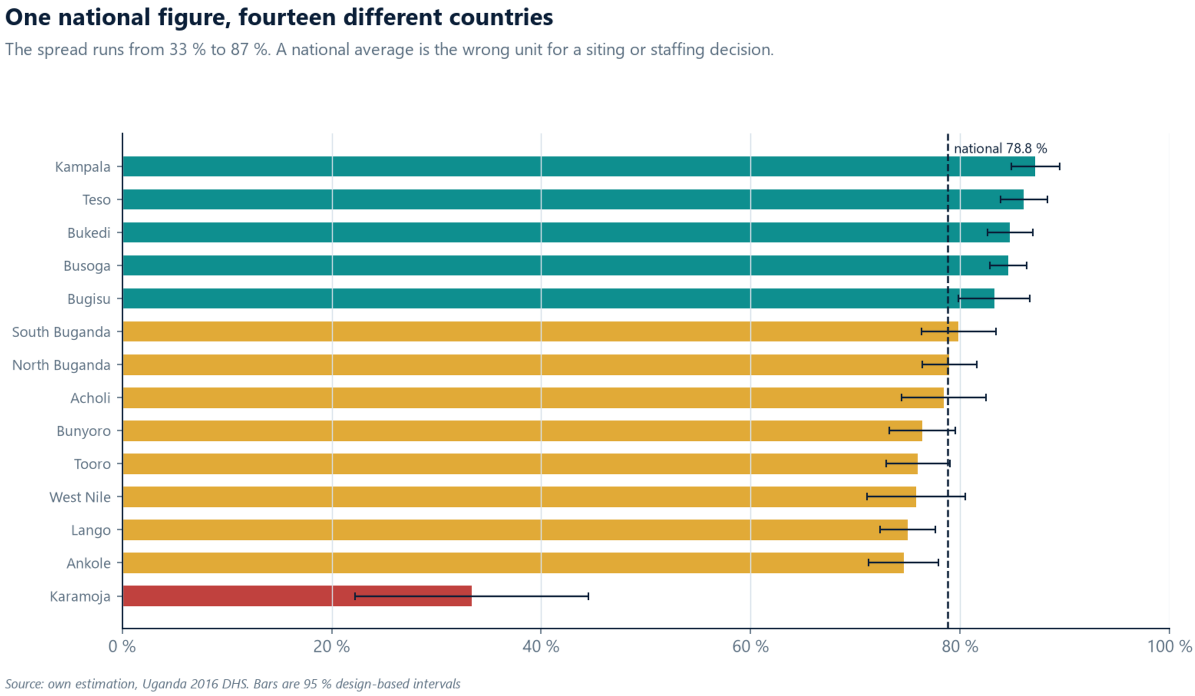


Figure 9. Primary net attendance by region: the range within one country. — Uganda 2016 DHS, own estimation

Region	Primary NAR	95 % interval	deff	n
Karamoja	33.4 %	22.3–44.5 %	16.54	1,143
Ankole	74.6 %	71.3–77.9 %	2.25	1,504
Lango	75.0 %	72.3–77.6 %	1.46	1,499
West Nile	75.8 %	71.1–80.5 %	4.87	1,571
Tooro	76.0 %	72.9–79.0 %	1.98	1,492
Bunyoro	76.4 %	73.3–79.6 %	1.78	1,240
Acholi	78.4 %	74.4–82.5 %	3.45	1,389
North Buganda	79.0 %	76.4–81.6 %	1.58	1,507
South Buganda	79.9 %	76.3–83.4 %	2.67	1,293
Bugisu	83.2 %	79.8–86.7 %	2.62	1,197
Busoga	84.6 %	82.8–86.4 %	1.27	1,993
Bukedi	84.8 %	82.6–86.9 %	1.43	1,508
Teso	86.1 %	83.8–88.3 %	1.66	1,533
Kampala	87.2 %	84.9–89.5 %	0.86	695

Table 5. Uganda 2016 DHS (UG2016DHS)

5.5.4 Over-age enrolment

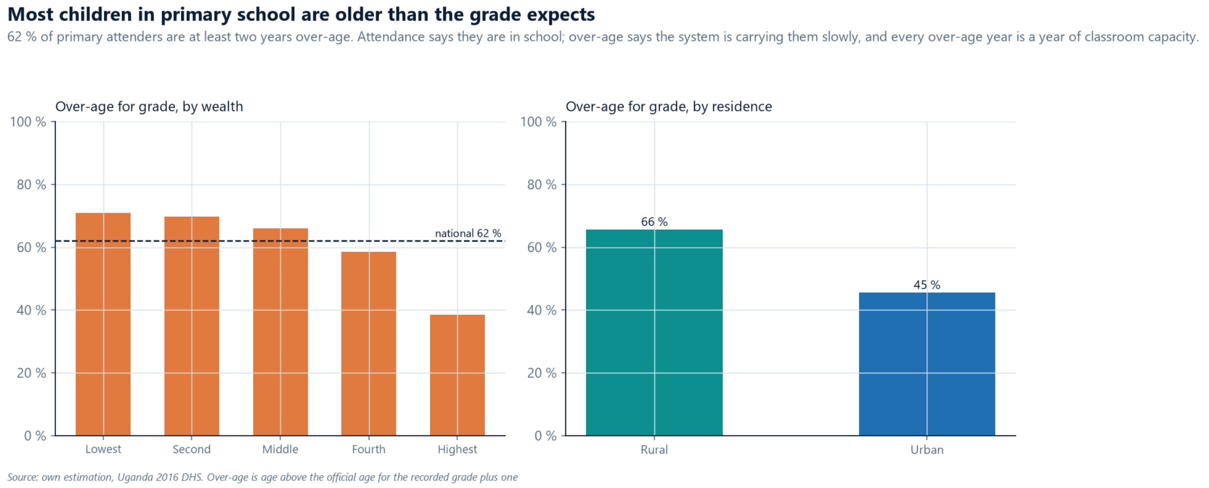


Figure 10. Over-age enrolment: the hidden inefficiency inside a healthy attendance rate. — Uganda 2016 DHS, own estimation

62.0 % of primary attenders are at least two years older than the official age for the grade they are in. Attendance says they are in school. Over-age says the system is carrying each of them for longer than it planned, and every over-age year is a year of classroom capacity and a year of teacher time. A capacity projection built on a net attendance rate and an on-time grade progression will understate demand systematically.

5.6 What this course did not estimate, and why

Model	Why not
3. Age at entry	Needs age at first entry, which neither recode carries directly. Over-age for grade is used as the observable consequence instead, and labelled as such.
5. Chronic absenteeism	Needs days attended, which is not in either recode.
8–10. Promotion, dropout hazard, competing-risks exit	Need either two observations of the same child or a retrospective schooling history. The household panel has waves; wiring them is the natural extension of this chapter.
11–13. Completion and tertiary entry	Estimable in principle from attainment; not run here because the completion definitions differ between the two surveys and the comparison would be definitional, not substantive.

5.8 Every model in this family the course does not estimate

With the input each one needs that is not present. This is the family's data request, and it is more useful than a proxy.

#	Model	Dependent variable	What it needs that is missing
2	Ever-attendance / school-entry model	Binary ever-attended indicator among relevant ages	age; sex; birth order; poverty; parental schooling; distance/travel time; disability; region
3	Age-at-entry model	Age at first school entry	sex; poverty; parental education; preschool attendance; distance; siblings; rurality
5	Chronic absenteeism model	Binary indicator above a defined absenteeism threshold	attendance history where available; travel time; poverty; household shocks; school calendar; school quality proxies

#	Model	Dependent variable	What it needs that is missing
7	Grade repetition model	Repeat indicator	age-for-grade; attendance; poverty; previous performance; class size; teacher availability; school type
8	Promotion model	Promoted to next grade	attendance; age-for-grade; learning score; household welfare; school quality; teacher inputs
9	Dropout hazard model	Time/grade until dropout	poverty; costs; distance; age-for-grade; child work; household shocks; gender; marriage risk proxies; school quality
10	Competing-risks school-exit model	Exit to work, inactivity, marriage/household work, migration or other state	age; sex; education level; household status; local labour market; marital status
11	Primary completion model	Primary completion indicator	entry age; sex; poverty; parents education; distance; repetition history; region
12	Secondary completion model	Secondary completion indicator	primary completion; gender; household resources; school access; school costs; labour demand; region
13	Tertiary entry model	Tertiary participation/entry	secondary completion; household income; fees; distance; sex; field availability; urban/rural

5.7 What this family feeds

Demand projection, the out-of-school model, cash-transfer microsimulation and spatial targeting.

6. B. Household Education Demand, Costs and Affordability

Specification chapter	6
Models	15–24 (10 cards)
Estimated in this course	5 on microdata
Feeds	Fee-abolition simulation, bursary design, the poverty link and the financing envelope.

6.1 What this family estimates

What a household pays, what share of its budget that is, when the payment becomes catastrophic, and how demand responds to fees.

6.2 The trap

READ THIS BEFORE THE MODELS
The denominator. Per student and per household with students differ by the number of students in the household, and a figure quoted under the wrong one overstates the cost of educating a child by that factor.

6.3 The models

#	Model	Dependent variable	Method	Status
15	Years-of-schooling model	Completed schooling years	OLS, Poisson/ negative binomial or hurdle/count model depending distribution	**Estimated**
16	School-type choice model	Discrete school-type choice	Multinomial or nested logit	Specified
17	Household education expenditure two-part model	Positive-spending indicator and positive expenditure amount	Two-part model	**Estimated**
18	Education budget-share model	Education expenditure share	Working-Leser or fractional regression	**Estimated**
19	Almost Ideal Demand System for education	Budget shares for multiple commodity groups	AIDS/QUAIDS with adding-up, homogeneity and symmetry restrictions	Specified
20	Education affordability model	Education burden ratio and high-burden indicator	Descriptive distribution plus logit for high-burden risk	**Estimated**
21	Catastrophic education expenditure model	Catastrophic expenditure indicator	Microsimulation and survey-weighted incidence	**Estimated**
22	Fee elasticity of enrolment	Enrolment or attendance	Logit marginal effects, IV where price endogeneity is credible, or quasi-experimental design	Specified
23	Private tutoring demand model	Tutoring participation and expenditure	Two-part model	Specified
24	School feeding participation model	Attendance/enrolment and feeding participation	Difference-in-differences, matching or multilevel regression depending design	Specified

Table 6. Family B — all 10 cards

6.4 The specifications that carry the family

Model 17. Household education expenditure two-part model

Estimate whether households spend on education and how much they spend conditional on positive expenditure.

Model 17
$\Pr(E\_h>0)=F(Z\_h\gamma)$; $\ln(E\_h|E\_h>0)=X\_h\beta+\varepsilon\_h$
Positive-spending indicator and positive expenditure amount
(27)

Key inputs	income/consumption; number of students; school levels; school type; urban/rural; fees and transport
Method	Two-part model; hurdle GLM; survey-weighted estimation.
Diagnostics	Smearing/retransformation, heteroskedasticity, prediction by quintile.
Outputs	Expected household education spending and cost incidence.
Integration	Feeds affordability, poverty and fee-reform simulations.

Model 18. Education budget-share model

Estimate the share of household expenditure allocated to education.

Model 18 w_edu=α+β ln(X/P)+γZ+ε *Education expenditure share* (28)

Key inputs	real total expenditure; prices/fees; number and age of children; school type; location
Method	Working-Leser or fractional regression.
Diagnostics	Bounds, heteroskedasticity, Engel-curve shape.
Outputs	Budget share and expenditure elasticity.
Integration	Quantifies household burden and demand response.

Model 21. Catastrophic education expenditure model

Estimate incidence of education spending that pushes household non-education resources below a welfare threshold.

Model 21 Cat_h=1[(C_h-EduExp_h)<z_h] *Catastrophic expenditure indicator* (29)

Key inputs	consumption aggregate; education expenditure; poverty threshold
Method	Microsimulation and survey-weighted incidence.
Diagnostics	Threshold sensitivity and counterfactual checks.
Outputs	Incidence, depth and distribution of catastrophic education spending.
Integration	Links education financing to poverty protection.

Model 22. Fee elasticity of enrolment

Estimate responsiveness of participation to direct schooling costs.

Model 22 ln(Q_i)=α+η ln(Fee_i)+X_iβ+ε_i *Enrolment or attendance* (30)

Key inputs	fees; other costs; income; school quality/access; grade
Method	Logit marginal effects, IV where price endogeneity is credible, or quasi-experimental design.
Diagnostics	Instrument strength where IV is used, placebo and robustness checks.
Outputs	Fee elasticity and predicted participation response.
Integration	Parameterizes fee-removal scenarios.

Model 15. Years-of-schooling model

Estimate completed years of schooling as a compact human-capital outcome.

Model 15 SchoolYears_i=α+X_iβ+ε_i *Completed schooling years* (31)

Method. OLS, Poisson/negative binomial or hurdle/count model depending distribution

Model 16. School-type choice model

Model household choice among public, private, religious/community and no school.

Model 16
Pr(Y_i=j)=exp(X_iβ_j)/Σ_k exp(X_iβ_k)
Discrete school-type choice
(32)

Method. Multinomial or nested logit

Model 19. Almost Ideal Demand System for education

Estimate expenditure and price substitution between education and other household consumption categories.

Model 19
w_i=α_i+Σ_j γ_ij ln p_j+β_i ln(X/P)+ε_i
Budget shares for multiple commodity groups
(33)

Method. AIDS/QUAIDS with adding-up, homogeneity and symmetry restrictions

Model 20. Education affordability model

Measure financial stress caused by education costs.

Model 20
Burden_h=EducationExpenditure_h/TotalConsumption_h
Education burden ratio and high-burden indicator
(34)

Method. Descriptive distribution plus logit for high-burden risk

Model 23. Private tutoring demand model

Estimate determinants and distribution of household spending on tutoring.

Model 23
Pr(Tutor_i=1)=F(X_iβ); ln(Cost_i|Tutor)=Z_iγ+ε_i
Tutoring participation and expenditure
(35)

Method. Two-part model

Model 24. School feeding participation model

Estimate association/effect of school feeding on attendance and enrolment where program data permit.

Model 24
Y_i=α+τ Feed_i+X_iβ+ε_i
Attendance/enrolment and feeding participation
(36)

Method. Difference-in-differences, matching or multilevel regression depending design

6.5 What this course estimated

Estimated on Uganda LSMS-ISA National Panel Survey 2019/20 (UNPS): 3,062 households, of which 2,243 have at least one currently attending student — 72.3 % of households, weighted.

Quantity	**Estimate**	**Model**
Households with a student	72.3 %	—

Quantity	Estimate	Model
Students per such household	2.69	—
Spending per student, mean	375k UGX	17
Spending per student, median	144k UGX	17
Education budget share, mean	9.3 %	18
Catastrophic at 10 % of consumption	32.2 %	21
Catastrophic at 20 %	12.2 %	21
Concentration index of spending	0.4134	18
Concentration index of the budget share	0.1448	18

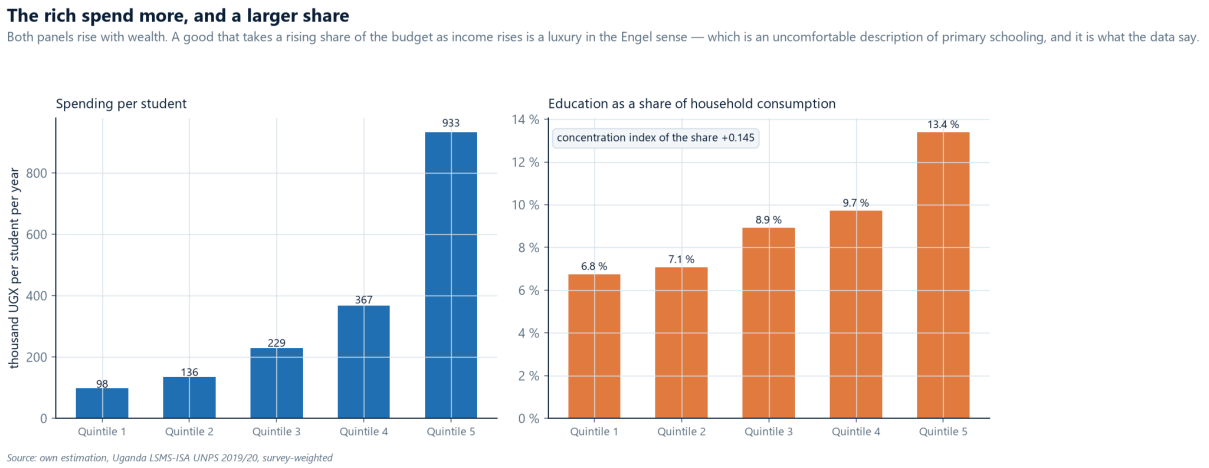


Figure 11. What a household pays per student, and what share of its consumption that is. — Uganda LSMS-ISA UNPS 2019/20, own estimation

6.5.1 The denominator case

This is the most important half-page in the manual, and it is not about an estimator.

Way of dividing	Result
Mean of per-student ratios	375k UGX
Total spending ÷ total students	382k UGX
Among spenders, mean per student	382k UGX
Among spenders, total ÷ total	389k UGX
Per household with students	**1.03 m UGX**

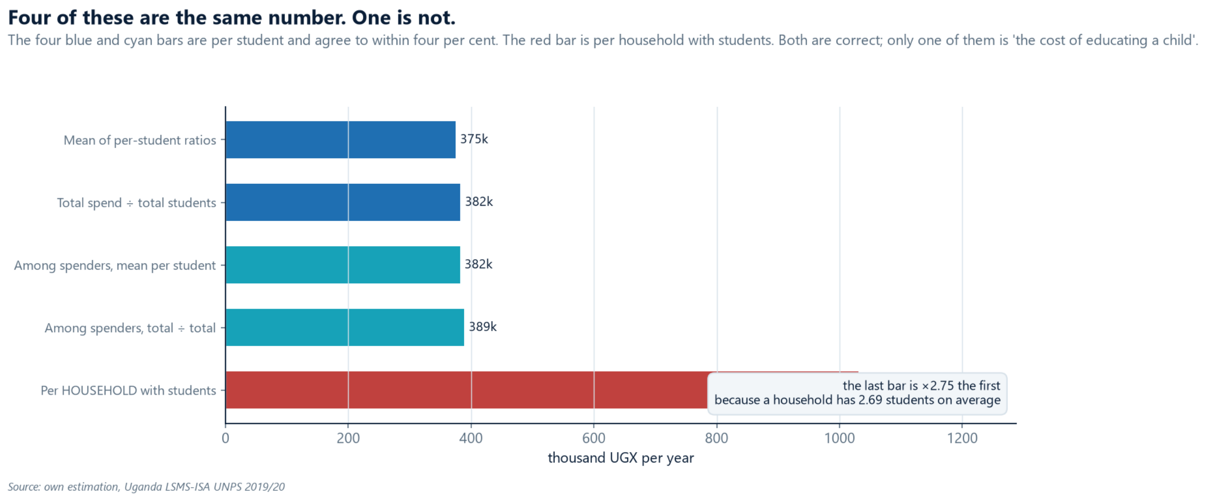


Figure 12. One number, five denominators: which one is 'per student'. — Uganda LSMS-ISA UNPS 2019/20, own estimation

EG-12 — RIGHT ARITHMETIC, WRONG LABEL

The estate's own data audit reports household education spending of about one million shillings per student per year. Re-deriving from the same file gives 1.03 m UGX per household with students and 375k UGX per student, and the mean household has 2.69 students. The figure is right. The denominator in its label is not, and a fee-abolition costing built on it would be nearly three times too large.

Nothing about this is exotic. It is what happens when a quantity is computed correctly and described quickly, and it is why two of this course's seven automatic fails are denominator failures.

6.5.2 The two-part model

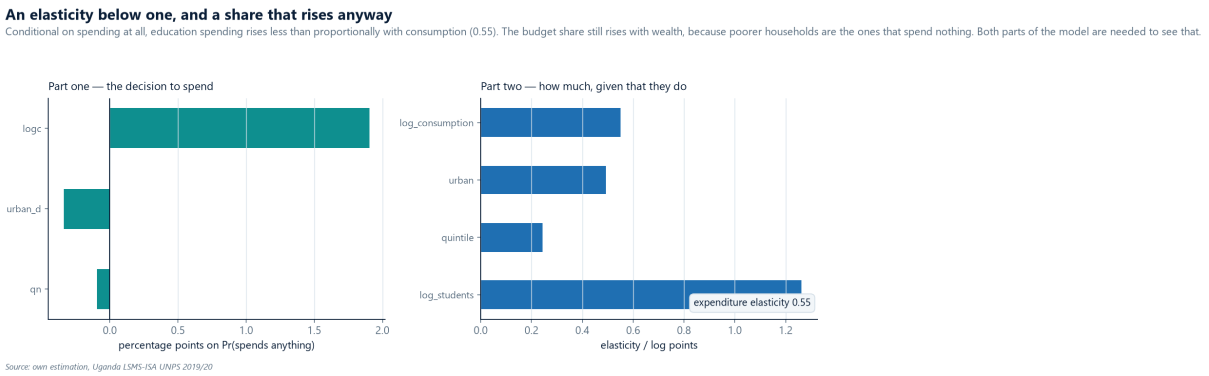


Figure 13. The two-part model: who spends, and how much. — Uganda LSMS-ISA UNPS 2019/20, own estimation

Model 17 is a two-part specification: a participation equation for whether the household spends anything, then a level equation for how much, conditional on spending. Both parts are needed and they say different things.

Term	Coefficient
(Intercept)	2.1739
log_consumption	0.5507
urban	0.4925
quintile	0.2441
log_students	1.2620

Table 7. Level equation, n = 2,200, weighted least squares on log spending

The expenditure elasticity is 0.551 — below one. Conditional on spending at all, education spending rises less than proportionally with consumption, which is the signature of a necessity. And yet the budget share rises with wealth, because poorer households are disproportionately the ones that spend nothing. Either fact alone would mislead; the two-part structure is what lets you hold both.

6.5.3 Catastrophic expenditure, and why it is lowest among the poor

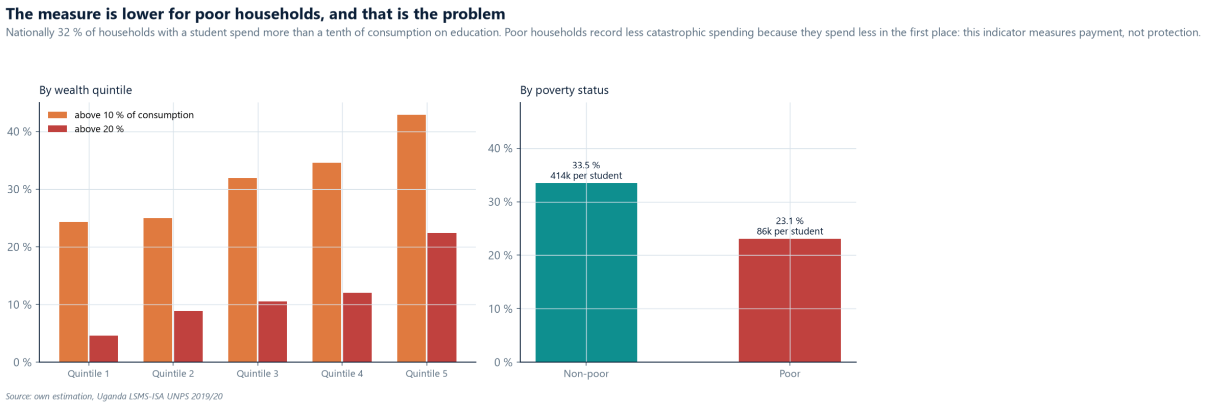


Figure 14. Catastrophic education expenditure, and why it is lowest among the poor. — Uganda LSMS-

ISA UNPS 2019/20, own estimation

Group	Catastrophic at 10 %	Budget share	Spend per student	n
Non-poor	33.5 %	9.7 %	414k UGX	1,917
Poor	23.1 %	6.5 %	86k UGX	326

Poor households record less catastrophic education expenditure than non-poor ones. That is not protection. It is exclusion showing up in a payment indicator: a household that cannot pay does not register a catastrophic payment. An indicator built on spending will always be blind to the household that spends nothing because it has withdrawn the child, which is precisely the household a bursary is for.

6.5.4 The impoverishing effect

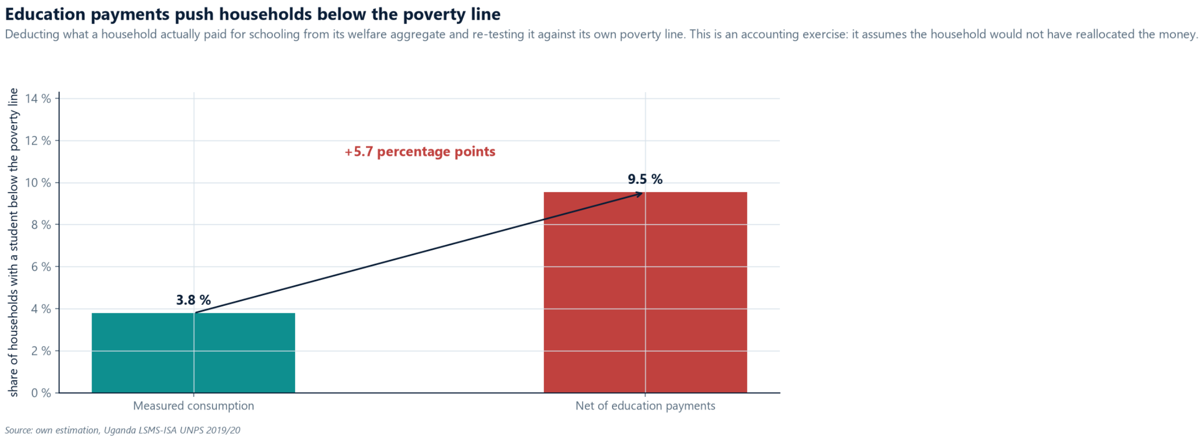


Figure 15. Poverty before and after education payments are deducted. — Uganda LSMS-ISA UNPS 2019/20, own estimation

Deducting what each household actually paid for schooling from its welfare aggregate and re-testing it against its own poverty line moves the poverty rate among households with students from 3.8 % to 9.5 % — an impoverishing effect of +5.7 pp.

WHAT THIS CALCULATION ASSUMES
That the household would not have reallocated. It is an accounting exercise on observed payments, not a behavioural model of what a family would do if fees vanished. It bounds the static burden and it is the right input to a fee-abolition discussion; it is not the answer to one.

6.6 What this course did not estimate

Model	Why not
16. School-type choice	The panel records day, boarding or both, not government/ private in a form comparable across waves.
19. Almost Ideal Demand System	Needs education prices, which no household survey observes; fees paid are an expenditure, not a price.
22. Fee elasticity of enrolment	Needs price variation with an identification strategy. The cross-sectional variation in fees paid is choice, not price.
23–24. Private tutoring and school feeding	Neither is separately itemised in the expenditure block.

6.8 Every model in this family the course does not estimate

With the input each one needs that is not present. This is the family's data request, and it is more useful than a proxy.

#	Model	Dependent variable	What it needs that is missing
16	School-type choice model	Discrete school-type choice	household income; fees; distance by school type; quality proxies; religion only where ethically and legally appropriate; urban/rural
19	Almost Ideal Demand System for education	Budget shares for multiple commodity groups	commodity-group prices; real household expenditure; demographics
22	Fee elasticity of enrolment	Enrolment or attendance	fees; other costs; income; school quality/access; grade
23	Private tutoring demand model	Tutoring participation and expenditure	income; exam grade; school type; learning score; urban/rural
24	School feeding participation model	Attendance/enrolment and feeding participation	program exposure; poverty; distance; school fixed effects where possible

6.7 What this family feeds

Fee-abolition simulation, bursary design, the poverty link and the financing envelope.

7. C. Child Labour, Social Inclusion and Education Inequality

Specification chapter	7
Models	25–32 (8 cards)
Estimated in this course	2 on microdata
Feeds	Targeting rules, equity floors in the optimisation layer, and the deprivation index.

7.1 What this family estimates

Who is excluded and why: work instead of school, disability, gender, location, birth order and household shocks.

7.2 The trap

READ THIS BEFORE THE MODELS
A decomposition is an accounting identity, not a cause. The unexplained share of a gender gap is what the covariates did not account for — it is not discrimination measured.

7.3 The models

#	Model	Dependent variable	Method	Status
25	Child labour versus schooling multinomial model	School only, work only, school+work, neither	Multinomial logit/probit	Specified
26	Child work-hours model	Weekly child work hours	Poisson/negative binomial or two-part model	Specified
27	Household shock and schooling resilience model	Enrolment, attendance, dropout or expenditure	Panel fixed effects or difference-in-differences when LSMS panel data exist	Specified
28	Disability and education exclusion model	Enrolment, attendance, attainment or completion	Survey regression with interactions and decomposition	Specified
29	Gender education gap decomposition	Education outcome gap by sex	Oaxaca-Blinder for continuous outcomes	**Estimated**

#	Model	Dependent variable	Method	Status
30	Rural-urban education gap decomposition	Enrolment/ completion/ attainment gap	Oaxaca-Blinder/ reweighting/ decomposition	**Estimated**
31	Intergenerational education mobility model	Child education years/ level	OLS/ordered models	Specified
32	Sibling competition / resource dilution model	Education expenditure, enrolment or attainment	Household fixed effects where sibling-level data permit	Partly — sibling count

Table 8. Family C — all 8 cards

7.4 The specifications that carry the family

Model 25. Child labour versus schooling multinomial model

Model mutually exclusive child activity states.

Model 25
$\Pr(Y_i=j)=\exp(X_i\beta_j)/\Sigma_k \exp(X_i\beta_k)$
School only, work only, school+work, neither
(37)

Key inputs	age; sex; poverty; farm ownership; season; household labour demand; distance; parent education
Method	Multinomial logit/probit.
Diagnostics	IIA tests, marginal effects, seasonal sensitivity.
Outputs	Probability of each activity state.
Integration	Links education to agriculture, household labour and social protection.

Model 29. Gender education gap decomposition

Separate observed gender gaps into differences in characteristics and coefficients.

Model 29
$\Delta Y=(\bar{X}_g-\bar{X}_b)\beta^* + \bar{X}_g(\beta_g-\beta^*) + \bar{X}_b(\beta^*-\beta_b)$
Education outcome gap by sex
(38)

Key inputs	household characteristics; access; cost; parent education; location
Method	Oaxaca-Blinder for continuous outcomes; nonlinear decomposition for binary outcomes.
Diagnostics	Reference-coefficient sensitivity and reweighting.
Outputs	Explained/unexplained portions of gender gaps.
Integration	Feeds gender-targeted policy design.

Model 31. Intergenerational education mobility model

Estimate persistence of educational attainment across generations.

Model 31
$EduChild_i=\alpha+\beta_m EduMother_i+\beta_f EduFather_i+X_i\gamma+\varepsilon_i$
Child education years/level
(39)

Key inputs	mother education; father education; child cohort; sex; location; wealth
Method	OLS/ordered models; rank-rank mobility measures.
Diagnostics	Cohort sensitivity and measurement error.
Outputs	Intergenerational elasticity and mobility ranks.
Integration	Creates EDIS Education Mobility Index.

Model 32. Sibling competition / resource dilution model

Estimate how household child composition influences schooling investments.

Model 32
Y_ih=α+β NChildren_h+γ BirthOrder_i+X_iδ+ε_i
Education expenditure, enrolment or attainment
(40)

Key inputs	number of children; birth order; age spacing; income; parent education
Method	Household fixed effects where sibling-level data permit.
Diagnostics	Within-household identification and age controls.
Outputs	Resource-dilution effects.
Integration	Improves household education-demand simulation.

Model 26. Child work-hours model

Estimate intensive margin of child labour and its schooling trade-off.

Model 26
E(Hours_i|X)=exp(X_iβ)
Weekly child work hours
(41)

Method. Poisson/negative binomial or two-part model

Model 27. Household shock and schooling resilience model

Estimate how drought, illness, job loss, price shocks or conflict affect education participation.

Model 27
Y_it=α_i+λ_t+β Shock_it+X_itγ+ε_it
Enrolment, attendance, dropout or expenditure
(42)

Method. Panel fixed effects or difference-in-differences when LSMS panel data exist

Model 28. Disability and education exclusion model

Quantify education gaps associated with disability status.

Model 28
Y_i=α+β Disability_i+X_iγ+ε_i
Enrolment, attendance, attainment or completion
(43)

Method. Survey regression with interactions and decomposition

Model 30. Rural-urban education gap decomposition

Quantify how access, poverty and family characteristics explain rural-urban differences.

Model 30
ΔY=Explained+Unexplained
Enrolment/completion/attainment gap
(44)

Method. Oaxaca-Blinder/reweighting/decomposition

7.5 What this course estimated

Model 29 is a decomposition of the gender gap. Applied to attendance among 13–18 year olds, the gap is +3.0 pp in favour of boys (74.0 % against 70.9 %), on 6,205 boys and 6,126 girls.

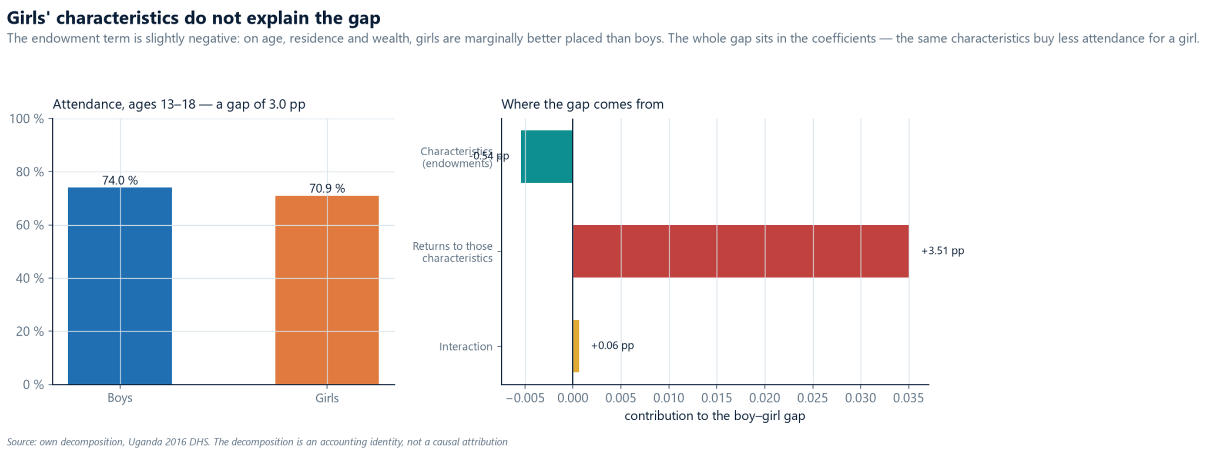


Figure 16. The secondary-age gender gap, split into what girls bring and what the system does with it. — Uganda 2016 DHS, Oaxaca–Blinder on a weighted linear probability model

Component	Contribution	What it means
Endowments	-0.5 pp	What the gap would be if boys and girls had the same returns and differed only in age, residence and wealth. It is negative: girls' characteristics are marginally *better*.
Coefficients	+3.5 pp	What the gap would be if they had the same characteristics and differed only in what those characteristics buy. It is the whole of the gap.
Interaction	+0.1 pp	The residual cross term.
Total	**+3.0 pp**	The identity closes exactly, which is the check.

WHAT THE COEFFICIENT TERM IS NOT
It is not discrimination measured. It is the part of the gap that the included covariates do not account for — omitted variables, functional form and measurement error all live in it. Add one covariate and it moves. The honest sentence is 'the gap is not explained by age, residence or wealth', and that sentence is already a useful finding.

7.5.1 Inequality of attainment

Model 36's machinery applies to attainment as readily as to learning. Years of schooling among adults aged 20 and over have a weighted Gini of 0.4053 around a mean of 6.47 years.

A Gini of that size on a quantity bounded below at zero is driven mostly by the share of adults with none at all — 13.3 % of those aged 15 and over. Inequality of schooling in this setting is not mainly about how far the educated get; it is about how many never start.

7.5.2 The rural–urban gap

Indicator	Urban	Rural	Gap
Primary net attendance	82.6 %	77.9 %	+4.7 pp
Secondary net attendance	31.7 %	14.3 %	+17.4 pp
Out of school, primary age	7.1 %	11.3 %	-4.2 pp
Over-age for grade	45.5 %	65.6 %	-20.1 pp
No education, 15+	7.6 %	15.2 %	-7.6 pp

Table 9. Uganda 2016 DHS (UG2016DHS), model 30

7.6 What this course did not estimate

Model	Why not
25–26. Child labour and work hours	The health-and-demography recode has no child labour module. The household panel's labour module starts at an age above most child work, and its hours question is asked of the main job only.
27. Household shock and resilience	Needs a shock module linked across waves. The panel has both; wiring them is the natural extension.
28. Disability and exclusion	The functioning questions are not in the recodes used here.
31. Intergenerational mobility	Needs parent and child attainment on the same record. The roster supports it where both live in the household, which truncates the sample towards younger children.

7.8 Every model in this family the course does not estimate

With the input each one needs that is not present. This is the family's data request, and it is more useful than a proxy.

#	Model	Dependent variable	What it needs that is missing
25	Child labour versus schooling multinomial model	School only, work only, school+work, neither	age; sex; poverty; farm ownership; season; household labour demand; distance; parent education
26	Child work-hours model	Weekly child work hours	enrolment; farm season; household enterprise; poverty; siblings; gender
27	Household shock and schooling resilience model	Enrolment, attendance, dropout or expenditure	shock exposure; baseline poverty; assets; social protection; location
28	Disability and education exclusion model	Enrolment, attendance, attainment or completion	disability domains; age; sex; poverty; access; region
31	Intergenerational education mobility model	Child education years/level	mother education; father education; child cohort; sex; location; wealth
32	Sibling competition / resource dilution model	Education expenditure, enrolment or attainment	number of children; birth order; age spacing; income; parent education

7.7 What this family feeds

Targeting rules, equity floors in the optimisation layer, and the deprivation index.

8. D. Learning Outcomes and School Quality

Specification chapter	8
Models	33–42 (10 cards)
Estimated in this course	0 on microdata
Feeds	Learning-adjusted years of schooling, teacher policy and the infrastructure investment case.

8.1 What this family estimates

What inputs produce learning, which schools add value, and how unequal learning is.

8.2 The trap

READ THIS BEFORE THE MODELS
This course estimates none of these ten models, because no pupil-level assessment data exists on any platform it reaches. The inputs are all available and the outcome is not, which is a different problem from a model that fails.

8.3 The models

#	Model	Dependent variable	Method	Status
33	Learning production function	Standardized test score	Multilevel linear model	Specified
34	Foundational literacy probability model	Binary literacy proficiency	Logit/multilevel logit	Specified
35	Foundational numeracy probability model	Binary numeracy proficiency	Logit/multilevel logit	Specified
36	Learning inequality decomposition	Test score distribution	Variance decomposition, Theil-type indices, quantile gaps	Specified
37	Teacher effectiveness/ value-added model	Achievement gain	Teacher value-added model with shrinkage/ empirical Bayes	Specified
38	Class-size effect model	Test score	Multilevel regression	Specified
39	Textbook and learning-material effect model	Test score/literacy/ numeracy	Multilevel regression	Specified
40	School infrastructure quality index	Composite school infrastructure score	PCA/factor analysis or transparent policy weights	Specified
41	School quality latent-factor model	Latent quality factor	Confirmatory factor analysis / item-response style latent model	Specified
42	Learning-adjusted years of schooling	Learning-adjusted schooling years	Deterministic index with uncertainty propagation	Specified

Table 10. Family D — all 10 cards

8.4 The specifications that carry the family

Model 33. Learning production function

Estimate how student, household, teacher and school inputs combine to produce learning outcomes.

Model 33
Score_is=α+X_iβ+S_sγ+u_s+ε_is
Standardized test score
(45)

Key inputs	prior achievement if available; household wealth; attendance; teacher qualifications; class size; books; electricity; water/ sanitation
Method	Multilevel linear model; value-added specification with lagged score when panel tests exist.
Diagnostics	ICC, residuals, school-effect stability, omitted-variable sensitivity.
Outputs	Input elasticities and predicted learning.

Key inputs	prior achievement if available; household wealth; attendance; teacher qualifications; class size; books; electricity; water/ sanitation
Integration	Links infrastructure and teacher policy to learning outcomes.

Model 37. Teacher effectiveness/value-added model

Estimate teacher contribution to student learning where repeated scores and teacher links exist.

Model 37
Score_it=α+ρ Score_i,t-1+θ_teacher+X_iβ+ε_it
Achievement gain
(46)

Key inputs	lagged score; teacher id; student covariates; school fixed effects
Method	Teacher value-added model with shrinkage/empirical Bayes.
Diagnostics	Stability across years, classroom sorting tests, minimum sample sizes.
Outputs	Teacher-effect distributions with uncertainty.
Integration	Supports teacher development, not punitive ranking.

Model 40. School infrastructure quality index

Construct a multidimensional measure of school infrastructure.

Model 40
Infra_s=Σ_k w_k z_{ks}
Composite school infrastructure score
(47)

Key inputs	classroom condition; WASH; electricity; internet; laboratory; library; furniture
Method	PCA/factor analysis or transparent policy weights.
Diagnostics	Internal consistency, loading stability, missingness.
Outputs	Infrastructure index and component gaps.
Integration	Used in learning, access and capital-budget models.

Model 42. Learning-adjusted years of schooling

Combine schooling quantity and learning quality into a human-capital metric.

Model 42
LAYS=ExpectedYearsSchooling × (LearningIndex/Benchmark)
Learning-adjusted schooling years
(48)

Key inputs	expected schooling years; standardized learning score; benchmark
Method	Deterministic index with uncertainty propagation.
Diagnostics	Benchmark sensitivity and score comparability.
Outputs	Learning-adjusted schooling by cohort/region.
Integration	Stronger human-capital input for growth/CGE than years alone.

Model 34. Foundational literacy probability model

Estimate probability of meeting a literacy proficiency threshold.

Model 34
Pr(Literate_i=1)=F(X_iβ+S_sγ)
Binary literacy proficiency
(49)

Method. Logit/multilevel logit

Model 35. Foundational numeracy probability model

Estimate probability of meeting a numeracy proficiency threshold.

Model 35
Pr(Numerate_i=1)=F(X_iβ+S_sγ)
Binary numeracy proficiency
(50)

Method. Logit/multilevel logit

Model 36. Learning inequality decomposition

Measure dispersion in learning and attribute it to between- and within-group components.

Model 36
Var(Y)=Var(E[Y|G])+E[Var(Y|G)]
Test score distribution
(51)

Method. Variance decomposition, Theil-type indices, quantile gaps

Model 38. Class-size effect model

Estimate association or causal effect of class size on learning.

Model 38
Score_is=α+β ClassSize_s+X_iγ+ε_is
Test score
(52)

Method. Multilevel regression; IV/RD only when a credible rule creates exogenous class-size variation

Model 39. Textbook and learning-material effect model

Estimate learning differences associated with access to textbooks/materials.

Model 39
Y_i=α+β Materials_i+X_iγ+S_sδ+ε_i
Test score/literacy/numeracy
(53)

Method. Multilevel regression; program evaluation design if applicable

Model 41. School quality latent-factor model

Estimate a latent school-quality measure from multiple observed outcomes and inputs.

Model 41
y_{ks}=λ_k Q_s+ε_{ks}
Latent quality factor
(54)

Method. Confirmatory factor analysis / item-response style latent model

8.5 What this course estimated: none of them

This is the shortest results section in the manual and the most important one to read carefully. Every model in this family has a dependent variable that is a measured learning outcome — a test score, a proficiency threshold, a value-added residual. No pupil-level assessment data exists on any platform this course can reach.

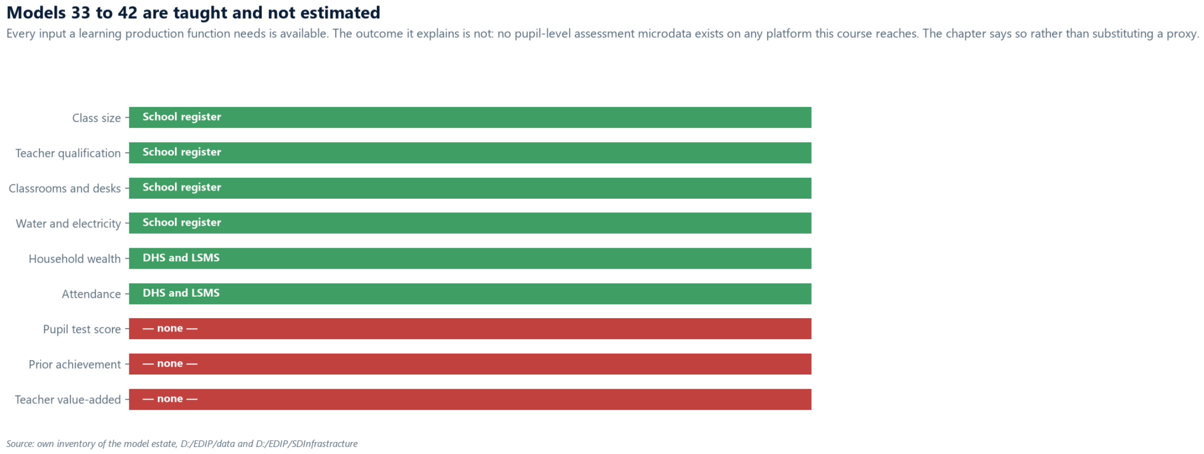


Figure 17. The learning chapter's right-hand side exists; its left-hand side does not. — Own inventory of every data platform this course reaches

Source examined	What it holds
model country bundles learning.csv	rows 54, note one aggregate row per country-region-year, no pupil-level scores
learning_assessments raw	files 54

No pupil-level assessment microdata is present in any platform this course reaches. Models 33 to 42 are taught as specifications with the data they would require, and the chapter states plainly that it estimates none of them. The school-quality inputs those models need - class size, teacher qualification, textbooks, infrastructure - are available from the school register, so the right-hand side exists and the left-hand side does not.

8.6 Why this is not the same as a model that failed

A model that fails has been fitted and produced something unusable — separation, non-convergence, a coefficient with the wrong sign. You learn something from it. A model that cannot be fitted because its outcome does not exist teaches something different and more useful: it tells you exactly what to go and collect.

What model 33 needs	Available?	Where
Standardised test score	**No**	— nowhere —
Prior achievement	**No**	— nowhere —
Household wealth	Yes	Both surveys
Attendance	Yes	Both surveys
Teacher qualification	Partly	School register carries teacher counts, not qualifications
Class size	Yes	School register: pupils, teachers and classrooms per school
Books and materials	Partly	The register carries desks; the country bundles carry a materials index
Electricity and water	Yes	School register carries both as indices

THE ONE SENTENCE THIS CHAPTER EXISTS TO TEACH
Every right-hand-side variable a learning production function needs is available. The left-hand side is not. That is a procurement statement, not a modelling failure, and the correct output of this chapter is a data request rather than an estimate.

8.7 What would change this

- A national learning assessment with pupil records, school identifiers and — for value added — a lagged score. Two rounds of the same pupils turn models 33 and 37 from specification into estimate.
- A regional assessment programme with a school identifier that joins to the register. That alone would let models 38, 39 and 41 run against real class sizes and real infrastructure.
- A household-survey literacy and numeracy module. It would not support a production function, but it would support models 34 and 35 at the population level, which is what an SDG indicator needs.

Until one of those arrives, this family is taught as specification and the estate's learning module runs on assumed parameters. Chapter 15 says the same thing from the engine's side.

8.8 Every model in this family the course does not estimate

With the input each one needs that is not present. This is the family's data request, and it is more useful than a proxy.

#	Model	Dependent variable	What it needs that is missing
33	Learning production function	Standardized test score	prior achievement if available; household wealth; attendance; teacher qualifications; class size; books; electricity; water/sanitation
34	Foundational literacy probability model	Binary literacy proficiency	age; grade; attendance; home language; books at home; teacher/school inputs
35	Foundational numeracy probability model	Binary numeracy proficiency	grade; attendance; household resources; teacher maths training; class size
36	Learning inequality decomposition	Test score distribution	scores; school; district; wealth quintile; gender
37	Teacher effectiveness/value-added model	Achievement gain	lagged score; teacher id; student covariates; school fixed effects
38	Class-size effect model	Test score	class size; grade; teacher quality; school resources; student background
39	Textbook and learning-material effect model	Test score/literacy/numeracy	textbook access; student-book ratio; grade; household and school inputs
40	School infrastructure quality index	Composite school infrastructure score	classroom condition; WASH; electricity; internet; laboratory; library; furniture
41	School quality latent-factor model	Latent quality factor	learning; attendance; completion; teacher presence; infrastructure
42	Learning-adjusted years of schooling	Learning-adjusted schooling years	expected schooling years; standardized learning score; benchmark

8.8 What this family would feed

Learning-adjusted years of schooling, teacher policy and the infrastructure investment case.

9. E. Education, Employment, Earnings and Skills

Specification chapter	9
Models	43–52 (10 cards)
Estimated in this course	7 on microdata
Feeds	The CGE skilled-labour mapping, rate-of-return analysis and TVET planning.

9.1 What this family estimates

What schooling is worth in the labour market: employment, formality, occupation, earnings and the transition from school to work.

9.2 The trap

READ THIS BEFORE THE MODELS
In a country where most workers are unpaid on a household farm, a wage equation describes the one worker in seven who has a wage. Selection is the whole problem, and a selection correction without an exclusion restriction is not a correction.

9.3 The models

#	Model	Dependent variable	Method	Status
43	Mincer returns-to-schooling model	Log hourly or monthly wage	Survey-weighted OLS	**Estimated**
44	Returns by education level model	Log wage	OLS/quantile regression	**Estimated**
45	Heckman wage-selection model	Employment selection and conditional log wage	Heckman two-step or full-information ML	**Estimated**
46	Employment probability model	Employment status	Logit/probit with survey design	**Estimated**
47	Formal-employment probability model	Formal versus informal employment	Logit/probit	**Estimated**
48	Education-to-occupation multinomial model	Occupation group	Multinomial logit or nested occupational choice	**Estimated**
49	Education-to-industry allocation model	Industry of employment	Multinomial model or empirical transition matrix	Specified
50	Skills mismatch model	Mismatch category	Realized-matches approach, modal education or external occupation-skill mapping	Specified
51	Youth NEET model	NEET indicator	Survey-weighted logit/probit	**Estimated**
52	School-to-work transition duration model	Duration to employment	Survival model	Specified

Table 11. Family E — all 10 cards

9.4 The specifications that carry the family

Model 43. Mincer returns-to-schooling model

Estimate wage returns to each additional year of schooling.

Model 43
ln(w_i)=α+β SchoolYears_i+γ Exp_i+δ Exp_i^2+X_iθ+ε_i
Log hourly or monthly wage
(55)

Key inputs	schooling years; potential/actual experience; sex; region; occupation; industry; formal status
Method	Survey-weighted OLS; quantile regression for distributional heterogeneity.
Diagnostics	Heteroskedasticity, functional form, wage outliers and sample selection.
Outputs	Return per year of schooling and subgroup returns.
Integration	Converts education outcomes into household income and productivity effects.

Model 45. Heckman wage-selection model

Correct wage equations for non-random selection into observed wage employment.

Model 45
Pr(Employed_i=1)=Φ(Z_iγ); ln(w_i)=X_iβ+ρλ_i+ε_i
Employment selection and conditional log wage
(56)

Key inputs	education; experience; household composition; non-labour income; local labour conditions
Method	Heckman two-step or full-information ML.
Diagnostics	Exclusion restriction, inverse-Mills significance, sensitivity.
Outputs	Selection-corrected returns to schooling.
Integration	Improves labour-income microsimulation.

Model 47. Formal-employment probability model

Estimate probability that workers enter formal employment.

Model 47
Pr(Formal_i=1)=F(Edu_iβ+X_iγ)
Formal versus informal employment
(57)

Key inputs	education level; occupation; industry; urban/rural; sex; age
Method	Logit/probit.
Diagnostics	Definition sensitivity and sector interactions.
Outputs	Formalization premium associated with education.
Integration	Supports tax-base and productivity channels in CGE/fiscal models.

Model 51. Youth NEET model

Estimate probability that youth are not in employment, education or training.

Model 51
Pr(NEET_i=1)=F(Edu_iβ+Household_iγ+Location_iδ)
NEET indicator
(58)

Key inputs	age; sex; education; wealth; marital status; disability; local labour conditions
Method	Survey-weighted logit/probit.
Diagnostics	Calibration and age-band sensitivity.
Outputs	NEET risk and geographic concentration.
Integration	Connects education, youth employment and social policy.

Model 44. Returns by education level model

Estimate wage premia for primary, secondary, TVET and tertiary education.

Model 44
ln(w_i)=α+Σ_e β_e D_{ie}+X_iγ+ε_i
Log wage
(59)

Method. OLS/quantile regression

Model 46. Employment probability model

Estimate how education changes probability of employment.

Model 46
Pr(Employed_i=1)=F(Edu_iβ+X_iγ)
Employment status
(60)

Method. Logit/probit with survey design

Model 48. Education-to-occupation multinomial model

Estimate occupational destination conditional on education and demographics.

Model 48
Pr(Occupation_i=j)=exp(X_iβ_j)/Σ_k exp(X_iβ_k)
Occupation group
(61)

Method. Multinomial logit or nested occupational choice

Model 49. Education-to-industry allocation model

Estimate distribution of workers across economic sectors by education.

Model 49
Pr(Sector_i=s)=exp(X_iβ_s)/Σ_k exp(X_iβ_k)
Industry of employment
(62)

Method. Multinomial model or empirical transition matrix

Model 50. Skills mismatch model

Measure overeducation, undereducation and match relative to occupational requirements.

Model 50
Mismatch_i = compare(Edu_i, RequiredEdu_{occupation})
Mismatch category
(63)

Method. Realized-matches approach, modal education or external occupation-skill mapping

Model 52. School-to-work transition duration model

Estimate time between leaving education and first stable employment.

Model 52
h_i(t)=h_0(t)exp(Edu_iβ+X_iγ)
Duration to employment
(64)

Method. Survival model; competing risks for formal/informal employment

9.5 What this course estimated

Estimated on Uganda LSMS-ISA National Panel Survey 2019/20 (UNPS). The working-age sample is 8,039 people aged 15 to 64, of whom 70.1 % did some work. The composition of that work is the whole story of this family.

Quantity	Value
Employment rate, 15–64	70.1 %
Of workers: on a household farm	51.8 %
Of workers: in paid employment	27.8 %
Paid employees in the sample	1,297

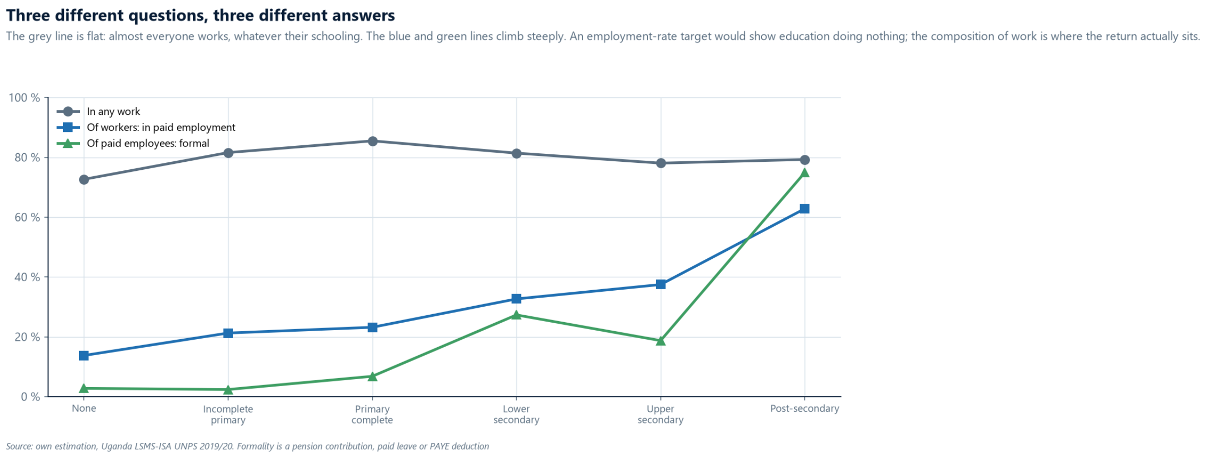


Figure 18. Education does not change whether you work. It changes what work you get. — Uganda LSMS-ISA UNPS 2019/20, own estimation

EG-09 — THE SELECTION PROBLEM IS THE SUBJECT, NOT A CAVEAT
A wage equation can only be estimated on people who have a wage. That is 14.3 % of the working-age sample. Most workers here are unpaid on a household farm and have no wage to explain. Every number in the next two sections is a statement about paid employment, and saying so is not modesty — it is the definition of what was estimated.

9.5.1 The Mincer equation

Term	Coefficient	Standard error
(Intercept)	5.4803	0.1183
Years of schooling	0.1183	0.0074
Experience	0.0626	0.0088
Experience squared	-0.0007	0.0002
Female	-0.2854	0.0627
Urban	-0.0349	0.0603

Table 12. Log hourly earnings. n = 914 (trimmed from 933 at the 1st and 99th percentiles), R^2 = 0.290

The return to a year of schooling is 11.8 % within paid employment, with the familiar concave experience profile and a female coefficient of -0.285 log points.

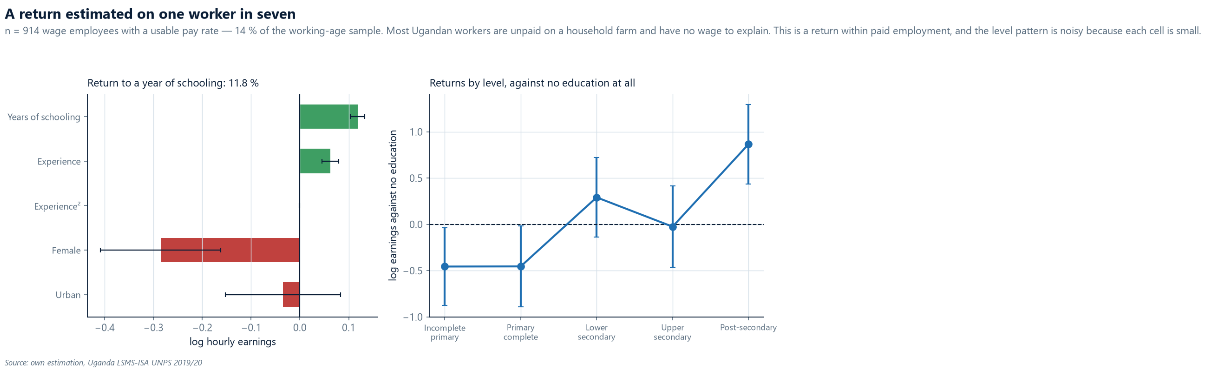


Figure 19. The Mincer equation on real Ugandan wage data. — Uganda LSMS-ISA UNPS 2019/20, own estimation

9.5.2 The selection correction, run twice

Model 45 is the two-step selection correction. It is worth running twice, because the difference between the two runs is the single clearest demonstration in this manual of what identification means.

Specification	Return per year	Inverse Mills coefficient	Excluded from the wage equation
Least squares, no correction	11.8 %	—	—
Two-step, no exclusion restriction	34.3 %	5.90 (se 1.65)	nothing
Two-step, with an exclusion restriction	11.3 %	-0.12 (se 0.27)	household size, children under five

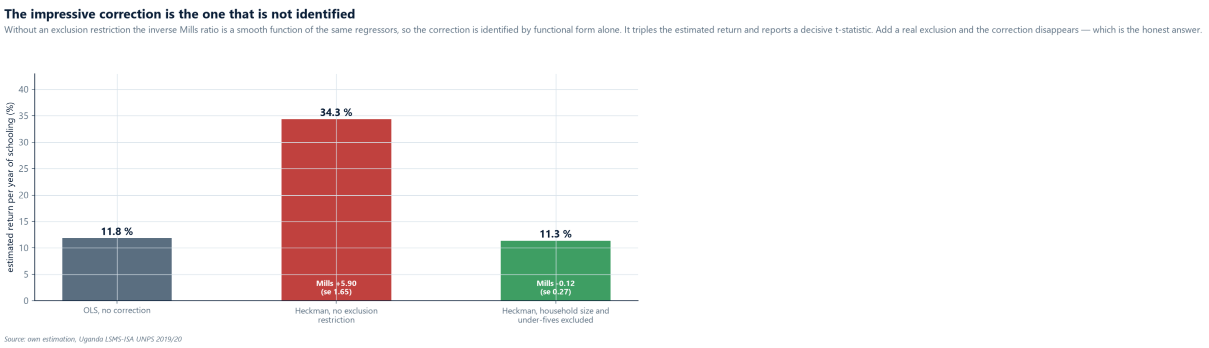


Figure 20. What an exclusion restriction is worth. — Uganda LSMS-ISA UNPS 2019/20, own estimation

EG-10 — THE IMPRESSIVE RESULT IS THE UNIDENTIFIED ONE
With the same regressors in both equations, the inverse Mills ratio is a smooth function of the covariates already in the wage equation. Identification then rests entirely on the non-linearity of the normal hazard — which is to say, on functional form. The correction triples the estimated return and reports a decisive t-statistic, and it means nothing.

Add two variables that plausibly move the decision to take paid work but not the wage once in it, and the correction collapses to nothing. That is the honest answer: on this sample, selection is not detectably biasing the return.

9.5.3 Returns by level, and why they are not monotone

Level	Coefficient against no education	Standard error
Incomplete primary	-0.4555	0.2141
Primary complete	-0.4541	0.2228
Lower secondary	0.2919	0.2201
Upper secondary	-0.0246	0.2235
Post-secondary	0.8666	0.2192

Table 13. n = 914, reference is no education

Post-secondary is far above everything else and the middle of the ladder is noisy, with upper secondary sitting below lower secondary. Each cell is a few hundred people at most and the standard errors overlap heavily. The right reading is that the return is concentrated at the top of the ladder, not that upper secondary is worth less than lower secondary.

9.5.4 What schooling actually changes

Level	In any work	Of workers: paid employment	Of paid employees: formal	n
None	72.6 %	13.8 %	2.8 %	567
Incomplete primary	81.6 %	21.2 %	2.4 %	2,791
Primary complete	85.5 %	23.2 %	6.8 %	930
Lower secondary	81.4 %	32.7 %	27.3 %	872
Upper secondary	78.1 %	37.5 %	18.7 %	672
Post-secondary	79.3 %	62.8 %	74.9 %	433

The first column is nearly flat. The second and third climb steeply. An education strategy evaluated against an employment rate would conclude that schooling does nothing; evaluated against the composition of employment it finds a large effect. Which target a ministry writes into its results framework decides what its own evidence will appear to say.

9.5.5 Youth not in education, employment or training

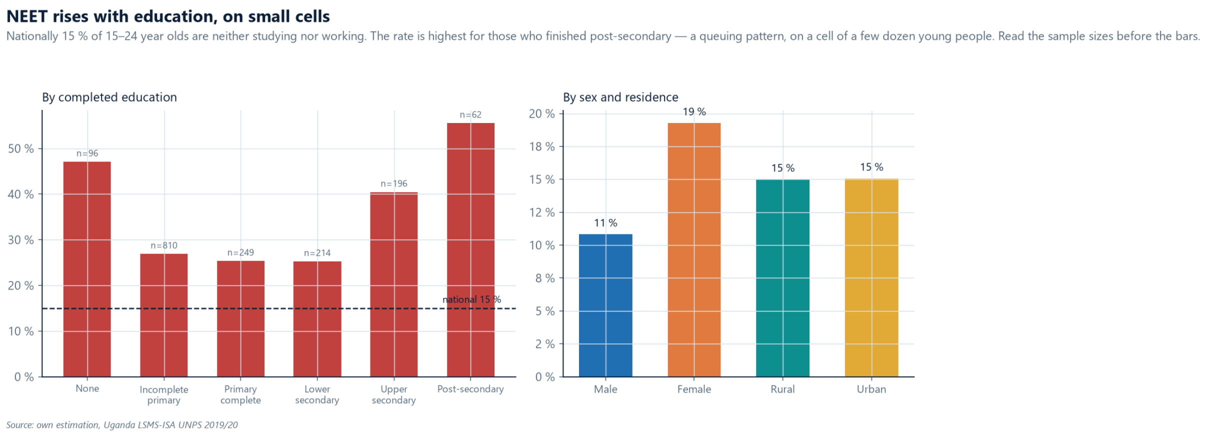


Figure 21. Youth not in education, employment or training. — Uganda LSMS-ISA UNPS 2019/20, own estimation

15.0 % of 15–24 year olds are neither studying nor working, on 3,306 young people. The rate rises with completed education, which is the queuing pattern — but the post-secondary cell is a few dozen people and the interval around it is wide. Read the sample sizes printed on the figure before the bars.

9.5.6 Education to occupation

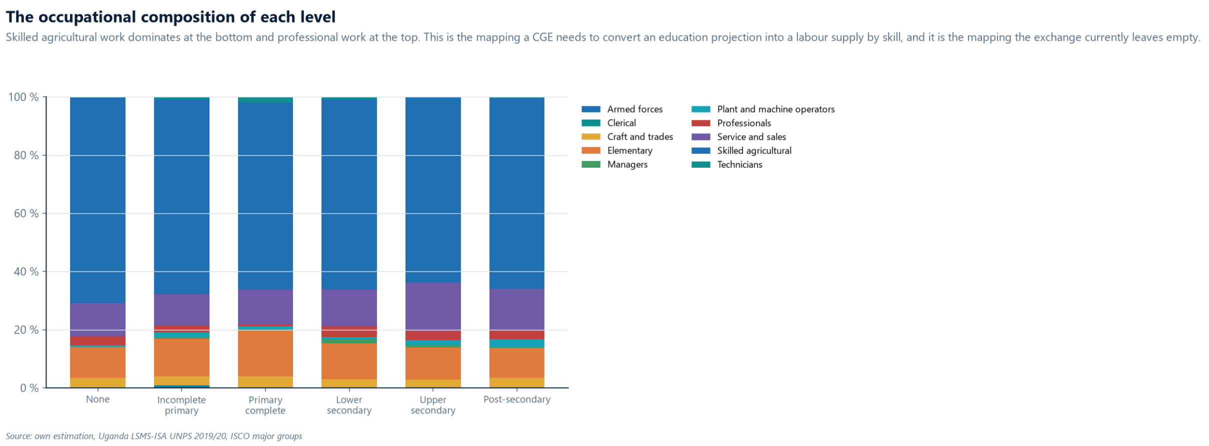


Figure 22. Where each education level ends up working. — Uganda LSMS-ISA UNPS 2019/20, own estimation

Model 48 maps education level to occupational destination. This is the table a computable general equilibrium model needs in order to convert an education projection into a labour supply by skill — and it is exactly the field the estate's exchange contract currently leaves empty. Chapter 22 returns to it.

9.6 What this course did not estimate

Model	Why not
49. Education to industry	The industry codes are present; the cell sizes by education level and industry fall below the disclosure threshold almost everywhere.
50. Skills mismatch	Needs a required-qualification measure for the job, which the module does not ask.
52. School-to-work transition duration	Needs the date of leaving school and the date of first job. Neither is asked.

9.8 Every model in this family the course does not estimate

With the input each one needs that is not present. This is the family's data request, and it is more useful than a proxy.

#	Model	Dependent variable	What it needs that is missing
49	Education-to-industry allocation model	Industry of employment	education; occupation; region; sex; age
50	Skills mismatch model	Mismatch category	education level; occupation; wage; industry
52	School-to-work transition duration model	Duration to employment	education; sex; location; field of study; household resources

9.7 What this family feeds

The CGE skilled-labour mapping, rate-of-return analysis and TVET planning.

10. F. Spatial Education Planning and Optimization

Specification chapter	10
Models	53–61 (9 cards)
Estimated in this course	3 on microdata
Feeds	The capital programme, teacher deployment and the MTEF capital line.

10.1 What this family estimates

Where the schools are, who can reach them, where the capacity gap is, and where the next school should go.

10.2 The trap

READ THIS BEFORE THE MODELS
Distance and crowding are different problems and they are not in the same places. A single 'underserved' ranking sends the wrong intervention to whichever county has only one of them.

10.3 The models

#	Model	Dependent variable	Method	Status
53	Nearest-school accessibility model	Network travel time/ distance	GIS network analysis / least-cost path	**Estimated**
54	Gravity-based education accessibility index	Continuous accessibility score	Spatial gravity model with calibrated decay	Specified
55	School catchment demand model	Potential students in catchment	GIS catchment overlay or Voronoi/ network catchments	**Estimated**
56	School capacity-gap model	Capacity surplus/ deficit	Deterministic calculation with scenario thresholds	**Estimated**
57	Optimal new-school location model	Binary site decisions and assignment	p-median, maximal covering or capacitated facility-location MILP	Illustrated
58	Equity-constrained school siting model	Site decisions with equity constraints	Multi-objective MILP	Specified
59	Teacher allocation optimization	Teacher assignments	Linear/MILP approximation of staffing objective	Specified
60	Subject-teacher allocation model	Teacher-subject-school assignment	Assignment/MILP	Specified
61	School consolidation/ expansion model	Facility investment decisions	Multi-period facility planning optimization	Specified

Table 14. Family F — all 9 cards

10.4 The specifications that carry the family

Model 53. Nearest-school accessibility model

Measure travel distance/time from settlements or households to nearest school by level.

Model 53
Access_i=min_j TravelTime(i,j)
Network travel time/distance
(65)

Key inputs	road/path network; school coordinates; settlement/ population grid; terrain where available
Method	GIS network analysis / least-cost path.
Diagnostics	Coordinate validation, road-network completeness and spot checks.
Outputs	Travel time to nearest primary/secondary/TVET school.
Integration	Core spatial determinant used in enrolment and siting models.

Model 54. Gravity-based education accessibility index

Measure access to multiple schools weighted by capacity and travel impedance.

Model 54
A_i=Σ_j Capacity_j exp(-λ T_ij)
Continuous accessibility score
(66)

Key inputs	school capacity; travel time matrix; decay parameter
Method	Spatial gravity model with calibrated decay.
Diagnostics	Sensitivity to λ and capacity definitions.
Outputs	Accessibility surfaces and population-weighted district scores.
Integration	Improves on nearest-facility indicators.

Model 56. School capacity-gap model

Compare potential or actual demand with seats/classroom capacity.

Model 56
Gap_s=Demand_s-Capacity_s
Capacity surplus/deficit
(67)

Key inputs	enrolment; classrooms; target class size; catchment population
Method	Deterministic calculation with scenario thresholds.
Diagnostics	Capacity-definition sensitivity.
Outputs	Seat/classroom deficits by school/district.
Integration	Prioritizes capital investments.

Model 57. Optimal new-school location model

Choose school sites that maximize coverage/minimize travel subject to a capital budget.

Model 57
min Σ_i Pop_i d_{i,j(i)} subject to Σ_j Cost_j x_j≤B
Binary site decisions and assignment
(68)

Key inputs	candidate sites; population demand; travel-time matrix; site costs; existing schools
Method	p-median, maximal covering or capacitated facility-location MILP.
Diagnostics	Optimality gap, alternative budgets, site robustness.
Outputs	Recommended school locations, beneficiaries and cost.
Integration	Flagship EDIS infrastructure optimization application.

Model 55. School catchment demand model

Estimate school-age population potentially served by each school.

Model 55
Demand_s=Σ_{i∈C_s} PopSchoolAge_i
Potential students in catchment
(69)

Method. GIS catchment overlay or Voronoi/network catchments

Model 58. Equity-constrained school siting model

Ensure new facilities improve access for poor/remote groups, not only total coverage.

Model 58
min Cost/Travel subject to Coverage_poor≥τ and Coverage_remote≥κ
Site decisions with equity constraints
(70)

Method. Multi-objective MILP

Model 59. Teacher allocation optimization

Allocate teachers across schools to reduce pupil-teacher disparities.

Model 59
min Σ_s |Students_s/T_s - PTR*| subject to Σ_s T_s=T_available
Teacher assignments
(71)

Method. Linear/MILP approximation of staffing objective

Model 60. Subject-teacher allocation model

Allocate secondary teachers by subject to meet timetable and curriculum requirements.

Model 60
min Σ_{s,k} Shortage_{sk} subject to teacher qualification and workload constraints
Teacher-subject-school assignment
(72)

Method. Assignment/MILP

Model 61. School consolidation/expansion model

Choose whether to expand, rehabilitate, consolidate or close facilities under demographic change.

Model 61
min TotalCost + λ TravelBurden subject to capacity and access standards
Facility investment decisions
(73)

Method. Multi-period facility planning optimization

10.5 What this course estimated

Three of this family's models can be estimated from the geography this course reaches: nearest-school accessibility, catchment demand and the capacity gap. The siting programme itself is a mixed-integer programme in the estate's optimisation layer, and chapter 21 runs it; this chapter prepares its inputs.

10.5.1 Where the schools are

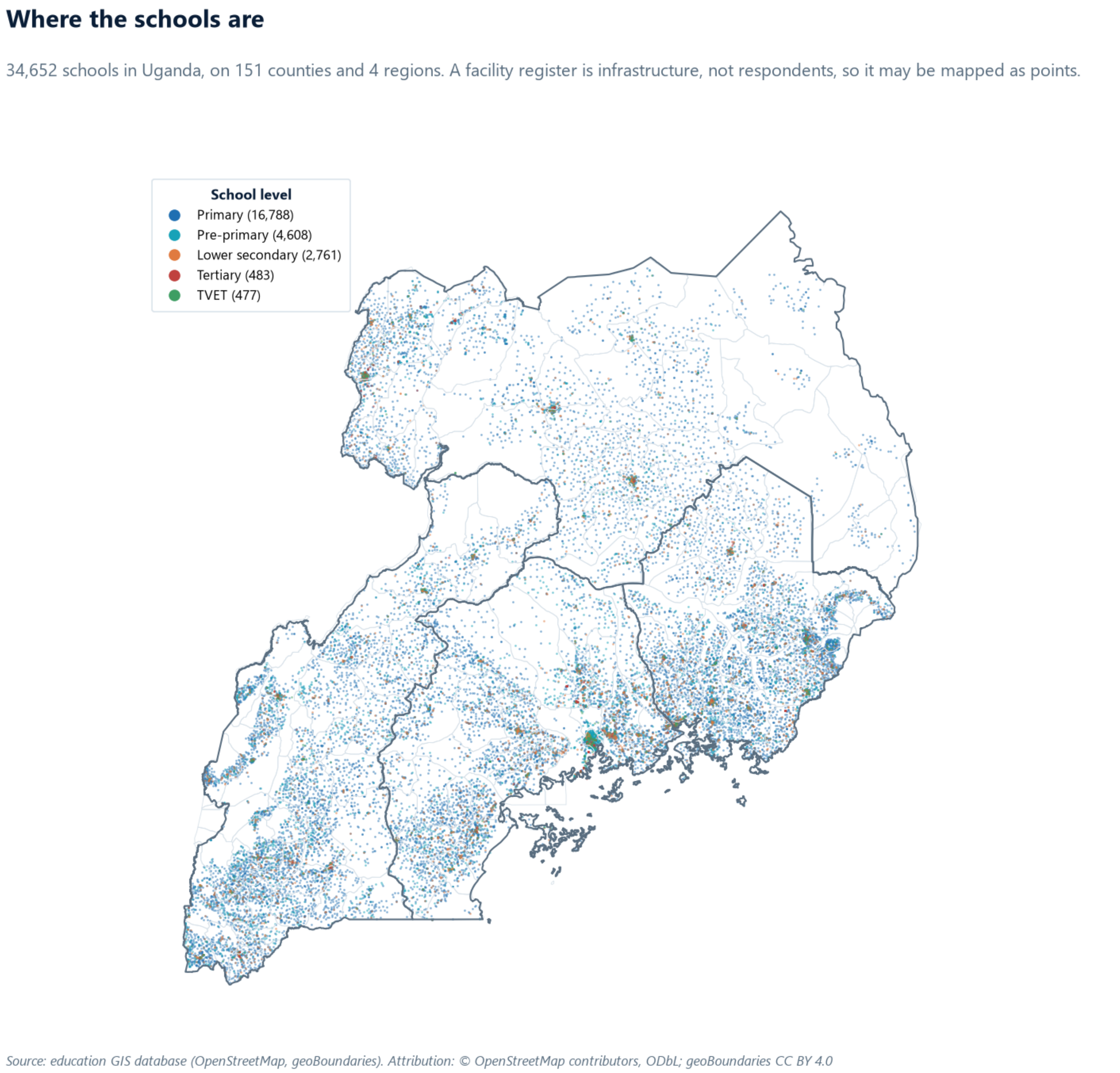


Figure 23. Every school the register places, by level. — Uganda school register (EMIS) and OpenStreetMap, via the education GIS database

Level	Schools
primary	16,788
unknown	8,988
pre primary	4,608
lower secondary	2,761
tertiary	483
tvet	477
secondary	342
upper secondary	205

Table 15. 34,652 schools placed by the education GIS database

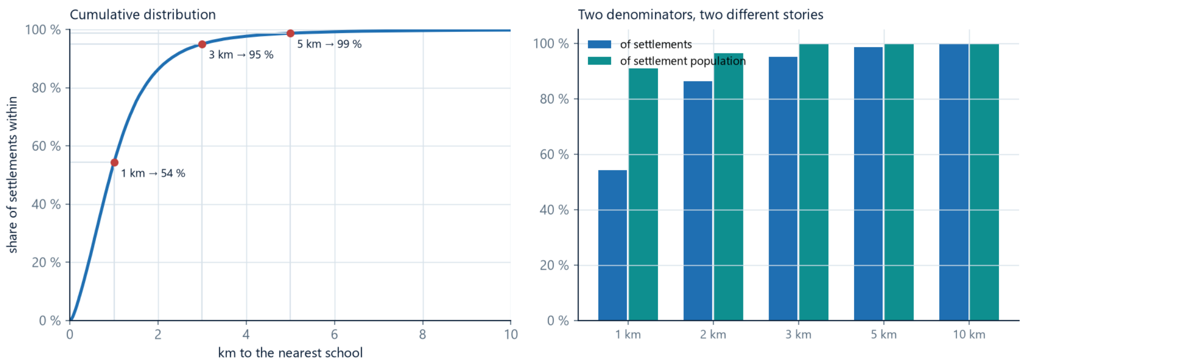
10.5.2 Model 53 — nearest-school accessibility

Distance was computed for 80,160 settlements to the nearest school of any level. The median settlement is 0.92 km from a school and the population-weighted mean is 0.44 km.

Threshold	Share of settlements	Share of settlement population
within 1 km	54.3 %	91.0 %
within 2 km	86.2 %	96.5 %
within 3 km	95.0 %	99.8 %
within 5 km	98.7 %	100.0 %
within 10 km	99.8 %	100.0 %

Coverage looks better when you weight by people

Settlements are scattered; people are not. Weighting by population moves the 1 km figure from 54 % to 91 %. Which denominator is right depends on whether you are siting schools or counting children.



Source: education GIS database. Distances are straight-line and the population field is the settlement register's own, which sums to well below the national total

Figure 24. The coverage curve: how much of the country each distance threshold buys. — Education GIS database, straight-line distance

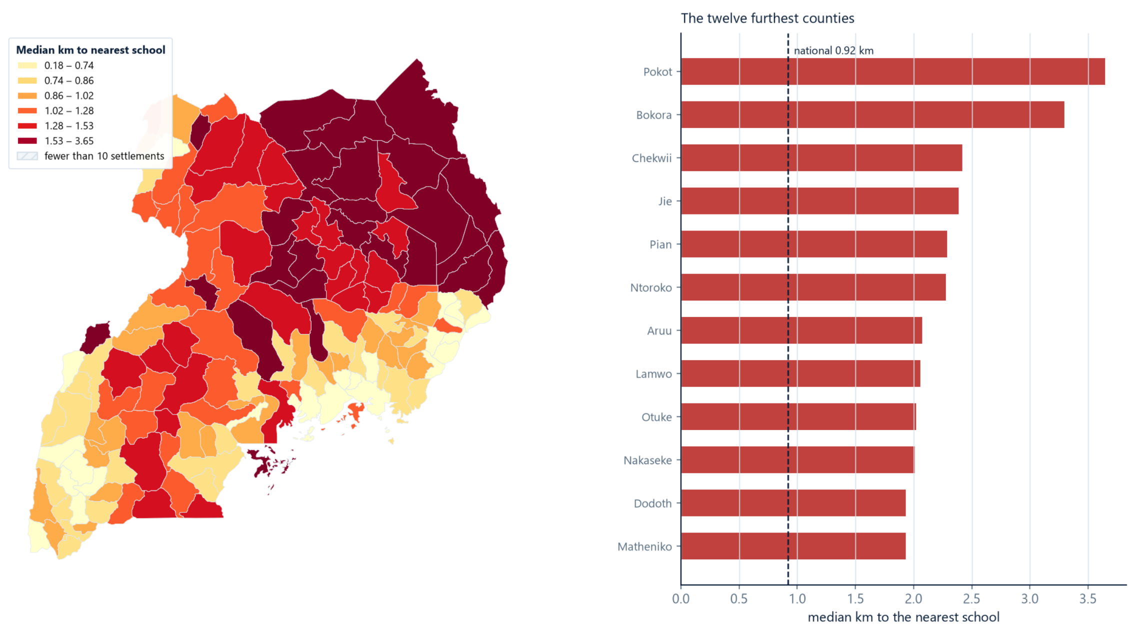
TWO DENOMINATORS, TWO DIFFERENT ANSWERS

Settlements are scattered and people are not. Weighting by population moves the one-kilometre figure from 54.3 % to 91.0 %. Which denominator is right depends on the decision: siting a school is about settlements, counting out-of-school children is about people. And the population field here sums to 5,570,285, well below the national total, so the second column is a share of the settlement population that carries a value (EG-14).

10.5.3 Where distance actually binds

Distance to school is a Karamoja problem

Median distance nationally is 0.92 km and 95 % of settlements are within 3 km. The counties in the north-east are three to four times the national median.



Source: education GIS database, straight-line distance from settlement to nearest school. Straight-line distance understates travel: it ignores roads, rivers and terrain

Figure 25. Median distance from a settlement to its nearest school. — Education GIS database, straight-line distance

County	Settlements	Median km	90th percentile km
Pokot	152	3.65	7.94
Bokora	268	3.30	10.28
Chekwii	155	2.42	4.15
Jie	180	2.39	5.75
Pian	64	2.29	8.93
Ntoroko	527	2.28	4.87
Aruu	655	2.07	4.05
Lamwo	384	2.06	5.02
Otuke	465	2.02	3.32
Nakaseke	400	2.01	5.49

Table 16. The ten counties furthest from a school

10.5.4 Model 56 — the capacity gap

The capacity gap converts recorded pupils into required classrooms at a norm class size and subtracts what exists.

Model 56
\mathrm{Required}_d = \frac{\mathrm{Pupils}_d}{\text{norm}}, \qquad \mathrm{Gap}_d = \max\!\left(0,\ \mathrm{Required}_d - \mathrm{Classrooms}_d\right)
classroom capacity gap
(74)

Quantity	Value
Norm class size (declared)	45
Classrooms recorded	90,803
Classrooms required at the norm	157,646
National gap	67,233
Districts in deficit	56

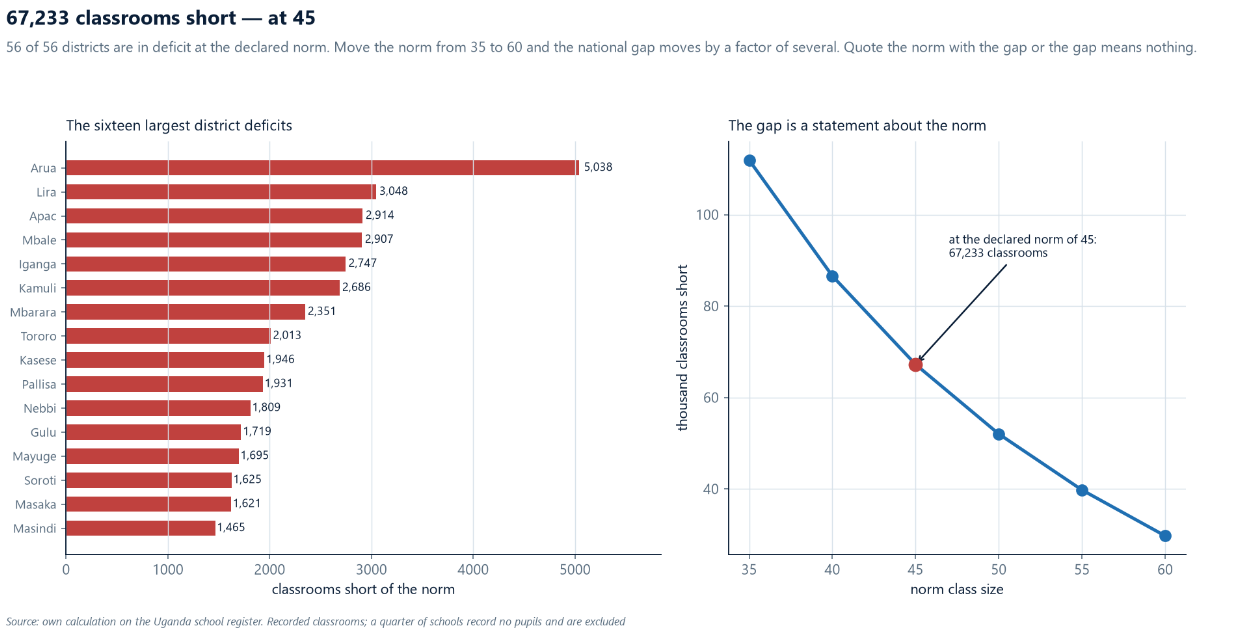


Figure 26. Classroom deficit by district at a norm of 45 pupils. — Uganda school register, own calculation

THE GAP IS A STATEMENT ABOUT THE NORM
Move the norm from 45 to 35 and the deficit grows by tens of thousands of classrooms; move it to 60 and much of it disappears. The norm is a policy parameter chosen by a ministry, not a property of the classrooms. A capacity gap quoted without it is not a planning number, and quoting one is an automatic fail in this course.

10.5.5 Crowding, and where it is

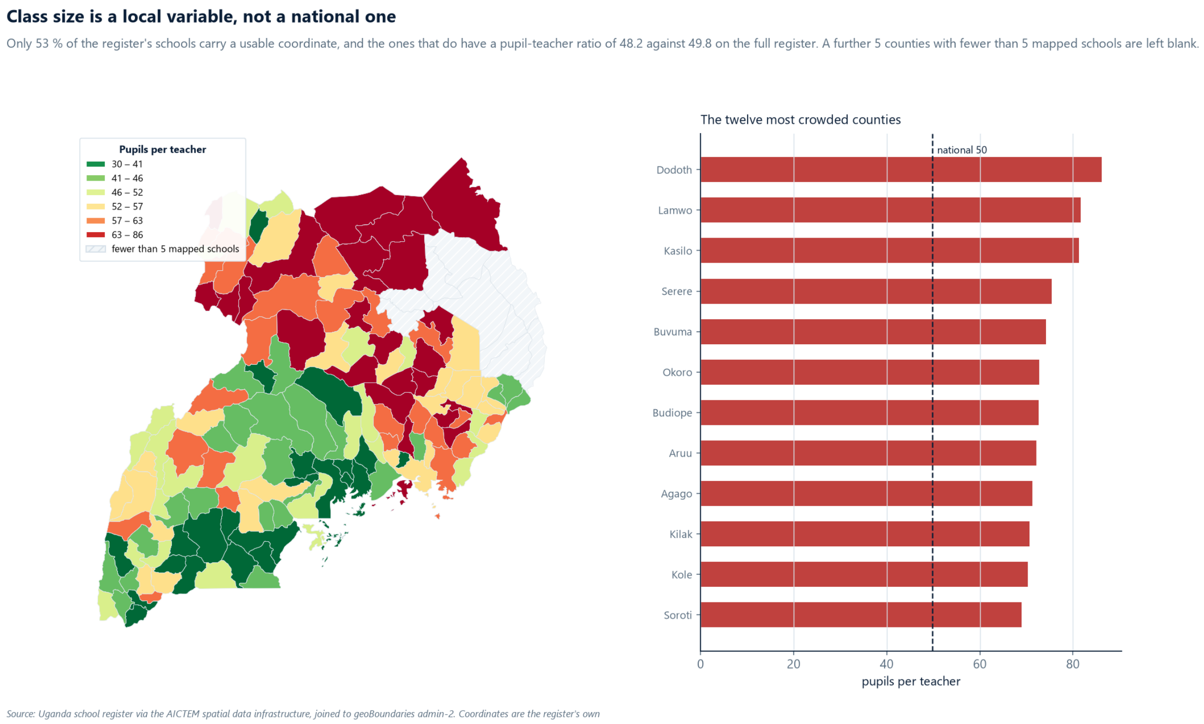


Figure 27. Pupil-teacher ratio by county, from the school register. — Uganda school register (EMIS), joined to admin-2 polygons

Quantity	Value
Aggregate pupil-teacher ratio	49.8
Aggregate pupil-classroom ratio	78.1
Median school pupil-teacher ratio	49.0
District range, pupil-teacher	29.5 to 81.9
District range, pupil-classroom	27.3 to 130.0

Only 53.5 % of pupil-bearing schools reach a county, and the schools that do are not a random subset: their aggregate pupil-teacher ratio is 48.2 against 49.8 on the full register, a bias of -1.63. A further 5 counties have fewer than 5 mapped schools and are left blank rather than shaded on a handful of records.

10.5.6 Distance and crowding are different problems

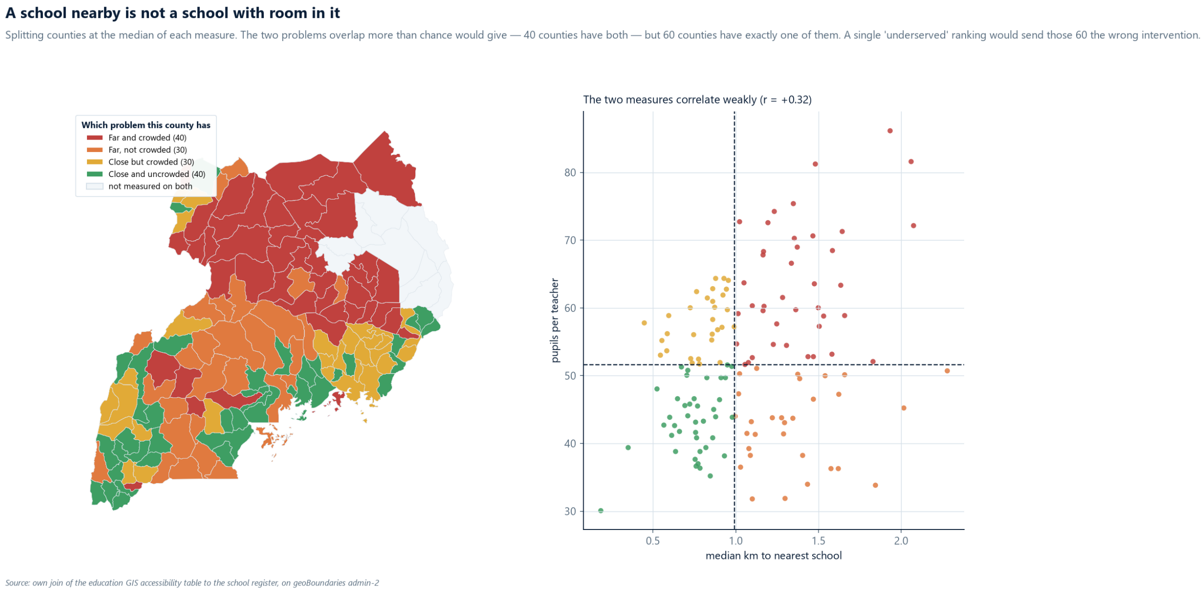


Figure 28. Distance and crowding are different problems in different places. — Education GIS database and the Uganda school register

Splitting counties at the median of each measure gives four groups, and the two problems overlap only partly. A county far from any school needs a school. A county with a school and eighty pupils per teacher needs teachers and classrooms. A single composite 'underserved' ranking collapses the two and sends whichever intervention the composite happened to weight more heavily.

10.5.7 An objective, and what it costs

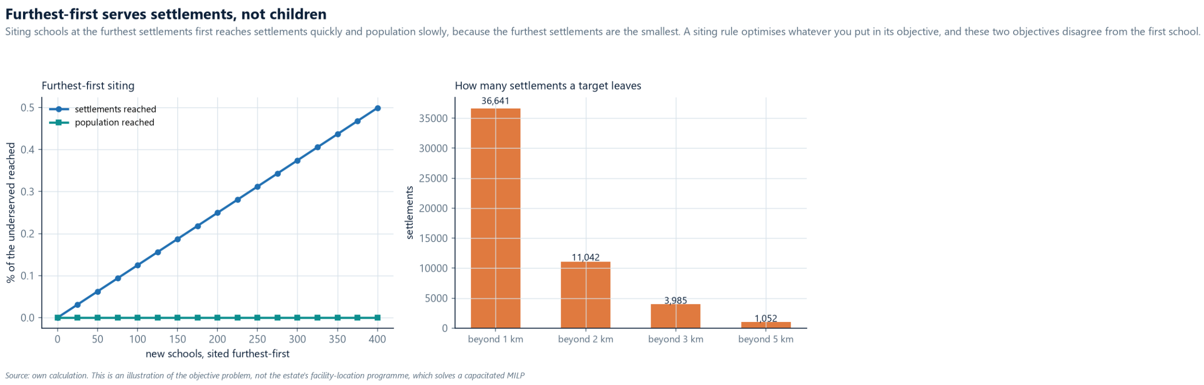


Figure 29. A siting sweep: how much a capital programme buys. — Own calculation on the Uganda accessibility surface

Siting schools furthest-settlement-first reaches settlements quickly and people slowly, because the furthest settlements are the smallest. Siting them largest-underserved-population-first does the reverse. Neither is wrong. The estate's facility location programme in chapter 21 solves a capacitated mixed-integer programme rather than a greedy rule, but it faces the same choice and it faces it explicitly, in the objective function.

10.6 What this course did not estimate

Model	Why not
54. Gravity accessibility index	Needs a distance-decay parameter. The estate implements the form; choosing the parameter is a calibration exercise that needs observed travel behaviour.
58–61. Equity-constrained siting, teacher and subject-teacher allocation, consolidation	Implemented in the optimisation layer and demonstrated in chapter 21 on the estate's country bundles, not estimated from microdata — they are programmes, not regressions.

EG-13 — EVERY DISTANCE IN THIS CHAPTER IS STRAIGHT-LINE
The estate implements a routable graph over a national road network and the spatial infrastructure holds derived travel surfaces. The table this course reads is straight-line distance. It ignores roads, rivers and terrain, and it always flatters access — most severely in exactly the places that are worst served.

10.8 Every model in this family the course does not estimate

With the input each one needs that is not present. This is the family's data request, and it is more useful than a proxy.

#	Model	Dependent variable	What it needs that is missing
54	Gravity-based education accessibility index	Continuous accessibility score	school capacity; travel time matrix; decay parameter
57	Optimal new-school location model	Binary site decisions and assignment	candidate sites; population demand; travel-time matrix; site costs; existing schools
58	Equity-constrained school siting model	Site decisions with equity constraints	poverty maps; remote-area flags; population; candidate sites
59	Teacher allocation optimization	Teacher assignments	enrolment; existing teachers; subject specialization; minimum staffing; transfer constraints
60	Subject-teacher allocation model	Teacher-subject-school assignment	teacher qualifications; subjects; classes; period requirements; locations
61	School consolidation/ expansion model	Facility investment decisions	projected population; school condition; operating costs; travel times; expansion costs

10.7 What this family feeds

The capital programme, teacher deployment and the MTEF capital line.

11. G. Education System Projection, Costing and Public Finance

Specification chapter	11
Models	62–68 (7 cards)
Estimated in this course	1 on microdata
Feeds	The MTEF envelope, the budget-allocation programme and the financing gap.

11.1 What this family estimates

How many pupils, teachers and classrooms the demography implies, what they cost, and who benefits from the spending.

11.2 The trap

READ THIS BEFORE THE MODELS
A unit cost and a norm class size are policy parameters, not measurements. The classroom gap is a statement about the norm as much as about the classrooms.

11.3 The models

#	Model	Dependent variable	Method	Status
62	Education cohort-flow projection model	Enrolment by grade and year	Deterministic cohort-component simulation with stochastic uncertainty option	From the run
63	Demographic education-demand model	School-age population and potential enrolment	Cohort-component demographic projection interface	From the run
64	Teacher requirement projection	Required teachers	Stock-flow projection	From the run
65	Classroom requirement projection	Required classrooms	Stock-flow capital needs model	**Estimated**
66	Education unit-cost model	Unit cost and cost per graduate	Accounting model with inflation/price normalization	From the run
67	Education expenditure projection model	Total education spending by level/year	Medium-term fiscal simulation	From the run
68	Benefit incidence of public education spending	Per-household/ quantile imputed education subsidy	Benefit-incidence analysis with concentration curves	Specified

Table 17. Family G — all 7 cards

11.4 The specifications that carry the family

Model 62. Education cohort-flow projection model

Project enrolment by grade using promotion, repetition, dropout and entry rates.

Model 62
E_{g,t+1}=Promoted_{g-1,t}+Repeaters_{g,t}+NewEntrants_{g,t+1}
Enrolment by grade and year
(75)

Key inputs	base enrolment; entry; promotion; repetition; dropout; population
Method	Deterministic cohort-component simulation with stochastic uncertainty option.
Diagnostics	Identity checks, historical backcast and transition consistency.
Outputs	Enrolment, entrants, repeaters, dropouts and graduates by year.
Integration	Core national education system projection engine.

Model 64. Teacher requirement projection

Project teachers needed to meet target pupil-teacher ratios and subject requirements.

Model 64
TeachersReq_{l,t}=Enroll_{l,t}/PTR_l
Required teachers
(76)

Key inputs	projected enrolment; target PTR; teacher attrition; training pipeline
Method	Stock-flow projection.
Diagnostics	Attrition and recruitment sensitivity.
Outputs	Annual teacher recruitment need and wage-bill implications.
Integration	Links education planning to fiscal module.

Model 66. Education unit-cost model

Estimate recurrent and capital cost per student and per graduate.

Model 66
UnitCost_l=Spending_l/Students_l; CostGraduate_l=Spending_l/Graduates_l
Unit cost and cost per graduate
(77)

Key inputs	budget/expenditure; students; graduates; teachers; capital spending
Method	Accounting model with inflation/price normalization.
Diagnostics	Classification reconciliation and denominator sensitivity.
Outputs	Unit costs by level and cost-efficiency ratios.
Integration	Inputs to budget projection and optimization.

Model 68. Benefit incidence of public education spending

Allocate public education subsidies across household welfare groups.

Model 68
Benefit_h=Σ_l Enrolment_{hl}×UnitSubsidy_l
Per-household/quantile imputed education subsidy
(78)

Key inputs	household enrolment; public/private status; unit subsidy by level; weights
Method	Benefit-incidence analysis with concentration curves.
Diagnostics	Incidence sensitivity to unit-subsidy assumptions.
Outputs	Share of subsidy accruing to each quintile and concentration index.
Integration	Evaluates progressivity of education budgets.

Model 63. Demographic education-demand model

Translate population projections into potential demand for each education level.

Model 63
PotentialEnroll_{l,t}=Pop_{age(l),t} × ParticipationRate_{l,t}
School-age population and potential enrolment
(79)

Method. Cohort-component demographic projection interface

Model 65. Classroom requirement projection

Estimate future classroom requirements by level and geography.

Model 65
ClassroomsReq_{l,t}=Enroll_{l,t}/TargetClassSize_l
Required classrooms
(80)

Method. Stock-flow capital needs model

Model 67. Education expenditure projection model

Project public education budget required under policy targets.

Model 67
Budget_t=Σ_l Students_{l,t}×UnitCost_{l,t}+CapitalNeed_t
Total education spending by level/year
(81)

Method. Medium-term fiscal simulation

11.5 What the committed run produced

This family is the system engine's own output rather than an estimate from microdata. The committed run reads a country bundle, projects the population, flows it through grades, derives a teacher and classroom requirement, and costs it.

Quantity	Value
Run identifier	edu_UGA_20260301T000000_664b64f1
Country	Uganda
Persons in the harmonised microdata	49,315
Total population	51,384,894
Base enrolment	8,840,589
Final gross enrolment ratio	1.060
Primary pupil-teacher ratio	42.66
Cumulative completers	5,666,411
Net present value of cost	12,304,687,708
Scenario	baseline
Scenario fingerprint	a4030be77b350176

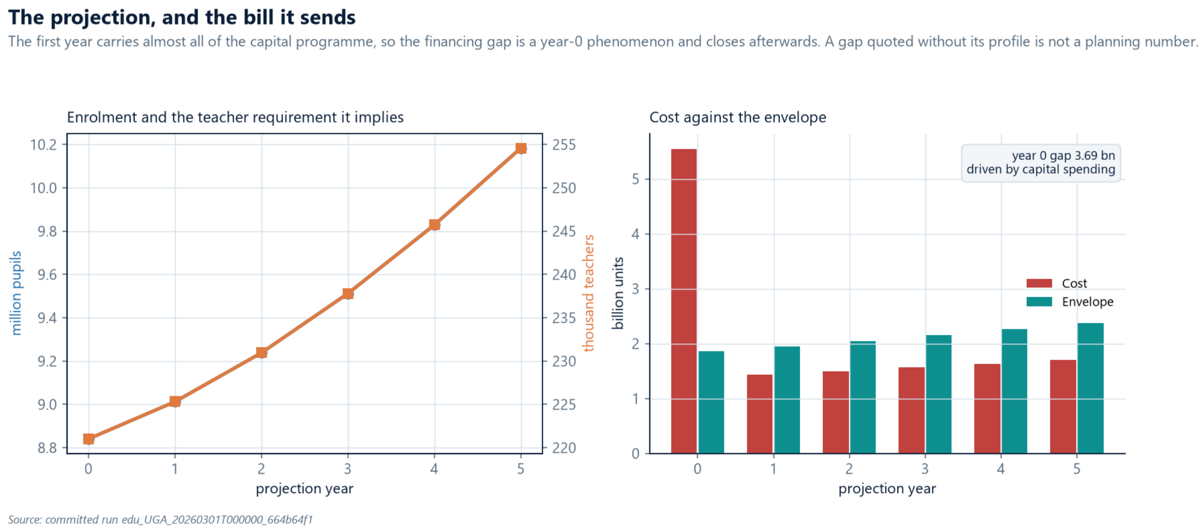


Figure 30. The committed run: enrolment, teachers and the financing gap. — Committed model run on Uganda

11.5.1 The financing profile, and why the gap is a year-0 phenomenon

Year	Cost	Envelope	Gap	Cumulative gap
0	5,546,382,850	1,859,574,879	3,686,807,971	3,686,807,971
1	1,437,695,512	1,952,553,623	-514,858,110	3,171,949,861
2	1,498,574,199	2,050,181,304	-551,607,105	2,620,342,756
3	1,563,766,899	2,152,690,369	-588,923,470	2,031,419,286
4	1,632,690,015	2,260,324,887	-627,634,872	1,403,784,414
5	1,704,807,525	2,373,341,132	-668,533,607	735,250,807

Almost the whole capital programme lands in the first year, so the financing gap is large at year zero and negative thereafter. A gap quoted as a single number — and this run's first-year gap is a large one — is not a planning figure. The profile is the planning figure, and the cumulative column is the one a finance ministry will read.

11.5.2 The classroom requirement, independently

Model 65 can also be computed directly from the school register: 90,803 classrooms exist and 157,646 are required at a norm of 45, a national deficit of 67,233. That figure and the run's capital programme are not the same quantity and should not be compared without reconciling the enrolment base first — which chapter 26 does.

11.6 The parameters that carry the result

Parameter	Where it comes from	What moves if it changes
Norm class size	Policy	The classroom requirement and the whole capital line
Pupil-teacher target	Policy	The teacher requirement, which is a restatement of the target rather than evidence about it
Unit cost per pupil	External costing study	Recurrent cost, and therefore the gap
Teacher salary	The public payroll	The largest single component of recurrent cost
Attrition rate	Administrative records	The recruitment path needed to hold the ratio
Discount rate	Finance ministry convention	The net present value, and the ranking of front-loaded against spread capital programmes

WHERE A PROJECTION'S UNCERTAINTY ACTUALLY LIVES
Not in a confidence interval. This is not an estimate from a sample and it has no sampling error. Its uncertainty lives in the parameter table above, and the honest way to report it is a sensitivity range over those parameters — not an interval, which would imply a sampling distribution that does not exist.

11.7 What this course did not do

Model	Why not
68. Benefit incidence of public spending	Needs unit subsidy by level joined to household use of public schools by quintile. The panel carries the second and the budget carries the first; joining them across the government/private split is the natural extension.
67. Expenditure projection under alternative scenarios	The scenario engine supports it; this course reports the baseline only, so that the parameter discussion above is not confounded with a scenario discussion.

11.8 Every model in this family the course does not estimate

With the input each one needs that is not present. This is the family's data request, and it is more useful than a proxy.

#	Model	Dependent variable	What it needs that is missing
62	Education cohort-flow projection model	Enrolment by grade and year	base enrolment; entry; promotion; repetition; dropout; population
63	Demographic education-demand model	School-age population and potential enrolment	UN/national population projections; age-specific participation; entry-age rules
64	Teacher requirement projection	Required teachers	projected enrolment; target PTR; teacher attrition; training pipeline
66	Education unit-cost model	Unit cost and cost per graduate	budget/expenditure; students; graduates; teachers; capital spending

#	Model	Dependent variable	What it needs that is missing
67	Education expenditure projection model	Total education spending by level/year	student projections; unit costs; teacher wages; capital plans; inflation
68	Benefit incidence of public education spending	Per-household/quantile imputed education subsidy	household enrolment; public/private status; unit subsidy by level; weights

11.8 What this family feeds

The MTEF envelope, the budget-allocation programme and the financing gap.

12. H. Policy Microsimulation, Human Capital and CGE Integration

Specification chapter	12
Models	69–72 (4 cards)
Estimated in this course	2 on microdata
Feeds	CDCGE, the poverty microsimulation and the cross-sector handover.

12.1 What this family estimates

What a policy does to households, how the human-capital stock evolves, and what the macro model receives.

12.2 The trap

READ THIS BEFORE THE MODELS
A microsimulation counterfactual is associational under the fitted model. It is not a treatment effect, and a quintile is not a policy lever.

12.3 The models

#	Model	Dependent variable	Method	Status
69	Cash-transfer education microsimulation	Counterfactual education and welfare outcomes	Static microsimulation with probabilistic or expected-value response	**Estimated**
70	Fee-removal and bursary simulation	Counterfactual enrolment, education spending and consumption	Microsimulation with behavioral response	Specified
71	Education-human-capital dynamic stock model	Labour force by education/skill category	Dynamic stock-flow simulation	Specified
72	Integrated education-CGE-poverty-budget model	GDP, sector output, factor wages, employment, fiscal balance, household welfare and poverty	Recursive dynamic CGE linked iteratively or sequentially to household microsimulation	**Estimated**

Table 18. Family H — all 4 cards

12.4 The specifications that carry the family

Model 69. Cash-transfer education microsimulation

Simulate effect of household transfers on enrolment, attendance and poverty.

Model 69
ΔP_i=P(Enroll|Income+Transfer)-P(Enroll|Income)
Counterfactual education and welfare outcomes
(82)

Key inputs	estimated enrolment model; transfer design; household welfare; eligibility rules
Method	Static microsimulation with probabilistic or expected-value response.
Diagnostics	Baseline replication, take-up sensitivity and confidence intervals.
Outputs	Additional students, fiscal cost, poverty and targeting metrics.
Integration	Links social protection, education and poverty modules.

Model 71. Education-human-capital dynamic stock model

Translate education flows into labour-force skill stocks over time.

Model 71
L_{e,t+1}=L_{e,t}+Graduates_{e,t}-Retirements_{e,t}-NetMigration_{e,t}
Labour force by education/skill category
(83)

Key inputs	graduates; age-specific labour participation; retirement; migration; skill mapping
Method	Dynamic stock-flow simulation.
Diagnostics	Mass-balance checks and labour-force reconciliation.
Outputs	Future skill supply by category and cohort.
Integration	Principal bridge from education projections to dynamic CGE factors.

Model 72. Integrated education-CGE-poverty-budget model

Evaluate economy-wide and distributional effects of education spending and reallocation.

Model 72
EduSpend_t → Capacity_t → Graduates_{e,t} → L_{e,t} & A_{s,t} → CGE equilibrium → HouseholdIncome_h → Poverty_h ; Share_{h,e} = LabourIncome_{h,e} / Σ_e LabourIncome_{h,e} ; Y_h^1 = Σ_e (1+Δw_e)Y_{h,e}^0 + Transfers_h^1 + OtherIncome_h^1 ; RealConsumption_h^1 = NominalResources_h^1 / HouseholdPriceIndex_h^1 ; EVI_i = w1 Poverty_i + w2 AccessDeficit_i + w3 OOSCRisk_i + w4 DropoutRisk_i + w5 CapacityGap_i + w6 LearningDeficit_i ; EDIS_EDUCATION/
data_raw/
data_harmonized/
metadata/
src/
harmonization/
participation/
household_demand/
learning/
labour/
spatial/
optimization/
projections/
microsimulation/
cge_link/
diagnostics/
config/
scenarios/
results/
tests/
docs/
GDP, sector output, factor wages, employment, fiscal balance, household welfare and poverty
(84)

Key inputs	education budget scenario; cohort-flow outputs; skill supply; returns/productivity parameters; SAM/CGE data; household factor-income shares
Method	Recursive dynamic CGE linked iteratively or sequentially to household microsimulation; scenario-based fiscal closure.
Diagnostics	SAM balance, CGE replication, closure report, labour-factor reconciliation, microsimulation income matching, scenario decomposition and sensitivity.
Outputs	Economy-wide growth, sector effects, skill wages, household welfare, poverty, inequality, tax revenue and fiscal returns to education.
Integration	Top-level EDIS policy engine for comparing education investments with alternative public spending uses.

Model 70. Fee-removal and bursary simulation

Simulate changes in participation and household welfare when school fees/costs are reduced.

Model 70
P_i^1=F(Fee_i-Subsidy_i,X_i); Welfare_h^1=C_h+Savings_h-ResidualCosts_h
Counterfactual enrolment, education spending and consumption
(85)

Method. Microsimulation with behavioral response

12.5 What this course estimated

Model 69 simulates a policy shock on real households through a fitted discrete-choice model and reports the distributional incidence. The shock used here is an upward move of one wealth quintile, applied to every household.

Quantity	**Value**
Baseline attendance, ages 6–17	85.7 %
Counterfactual attendance	89.2 %
Change	+3.5 pp
Persons in the simulation	31,019

Quantity	Value
Weighted population represented	30,750
Additional children reached	1,067
Incidence conserved across groups	yes

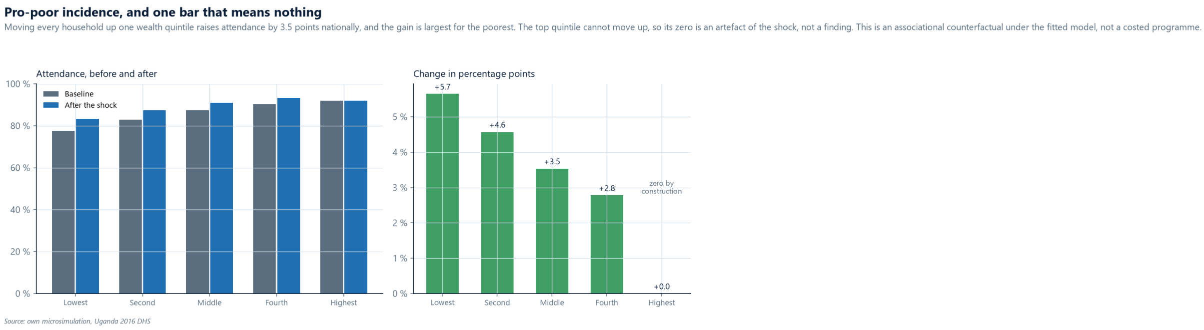


Figure 31. A wealth shock through the fitted model: who gains. — Uganda 2016 DHS, own microsimulation

Wealth quintile	Baseline	Counterfactual	Change	n
Lowest	77.6 %	83.3 %	+5.7 pp	7,629
Second	82.9 %	87.4 %	+4.6 pp	6,768
Middle	87.4 %	90.9 %	+3.5 pp	6,418
Fourth	90.4 %	93.2 %	+2.8 pp	5,822
Highest	91.9 %	91.9 %	+0.0 pp	4,382

TWO THINGS THIS SIMULATION IS NOT

It is not a treatment effect. It is an associational counterfactual under the fitted model: what the model predicts if a covariate takes a different value, holding its coefficients fixed. Nothing here identifies a causal parameter.

It is not a costed programme. A wealth quintile is not a policy lever. Linking a specific intervention — a cash transfer, fee abolition, a bursary — to the covariate change it would produce, and to its cost, is a separate exercise that families B and G supply the inputs for.

The top quintile's change is exactly zero because it cannot move up a quintile. That zero is an artefact of how the shock was defined and it is not a finding about rich households. It is left in the table deliberately: a simulation output that contains a mechanical zero and does not say so is how an artefact becomes a conclusion.

12.5.1 The conservation check

The sum of additional children reached across quintiles equals the national total, and the group weights sum to the survey population. That check either passes or the simulation has leaked people somewhere, and it is worth running every time.

12.6 Model 72 — the integrated handover

Model 72 is the integration point: education outputs, poverty outputs and the budget, handed to a macro model. The estate implements it as a versioned, validated exchange contract rather than as a second general-equilibrium model, which is the right decision and is stated as a design rule in its own documentation.

Figure 32. The education-to-macro exchange: what the contract asks for and what the run supplies. — CGE exchange contract and the committed Uganda run

Contract field	Supplied by the committed run?
year	yes
population	yes
enrolled	yes
teachers	yes
classrooms	yes
learning_score	**no**
completers	**no**
education_spending	yes
skilled_workers	**no**
human_capital_index	yes

human_capital_index is the run's gross enrolment ratio, an enrolment-based proxy the exchange documents as a proxy. It is not a human-capital measurement.

EG-07 AND EG-08 — WHAT A RECEIVING MODELLER MUST BE TOLD
Three of the contract's fields are empty in the committed run: learning score, completers and skilled workers. The human-capital index is the run's gross enrolment ratio, which the exchange documents as a proxy. And the operational coupling cannot be executed on this host, because the macro model's native stack is blocked by the machine's Application Control policy.

None of that makes the contract useless. It makes the handover note mandatory, and chapter 22 specifies what has to be in it.

12.7 The poverty link

The poverty side of model 72 is estimated in this course rather than assumed. Education payments are deducted from the welfare aggregate and the household is re-tested against its own poverty line.

Quantity	Value
Households with a student	2,243
Poverty rate as measured	3.8 %
Poverty rate net of education payments	9.5 %
Impoverishing effect	+5.7 pp
Concentration index of education spending	0.4134

Education spending is deducted from the monthly adult-equivalent welfare aggregate and the household is re-tested against its own poverty line. The difference is the impoverishing effect of education payments. It is an accounting exercise, not a behavioural one: it assumes the household would not have reallocated.

12.8 What this course did not estimate

Model	Why not
70. Fee removal and bursary simulation	The machinery is the same as model 69 and the expenditure elasticities come from family B. What is missing is a fee schedule by school type, which the survey does not observe — it observes what was paid, which is a choice.
71. Human-capital dynamic stock	Needs a learning-adjusted schooling measure, and family D cannot supply one.

12.8 Every model in this family the course does not estimate

With the input each one needs that is not present. This is the family's data request, and it is more useful than a proxy.

#	Model	Dependent variable	What it needs that is missing
70	Fee-removal and bursary simulation	Counterfactual enrolment, education spending and consumption	fee elasticity/model; baseline fees; subsidy design; household expenditure
71	Education-human-capital dynamic stock model	Labour force by education/ skill category	graduates; age-specific labour participation; retirement; migration; skill mapping

12.9 What this family feeds

CDCGE, the poverty microsimulation and the cross-sector handover.

Part III — The system engine
How a population becomes a requirement: cohort projection, grade flow, supply and capacity, teachers, cost, and the scenario machinery that moves them all at once.

13. Population and the demand for schooling

Everything downstream is a function of the age structure, not of the population total. A country whose population grows two per cent a year with a falling birth rate and one whose population grows two per cent a year with a rising one need different classroom programmes, and the totals are identical.

13.1 The projection

E-P-01
N[r,a+1,t+1] = N[r,a,t](1-m[r,a,t]) + M[r,a,t]
cohort-component projection
(86)

Survivors age into the next single-year cohort, deaths and net migration adjust the stock, and births enter at the bottom. The estate runs this by region and single year of age, which is what makes a school-age population extractable at all: a five-year age group cannot tell you how many six-year-olds arrive next September.

Module	Layer	Lines
‘Population.jl‘	demography	197

13.2 From population to demand

E-A-01
P_access = logistic(alpha0 + alpha1 income - alpha2 travel_time + policy - poverty and disability penalties)
Demand = school_age_population * P_access
Enrolment = min(Demand, Capacity)
access probability, demand and enrolment
(87)

Three things happen in those three lines and they are worth separating. An access probability is a behavioural model in income, travel time and policy. Demand is that probability applied to the school-age population. Enrolment is the **minimum** of demand and capacity — and that minimum is where a great deal of policy confusion begins.

WHY A FLAT ENROLMENT SERIES IS AMBIGUOUS
Enrolment equal to demand means the system is serving everyone who wants to come, and the constraint is behavioural — fees, distance, opportunity cost. Enrolment equal to capacity means the system is turning children away, and the constraint is physical. The two need opposite budgets, and the enrolment series alone cannot tell them apart. You need the demand series and the capacity series separately, which is why the run reports both.

13.3 What the survey layer contributes here

The access probability's parameters are exactly what family A estimates. The wealth gradient, the urban coefficient and the region effects in chapter 5's attendance model are the empirical content of the logistic access function — which is why a projection built on assumed access parameters, in a country where the survey exists, is a projection that has declined free information.

13.4 Grade flow

E-G-01
Promotion + Repetition + Dropout = 1
Completers = final_grade_enrolment * promotion_rate
Transitions are bounded by next-level capacity.
promotion, repetition, dropout
(88)

Within a grade the three transition rates sum to one. Completers are the final grade's promoted share, and transitions between levels are bounded by the next level's capacity — so a transition rate is partly a statement about supply, not purely about pupils.

Module	Layer	Lines
‘DropoutModel.jl‘	dropout	149

Chapter 5's over-age finding bears directly on this: 62.0 % of primary attenders are at least two years over-age, which is what a high repetition rate looks like after several years of accumulation. A grade-flow model calibrated to on-time progression will project fewer pupil-years than the system will actually carry.

14. Supply, capacity and the binding constraint

14.1 Three capacities and a minimum

E-S-01
SeatCapacity = Classrooms * NormClassSize * Shifts
TeacherCapacity = Teachers * PTRTarget
EffectiveCapacity = min(SeatCapacity, TeacherCapacity, DeskCapacity) * InfrastructureIndex
seat, teacher and desk capacity
(89)

Seat capacity is classrooms times a norm class size times shifts. Teacher capacity is teachers times a pupil-teacher target. Desk capacity is what furniture allows. Effective capacity is the smallest of the three, scaled by an infrastructure index — and which of the three is smallest decides what the money should buy.

If the binding constraint is...	...then the intervention is	...and building classrooms would
Seat capacity	Classroom construction	Work
Teacher capacity	Recruitment, deployment, training	Produce empty rooms
Desk capacity	Furniture procurement	Be an expensive way to buy desks
The infrastructure index	Water, electricity, sanitation, repair	Add unusable space to an unusable school

14.2 What the register says about the binding constraint

The school register lets this be checked rather than assumed. Aggregate pupil-teacher ratio is 49.8; aggregate pupil-classroom ratio is 78.1. Against a pupil-teacher target near 40 and a norm class size near 45, the classroom constraint is the tighter of the two nationally — but chapter 10's map shows it is not the tighter one everywhere, and a national comparison of two national ratios is exactly the reasoning that puts classrooms where teachers were needed.

14.3 Teachers as a stock, not a target

E-T-01
Teachers[t+1] = Teachers[t] + Graduates + Recruits - Attrition + Migration
teacher stock-flow
(90)

The teacher block is a stock-flow model: graduates and recruits in, attrition and migration out. It is the one part of the engine where the arithmetic is genuinely dynamic, and it is where most requirement projections go wrong.

A TEACHER REQUIREMENT IS A RESTATEMENT OF THE TARGET
Divide projected enrolment by a pupil-teacher target and you get a teacher requirement. That number contains no evidence about how many teachers are needed; it contains the target you chose. The evidence question is what a pupil-teacher ratio buys in learning, and family D — which cannot be estimated here — is where that question lives.

Module	Layer	Lines
‘TeacherDemand.jl‘	teachers	137

14.4 The learning block, and its status

E-L-01
Score[t+1] = Score[t] + beta1 instructional_time + beta2 attendance + beta3 teacher_quality + beta4 materials + beta5 digital_access - beta6 class_size
value-added learning
(91)

The estate implements a value-added learning form in instructional time, attendance, teacher quality, materials, digital access and class size. It is implemented, parameterised and not calibrated, because — as chapter 8 sets out — no measured assessment exists on any platform this course reaches. Any learning number this engine produces is a consequence of its assumed coefficients.

Module	Layer	Lines
‘LearningModel.jl‘	learning	172

15. Costing and public finance

15.1 The identity

E-F-01
Spend = UnitCost*Enrolment + Wage*Teachers + CapitalSpending
Spend <= Revenue + DonorFunds + HouseholdSpending + Borrowing
cost and the financing constraint
(92)

Unit cost times enrolment, plus the wage bill, plus capital, against revenue, donors, households and borrowing. Nothing in that identity is estimated from a household survey, and everything in it is a parameter someone chose.

15.2 Where each parameter comes from

Parameter	Source	Volatility
Unit cost per pupil	A costing study	High — it is the least stable number in the identity
Teacher salary	The public payroll	Low year to year, but a pay award moves the whole recurrent line
Capital cost per classroom	Procurement records	High, and locally variable
Envelope	The medium-term expenditure framework	Set by the finance ministry; a ceiling, not a forecast
Household contribution	Estimable — this course estimates it in chapter 6	Measured, with an interval
Discount rate	Finance ministry convention	Fixed by rule

Only one row of that table has an interval, and it is the one family B estimates. Household education spending is 375k UGX per student per year, and it is real money already in the system. A financing plan that treats it as zero is understating resources; one that treats it as permanent is assuming a burden that pushes +5.7 pp of households with students below the poverty line.

15.3 The committed run's financing profile

Year	Cost	Envelope	Gap
0	5,546,382,850	1,859,574,879	3,686,807,971
1	1,437,695,512	1,952,553,623	-514,858,110
2	1,498,574,199	2,050,181,304	-551,607,105
3	1,563,766,899	2,152,690,369	-588,923,470
4	1,632,690,015	2,260,324,887	-627,634,872
5	1,704,807,525	2,373,341,132	-668,533,607

The first year carries the capital programme and the gap is concentrated there. Any single-number 'financing gap' drawn from this run is a statement about the phasing of construction, not about the sustainability of the system.

Module	Layer	Lines
‘BudgetLoader.jl‘	finance	147
‘Costing.jl‘	finance	122
‘MTEF.jl‘	finance	93
‘CostBenefit.jl‘	costbenefit	144

15.4 The national budget as data

The estate reads the national approved budget directly rather than assuming an envelope: education vote line items, split into recurrent and capital. That is the right anchor for a costing exercise, because it is the number the finance ministry will check the model against.

16. Scenarios, diagnostics and the run record

16.1 What a scenario is in this estate

A scenario is a named, fingerprinted set of parameter changes applied to a baseline configuration. It is not a separate model. That distinction is what makes two runs comparable: if the code version and the input bundle match, the difference between two runs is the parameter set and nothing else.

Run record field	This run
Run identifier	edu_UGA_20260301T000000_664b64f1
Scenario	baseline
Scenario fingerprint	a4030be77b350176
Stages executed	harmonize, demography, teacher_demand, scenario

Every run also writes a configuration snapshot and a run log beside its outputs. A result whose configuration snapshot is missing cannot be reproduced, and this course treats an irreproducible result as no result.

16.2 Diagnostics that must be read

Check	What it catches
Grade-flow identity	Promotion plus repetition plus dropout not summing to one
Capacity identity	Effective capacity exceeding the minimum of its components
Enrolment bound	Enrolment above demand or above capacity
Finance identity	Cost components not summing to the total
Solver status	An optimisation reported as solved when it hit a time limit, with the optimality gap unreported
Population continuity	Cohorts appearing or vanishing between years

Module	Layer	Lines
‘DataQuality.jl‘	diagnostics	108
‘Acceptance.jl‘	acceptance	60

16.3 The baseline test suite

16.4 From a scenario to a defensible result

A number that cannot be traced to a run cannot be defended a year later, when the person who produced it has moved on and the question is whether the ministry should still be acting on it. The estate's service layer exists to make that traceability automatic rather than a matter of discipline.

Stage	What is created	What it records
A scenario is defined	A named parameter set, stored and versioned	Its own fingerprint, so two runs can be compared on whether the parameters were identical
A job is submitted	A queued unit of work with an identifier	Who submitted it and when; identical submissions are de-duplicated rather than run twice
The job runs	A run directory named for country, timestamp and code version	A configuration snapshot, a run log and a metadata file
Stages execute in order	Harmonise, project, derive requirements, cost, scenario	Which stages ran — a run that stopped early says so rather than shipping a partial result silently
Outputs are written	Tabular outputs and a summary	The run identifier, so any table can be traced back
A failure occurs	A recorded failure, not an exception that vanishes	The runner captures it without throwing, which is what the test suite checks

THE SCENARIO FINGERPRINT IS THE PART THAT MATTERS
Two runs of the same code on the same inputs with the same fingerprint must agree. If they do not, something outside the declared configuration changed and that is a defect. If the fingerprints differ, the runs are not comparable and any difference between them is uninterpretable — which is exactly the situation a 'before and after' slide usually hides.

16.5 The data-validation report

Before any model runs, the inputs are validated and the result is a report rather than a boolean. That distinction matters: a validator that returns pass or fail teaches nobody anything, and a validator that fails loudly on a warning stops work that could have proceeded with a caveat.

Check	What it catches	Severity
Required columns present	A bundle missing a table the stage needs	Blocking
Types and ranges	A rate outside zero to one, a negative population	Blocking
Identity checks	Promotion plus repetition plus dropout not summing to one	Blocking
Referential integrity	A region in one table and not in another	Blocking
Coverage	A table present but empty, or covering fewer regions than the population table	Warning — the run proceeds and the report says so
Plausibility	A pupil-teacher ratio in the hundreds, a school with more teachers than pupils	Warning, because it may be real

The two warning rows are the interesting ones. A table that is present but empty is how the accessibility figures in this course came to be straight-line: the road-network and travel-time tables exist in every country bundle and carry no rows for most of them. The validator says so; nothing stops.

16.6 What a run directory has to contain

Every committed run in this estate carries the same three files alongside its outputs, and a result whose directory is missing any of them is not reproducible.

File	What it is for
‘config_snapshot.toml‘	The configuration as it was at run time, not as it is now. The configuration file on disk today may have moved on.
‘run_metadata.json‘	Run identifier, code version, timestamp and the configuration sources that were read.
‘run.log‘	What happened, in order, including anything the run decided to skip.
‘summary.json‘	The headline quantities, so a reader can check they are looking at the run they think they are before opening any table.

For the run this manual quotes throughout:

Field	Value
Run identifier	edu_UGA_20260301T000000_664b64f1
Country	Uganda
Stages executed	harmonize, demography, teacher_demand, scenario
Scenario	baseline
Scenario fingerprint	a4030be77b350176
Persons in the harmonised microdata	49,315

16.7 The question to ask of any result

1. **Which run produced it?** If there is no run identifier, stop here.
2. **Which scenario, and what is its fingerprint?** Without the fingerprint you cannot tell whether a comparison is a comparison.

3. **Which stages ran?** A projection whose costing stage did not execute has no cost column, and an empty column is easily mistaken for a zero.
4. **What did the validator warn about?** Warnings do not stop a run, which means somebody has to read them.
5. **Is the configuration snapshot present?** It is what lets someone else get the same answer next year.

17. The estate module by module

This chapter is a map rather than an argument. Every module in the estate, by layer, so that any number produced anywhere in this course can be traced to the file that produced it.

17.Data — 11 modules, 1,654 lines

Loaders and harmonisers: DHS, LSMS, schools, catalogues.

Module	Lines
‘SchoolHarmonizer.jl‘	240
‘Schemas.jl‘	200
‘HouseholdHarmonizer.jl‘	189
‘SchoolLoader.jl‘	182
‘HouseholdLoader.jl‘	175
‘HouseholdSurveyCatalog.jl‘	165
‘UgandaLSMS.jl‘	162
‘LSMSHarmonizer.jl‘	156
‘DataValidation.jl‘	117
‘LSMSLoader.jl‘	64
‘Loader.jl‘	4

17.Services — 4 modules, 1,160 lines

Job service, scenario store, runner.

Module	Lines
‘Services.jl‘	957
‘JobService.jl‘	82
‘ScenarioStore.jl‘	67
‘AuditLog.jl‘	54

17.Web — 2 modules, 1,062 lines

The Genie application.

Module	Lines
‘Service.jl‘	1,011
‘CanonicalRoutes.jl‘	51

17.econometrics — 7 modules, 955 lines

The discrete-choice and regression engine behind the population-education models.

Module	Lines
‘DiscreteChoice.jl‘	250
‘Equity.jl‘	217
‘ExpenditureModel.jl‘	148
‘LinearModels.jl‘	111
‘AttainmentModel.jl‘	105
‘EnrolmentModel.jl‘	66
‘AttendanceModel.jl‘	58

17.integration — 6 modules, 902 lines

CGE, MTEF, BOE, SDI, unit-cost and cross-sector links.

Module	Lines
'SDIBridge.jl'	197
'MTEFBridge.jl'	193
'CGEExchange.jl'	185
'CrossSector.jl'	146
'UnitCostBridge.jl'	108
'BOEBridge.jl'	73

17.Core — 9 modules, 838 lines

Configuration, types, validation, scenarios, run context.

Module	Lines
'Indicators.jl'	226
'Config.jl'	141
'Auth.jl'	112
'Validation.jl'	93
'RunContext.jl'	82
'Scenario.jl'	70
'RunLog.jl'	53
'Paths.jl'	36
'Types.jl'	25

17.Models — 13 modules, 695 lines

The education system engine: demography, cohort flow, enrolment, supply, teachers, finance, learning, labour.

Module	Lines
'CohortFlow.jl'	67
'Demography.jl'	63
'PopulationProjection.jl'	62
'Enrolment.jl'	60
'Learning.jl'	57
'Supply.jl'	53
'GradeFlow.jl'	52
'Finance.jl'	50
'Teachers.jl'	50
'LabourMarket.jl'	49
'Impacts.jl'	48
'EducationIndicators.jl'	44
'Requirements.jl'	40

17.GIS — 3 modules, 385 lines

Spatial data, road network and climate risk.

Module	Lines
'SpatialData.jl'	276
'Network.jl'	81
'ClimateRisk.jl'	28

17.finance — 3 modules, 362 lines

Costing, budget and MTEF.

Module	Lines
‘BudgetLoader.jl‘	147
‘Costing.jl‘	122
‘MTEF.jl‘	93

17.Optimization — 9 modules, 317 lines

Mathematical programmes: school siting, teacher deployment, transport, budget, portfolio, transition.

Module	Lines
‘FacilityLocation.jl‘	145
‘Diagnostics.jl‘	71
‘InvestmentPortfolio.jl‘	37
‘SchoolLocation.jl‘	14
‘TeacherDeployment.jl‘	12
‘TransitionPlanning.jl‘	11
‘BudgetAllocation.jl‘	10
‘StudentTransport.jl‘	9
‘MultiObjectivePolicy.jl‘	8

17.accessibility — 2 modules, 256 lines

Road-network routing and catchment access.

Module	Lines
‘RoadNetwork.jl‘	134
‘Accessibility.jl‘	122

17.enrolment — 1 modules, 217 lines

Enrolment forecasting.

Module	Lines
‘ProgressionModel.jl‘	217

17.demography — 1 modules, 197 lines

Population projection.

Module	Lines
‘Population.jl‘	197

17.learning — 1 modules, 172 lines

Learning models.

Module	Lines
‘LearningModel.jl‘	172

17.dropout — 1 modules, 149 lines

Dropout and progression.

Module	Lines
‘DropoutModel.jl‘	149

17.costbenefit — 1 modules, 144 lines

Cost-benefit and rate of return.

Module	Lines
‘CostBenefit.jl‘	144

17.teachers — 1 modules, 137 lines

Teacher demand and deployment.

Module	Lines
‘TeacherDemand.jl‘	137

17.CDCGE — 1 modules, 135 lines

The dynamic CGE bridge.

Module	Lines
‘Bridge.jl‘	135

17.scenarios — 1 modules, 122 lines

Scenario engine.

Module	Lines
‘ScenarioEngine.jl‘	122

17.prioritization — 1 modules, 121 lines

Multi-criteria decision analysis.

Module	Lines
‘MCDA.jl‘	121

17.diagnostics — 1 modules, 108 lines

Solver and model diagnostics.

Module	Lines
‘DataQuality.jl‘	108

17.forecasting — 1 modules, 105 lines

Forecast engine.

Module	Lines
‘EnrolmentForecast.jl‘	105

17.reporting — 1 modules, 105 lines

Report generation.

Module	Lines
‘Report.jl‘	105

17.microsim — 1 modules, 95 lines

Household policy microsimulation.

Module	Lines
‘Microsimulation.jl‘	95

17.pipeline — 1 modules, 93 lines

The end-to-end pipeline.

Module	Lines
‘Pipeline.jl‘	93

17.tertiary — 1 modules, 89 lines

Post-basic and tertiary.

Module	Lines
‘PostBasic.jl‘	89

17.perf — 1 modules, 73 lines

Performance harness.

Module	Lines
‘Benchmark.jl‘	73

17.Simulation — 1 modules, 71 lines

The run engine.

Module	Lines
‘Engine.jl‘	71

17.acceptance — 1 modules, 60 lines

Acceptance checks.

Module	Lines
‘Acceptance.jl‘	60

17.API — 1 modules, 4 lines

REST routes.

Module	Lines
‘app.jl‘	4

Part IV — Geography and optimisation
The spatial infrastructure the estate reads, the accessibility it derives, and the nine mathematical programmes that turn a requirement into a decision.

18. The spatial data infrastructure

Geography enters this estate through a separate repository whose job is to be the authoritative spatial store for the whole decision-intelligence system, not just for education. That separation matters: the education model does not own the boundaries, the population grid or the road network, and it does not get to redefine them.

18.1 What the infrastructure holds

Sector layer group	Files
Administrative	3
Agriculture	0
Education	5
Energy	4
Environment	0
Health	2
ICT	0
Minerals	0
Population	0
Transport	0

The education group holds school registers by level — primary, secondary, post-primary, other and non-formal — as vector layers with pupils, teachers, classrooms, ownership and condition attributes attached.

18.2 The registry idea

Everything in that repository is driven from a single layer registry: one entry per spatial layer declaring its geometry type, coordinate system, primary key, required attributes, source, licence, quality grade and model linkage. That one declaration drives the data dictionary, the ingestion validation, the quality rules, the export column selection and the catalogue the models read.

WHY THIS IS THE RIGHT DESIGN
Because a layer's licence and quality grade travel with it into every downstream use. A model that consumes a layer knows whether it may be republished and how good it is, without anybody remembering to say so. The alternative — hard-coded mappings — loses that information at the first join.

18.3 The education GIS database

Table	Rows	What it is
‘pop_settlements‘	81,746	Settlement points with population
‘acc_settlement_school‘	80,160	Settlement-to-school distances
‘edu_school_source_records‘	48,278	Every source record, including those rejected, with the rule that rejected them
‘edu_schools‘	46,555	Matched school points, deduplicated across sources
‘ind_observations‘	7,713	National indicator observations
‘edu_match_candidates‘	3,189	Candidate matches considered during deduplication
‘meta_download_attempts‘	1,219	Every download attempt, successful or not
‘ref_admin_areas‘	918	Administrative boundaries with geometry
‘meta_coverage‘	648	What was attempted per country and dataset
‘meta_hdx_catalogue‘	124	Humanitarian data exchange catalogue
‘pop_raster_catalog‘	84	Population raster catalogue
‘ind_definitions‘	66	Indicator definitions
‘ref_countries‘	55	Country reference

Two of those tables are the ones a sceptical reader should look at first. ‘edu_school_source_records‘ keeps every rejected record with the rule that rejected it, and ‘meta_download_attempts‘ keeps every attempt including the failures. A pipeline that hides its failures cannot be audited, and this one does not.

18.4 The coordinate problem, stated precisely

Stage	Records	Share of the register
In the register	16,884	100 %
With a pupil count	12,758	75.6 %
With a coordinate inside the country	10,807	64.0 %
Placed inside a county polygon	6,821	40.4 %

LON/LAT are zero or out of country for the remainder; the register's counts are reliable, its per-school coordinates are not.

The estate's own data audit reaches the same conclusion independently: the register's counts by level are reliable and its per-school coordinates are not, because only a subset reprojects validly from the historic national grid. This manual quantifies what that costs — a pupil-teacher ratio of 48.2 on the mappable schools against 49.8 on all of them — and requires every map built on the register to say so.

19. Accessibility and catchments

19.1 Four measures, four questions

Measure	Answers	Blind to
School density per square kilometre	How much supply is here?	Where people are, and whether the schools have room
Nearest-school distance	How far is the closest school?	Whether it is full, and whether anyone else is competing for it
Gravity accessibility index	How much school is reachable, discounted by distance?	Capacity, unless it is in the numerator; needs a decay parameter
Floating catchment ratio	How much capacity is reachable per competing child?	Nothing much — it is the most complete of the four and the most data-hungry

This course computes the second and demonstrates the fourth's logic. The estate implements all four.

19.2 The coverage curve

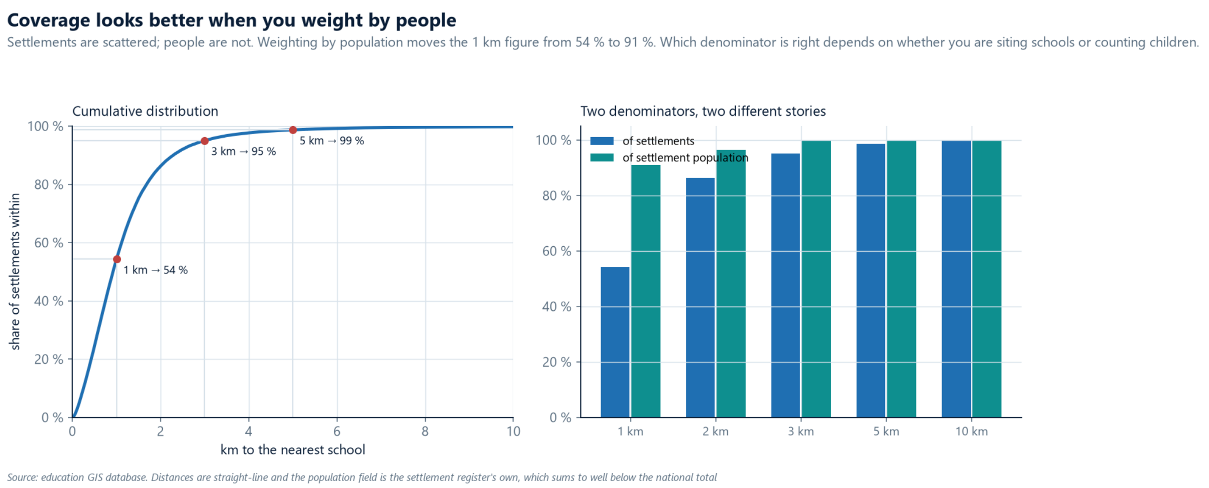


Figure 33. The coverage curve: how much of the country each distance threshold buys. — Education GIS database, straight-line distance

Median distance to the nearest school of any level is 0.92 km. That is a good number and it should be read with two qualifications. It is straight-line, so it is a lower bound on travel. And it is to a school of *any* level — the median distance to a secondary school is a different and much larger quantity, which is where the secondary net attendance rate of 17.8 % in chapter 5 starts to have a spatial explanation.

19.3 A catchment is a competition

The reason a floating catchment ratio beats a distance is that distance treats a school as inexhaustible. Two settlements five hundred metres from the same one-stream school are not both well served. The estate's catchment demand model (model 55) allocates settlement population to schools and then divides capacity by the population competing for it.

Module	Layer	Lines
‘RoadNetwork.jl‘	accessibility	134
‘Accessibility.jl‘	accessibility	122
‘SpatialData.jl‘	GIS	276
‘Network.jl‘	GIS	81
‘ClimateRisk.jl‘	GIS	28

19.4 Straight-line against routed

WHAT EG-13 COSTS, CONCRETELY
A settlement two kilometres from a school across a river with no bridge is, for a six-year-old, not two kilometres from a school. Straight-line distance cannot see that, and the error is not random: it is largest exactly where terrain is difficult, which is where access is worst. Every accessibility figure in this manual is therefore an optimistic bound, and the routed graph in the estate is the fix.

20. The optimisation suite

Nine mathematical programmes sit in the estate's optimisation layer. They differ in what they decide, but they share a structure worth learning once: an objective, a set of decisions, and constraints that make the decisions feasible.

20.1 The nine programmes

Programme	Decides	Type
School location	Which candidate sites to open	Mixed-integer — a site is opened or it is not
School consolidation and expansion	Which schools to close, merge or expand	Mixed-integer
Teacher deployment	How many teachers of each cadre go to each region	Mixed-integer
Subject-teacher allocation	Which subject specialists go where	Mixed-integer
Student transport	Which pupils travel to which school	Linear — flows are continuous
Transition planning	How much capacity to add at the next level	Linear
Budget allocation	How much of the budget each programme gets	Linear
Investment portfolio	Which investments enter the capital programme	Mixed-integer
Multi-objective policy	The weighted combination across learning, completion, equity and employment	Linear

Module	Layer	Lines
‘FacilityLocation.jl‘	Optimization	145
‘Diagnostics.jl‘	Optimization	71
‘InvestmentPortfolio.jl‘	Optimization	37
‘SchoolLocation.jl‘	Optimization	14
‘TeacherDeployment.jl‘	Optimization	12

Module	Layer	Lines
‘TransitionPlanning.jl‘	Optimization	11
‘BudgetAllocation.jl‘	Optimization	10
‘StudentTransport.jl‘	Optimization	9
‘MultiObjectivePolicy.jl‘	Optimization	8

20.2 The four that carry the capital decision

School location MILP

E-OPT-01
min fixed construction cost + weighted student travel time
subject to student assignment, school capacities, opening decisions and maximum travel time.
School location MILP
(93)

Teacher deployment MILP

E-OPT-02
min weighted shortages
subject to available teachers, regional-subject demand, salary budget and equity floors.
Teacher deployment MILP
(94)

Budget allocation LP

E-OPT-03
max learning + completion + equity benefits
subject to total budget, programme scale limits and regional equity floors.
Budget allocation LP
(95)

Student transport LP

E-OPT-04
min transport cost
subject to all students being assigned and school capacities.
Student transport LP
(96)

20.3 Why the integer ones are harder

A school is built or it is not. Half a school serves nobody, so the siting decision is a binary variable and the problem is a mixed-integer programme. Flows — pupils travelling, budget allocated — are continuous, so those problems are linear and solve quickly and exactly.

THE NUMBER THAT MUST TRAVEL WITH AN INTEGER RESULT
The optimality gap. An integer solver that stops at a time limit reports the best solution found and the bound it could not close. A plan described as optimal when the solver stopped at a four per cent gap is a plan described wrongly, and the estate's solver diagnostics module exists to make that gap visible rather than convenient to omit.

20.4 Infeasibility is information

A programme that returns infeasible has told you something precise: the constraints you wrote cannot all hold at once. A coverage floor of ninety-five per cent within three kilometres, a capital budget, and a maximum class size may simply be incompatible. The right response is to find which constraint to relax and by how much — not to loosen everything until something comes out. The estate handles infeasibility gracefully and reports it, which its test suite checks explicitly.

21. Siting, deployment and the capital decision

21.1 The objective is the decision

Every siting result in this course carries its objective function, because the objective is where the policy choice lives. The model does not choose it and cannot; it optimises whatever it is given, perfectly.

Objective	Serves	Sites schools
Minimise total cost	A finance ministry	Where they are cheapest to build and run, which is where people already are
Maximise coverage	A programme manager against an access target	Where they pick up the most currently-unserved children
Minimise population-weighted travel	Communities on average	Near large settlements, which may already have schools
Minimise the worst region's unmet share	An equity mandate	In the hardest, emptiest places, serving fewer children per shilling
Weighted multi-objective	A cabinet that has made its trade-off explicit	Wherever the weights say — which is why the weights are the political document, not the model

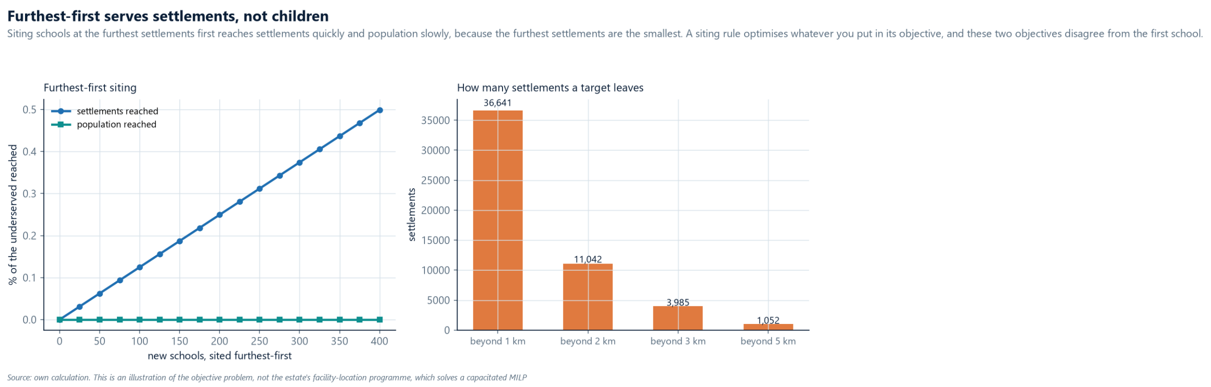


Figure 34. A siting sweep: how much a capital programme buys. — Own calculation on the Uganda accessibility surface

The illustration above is deliberately naive — a greedy furthest-first rule rather than the estate's capacitated programme — because it isolates the objective problem from the solver. Siting furthest-settlement-first reaches settlements quickly and children slowly, since the furthest settlements are the smallest. Two defensible objectives, opposite orderings, from the first school.

21.2 What a siting run needs before it can run

Input	Where it comes from in this course	Status
Demand origins with population	Settlement register	80,160 points
Existing supply with capacity	School register	Counts reliable, coordinates partial
Travel cost between them	Accessibility table	Straight-line only (EG-13)
Candidate sites	The estate's country bundles	Present
Construction and operating cost	Costing parameters	External, and the least stable input
The objective	The ministry	**Not a model output**

21.3 The capacity gap as a capital programme

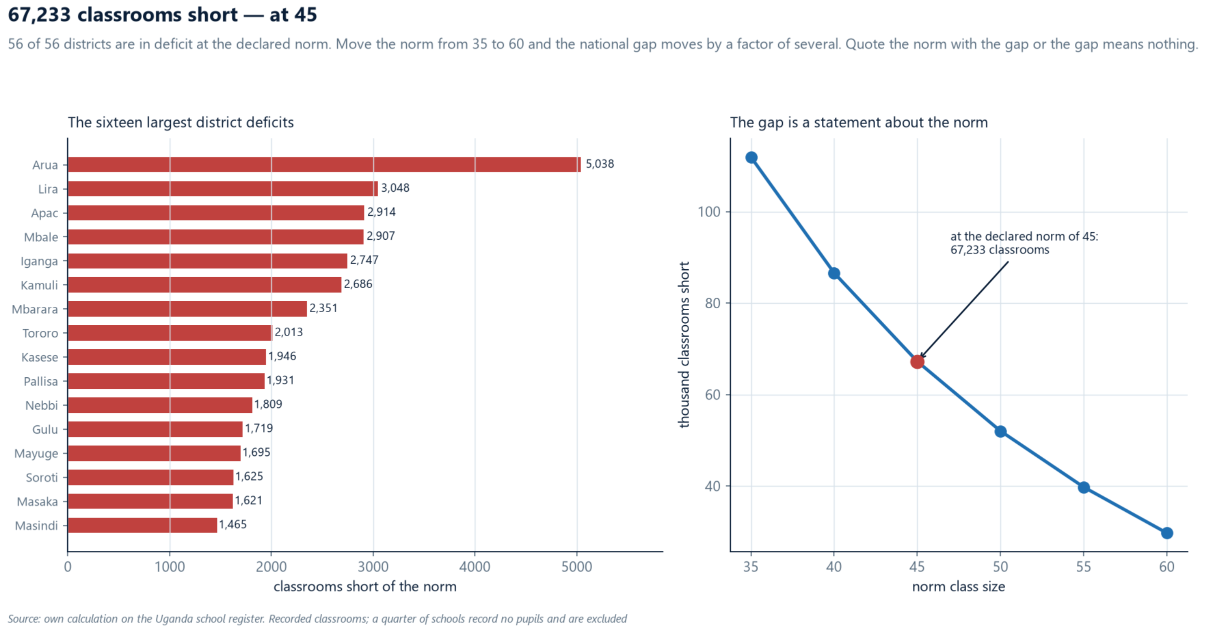


Figure 35. Classroom deficit by district at a norm of 45 pupils. — Uganda school register, own calculation

At the declared norm of 45 the national deficit is 67,233 classrooms across 56 districts. The sensitivity panel is the more useful half of that figure: the deficit is a smooth and steep function of the norm, so the first question to a ministry proposing a construction programme is not how many classrooms but at what class size.

21.4 Teacher deployment

The deployment programme minimises weighted shortages subject to available teachers, regional and subject demand, the salary budget and equity floors. Its inputs are the requirement from chapter 14 and the existing distribution from the register — and its most important property is that it can only move teachers who exist. A deployment optimum inside a national shortage redistributes the shortage; it does not remove it.

Chapter 10's map is the relevant evidence: the pupil-teacher ratio ranges from about 29 to about 82 across districts. A gap of that width inside one country is a deployment problem before it is a recruitment problem.

Part V — Integration
What leaves the education model: the macro exchange, the poverty link, the fiscal envelope, the cross-sector schemas — and the reconciliation problem that meets you whenever two of them describe the same thing.

22. The CDCGE exchange

An education model that stops at the sector boundary tells a cabinet how many classrooms it needs. An education model wired to a macro model tells it what the classrooms do to growth, revenue and the wage structure. The wiring is the hard part, and this estate does it with a contract rather than with a second model.

22.1 The design rule

NO SECOND CGE IS BUILT HERE
The exchange layer defines, validates and serialises an interface to an existing macro model. It contains no solver. That is a deliberate rule in the estate's own documentation, and it is the right one: two general-equilibrium models in one system will disagree, and nobody will be able to say which is authoritative.

22.2 The contract

Direction	Payload	Shape
Education → macro	‘edu_to_macro‘	Year-indexed rows over the legacy column set, each field with documented units, plus a human-capital index
Macro → education	‘macro_to_edu‘	Variable, period, baseline, scenario, percentage change

Every payload carries a semantic version, and compatibility is gated: same major version, minor no newer than the consuming build. A field rename or a unit change is a breaking major bump. That sounds bureaucratic until the first time a macro modeller silently receives spending in thousands where they expected millions.

22.3 What the committed run actually supplies

Figure 36. The education-to-macro exchange: what the contract asks for and what the run supplies. — CGE exchange contract and the committed Uganda run

Field	Supplied?	What it is here
year	yes	Projection year, 0-indexed
population	yes	Total population from the run summary
enrolled	yes	Total enrolment from the projection
teachers	yes	Teacher requirement at the target ratio
classrooms	yes	Derived from the teacher requirement
learning_score	**no**	— the run produces none —
completers	**no**	— not carried in the shipped tables —
education_spending	yes	Total cost from the costing block
skilled_workers	**no**	— needs a labour-supply mapping —
human_capital_index	yes	The gross enrolment ratio, used as a proxy

22.4 The two fields that matter most, and are empty

Skilled workers is the field a computable general equilibrium model most needs, because it is how education enters the production function. Filling it requires a mapping from education level to labour-force skill category — which is exactly what chapter 9's education-to-occupation table is.

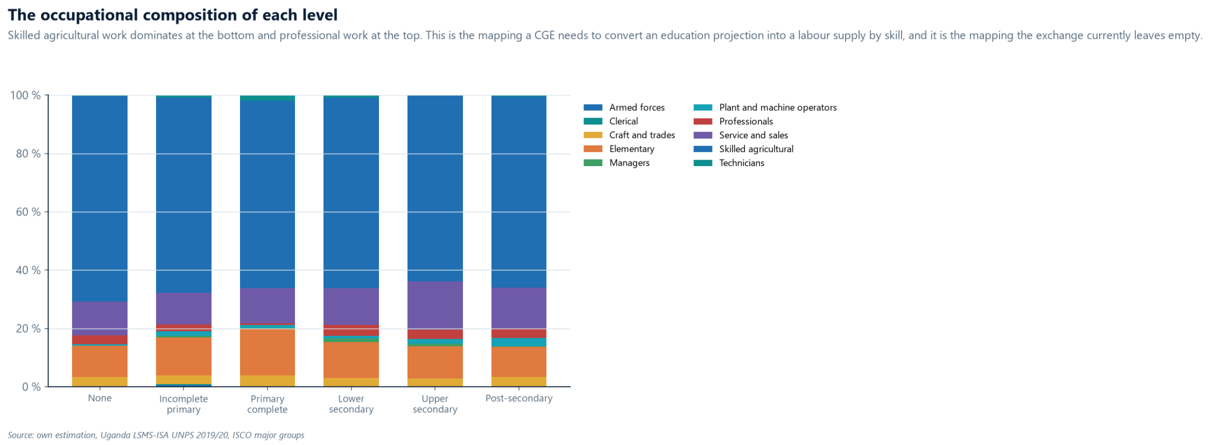


Figure 37. Where each education level ends up working. — Uganda LSMS-ISA UNPS 2019/20, own estimation

Learning score is empty for the reason chapter 8 gives. A human-capital index built on enrolment alone counts years in a classroom and not what happened in it, and the exchange documents that honestly rather than dressing the enrolment ratio in a better name.

22.5 What has to be in the handover note

- The run identifier, the scenario name and the scenario fingerprint, so the receiving modeller can ask for the same run again.
- Every empty contract field, named. A macro modeller who discovers an empty column at solve time will fill it with something, and you will not know what.
- The units of every populated field, and the base year of any monetary one.
- That the human-capital index is an enrolment proxy, in those words.
- That the coupling was not executed here, and why (EG-08).

22.6 The direction that does not exist yet

The contract defines a macro-to-education payload, and the bridge accepts it. What the estate does not do is iterate: there is no loop that feeds macro results back into the education projection and re-solves until the two agree. The handover is one-directional in practice. Calling it an integrated equilibrium would overstate it, and this manual does not.

23. Poverty, microsimulation and incidence

23.1 Two different questions

Question	Method	What it needs
Who pays, and does it hurt?	Accounting on observed payments	Expenditure, a consumption aggregate and a poverty line
Who would gain from a policy?	Microsimulation through a fitted model	A behavioural model and a covariate the policy moves

The first is measurement and the second is inference. Both are in this chapter and they must not be blurred.

23.2 The burden, measured

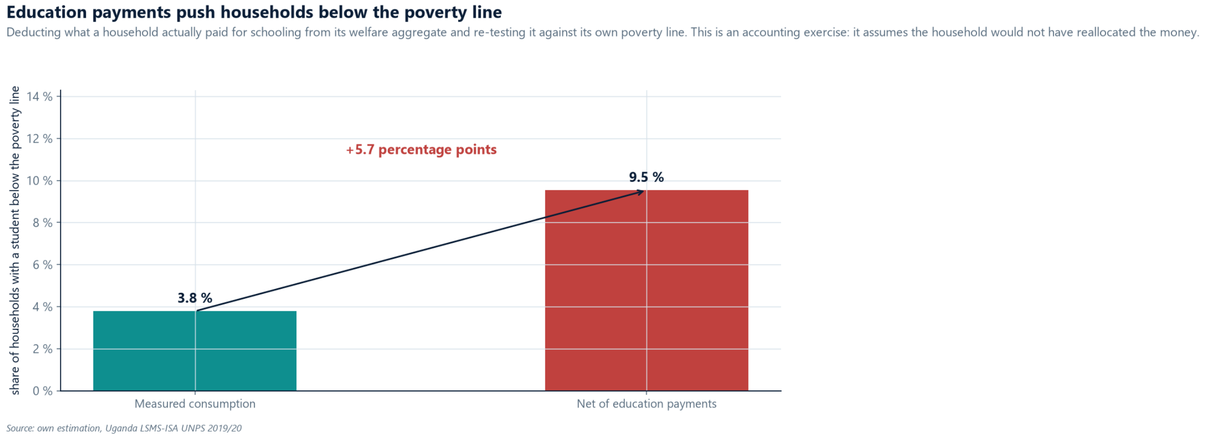


Figure 38. Poverty before and after education payments are deducted. — Uganda LSMS-ISA UNPS 2019/20, own estimation

Quantity	Value
Households with a student	2,243
Mean education budget share	9.3 %
Catastrophic at 10 % of consumption	32.2 %
Poverty rate, measured	3.8 %
Poverty rate net of education payments	9.5 %
Impoverishing effect	+5.7 pp
Concentration index of education spending	0.4134

Education spending is deducted from the monthly adult-equivalent welfare aggregate and the household is re-tested against its own poverty line. The difference is the impoverishing effect of education payments. It is an accounting exercise, not a behavioural one: it assumes the household would not have reallocated.

23.3 Why a catastrophic-spending indicator misses the excluded

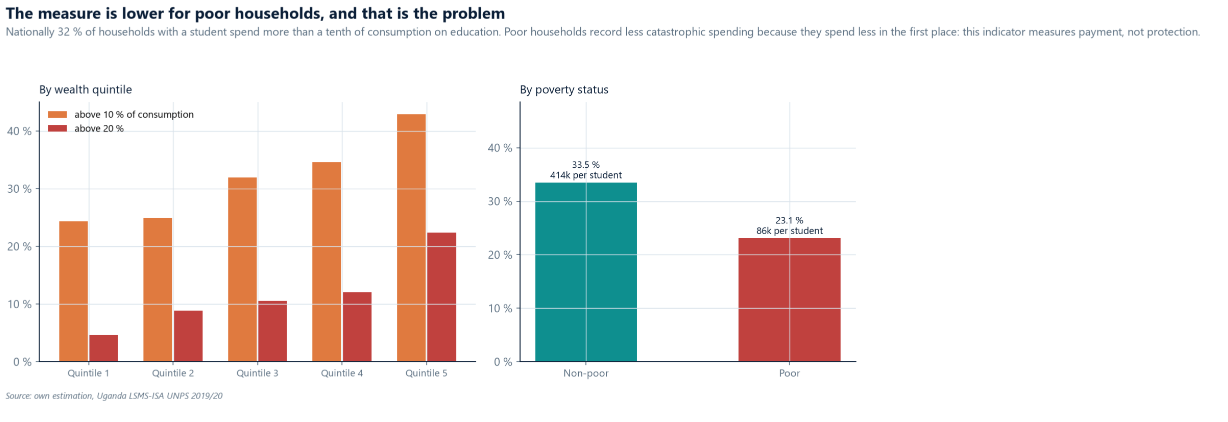


Figure 39. Catastrophic education expenditure, and why it is lowest among the poor. — Uganda LSMS-ISA UNPS 2019/20, own estimation

Poor households in this survey record *less* catastrophic education spending than non-poor ones, because they spend less. Any indicator defined on payments is blind to the household that pays nothing because the child is not there. The pairing that works is a payment indicator alongside a participation indicator: chapter 5's out-of-school rate by quintile is the other half of chapter 6's catastrophic rate by quintile, and neither is interpretable alone.

23.4 The microsimulation, and its four honest labels

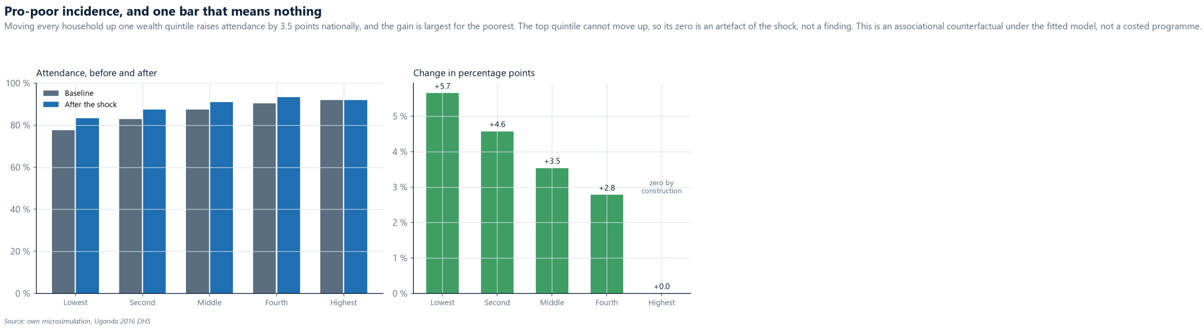


Figure 40. A wealth shock through the fitted model: who gains. — Uganda 2016 DHS, own microsimulation

Label	Statement
What it is	A counterfactual attendance rate under the fitted logit when one covariate is shifted, weighted to the population
What it is not	A treatment effect. No causal parameter is identified anywhere in it
What it assumes	That the coefficients hold at the shifted covariate value, and that nothing else moves
What is mechanical	The top quintile's zero change — it cannot move up a quintile

National attendance moves from 85.7 % to 89.2 %, +3.5 pp, and the gain is monotonically larger the poorer the quintile. That shape — not the magnitude — is what a targeting decision needs.

23.5 Turning it into something costed

The simulation above shifts a wealth quintile, which no ministry can do. Making it a policy takes three steps this course has the pieces for:

6. **Name the instrument.** A conditional transfer, fee abolition at a level, a bursary with an eligibility rule.
7. **Map the instrument to the covariate.** A transfer of a stated size moves household consumption by a stated amount; chapter 6's expenditure elasticity is what converts one into the other.
8. **Cost it.** Chapter 6 gives what households currently pay, chapter 15 gives the unit cost, and the product with the reached population from the simulation gives the bill.

What the course does not do is pretend the third step has been done when it has not. A distributional shape without a price is a targeting argument, not a budget submission.

24. The fiscal envelope and the budget

24.1 A ceiling is not a forecast

The medium-term expenditure framework gives the education sector a ceiling. It is a political commitment about what will be made available, not a prediction of what will be spent, and certainly not a statement about what is needed. The estate reads it as an input constraint and nothing more.

24.2 The approved budget as data

The estate also reads the national approved budget directly — education vote line items split into recurrent and capital. That is the anchor a costing exercise should be checked against, because it is the number the finance ministry recognises. A model whose recurrent cost is three times the approved recurrent budget is not describing the same system.

24.3 Household spending is part of the envelope

The financing identity in chapter 15 has a household term, and in this setting it is large. Households with a student spend 375k UGX per student per year on average. Two mistakes are available and both are common: treating that money as absent, which understates what the system already commands; and treating it as permanent, which assumes a burden that pushes +5.7 pp of households with students below the poverty line.

24.4 Where the money would have to come from

Year	Cost	Envelope	Gap	Cumulative gap
0	5,546,382,850	1,859,574,879	3,686,807,971	3,686,807,971
1	1,437,695,512	1,952,553,623	-514,858,110	3,171,949,861
2	1,498,574,199	2,050,181,304	-551,607,105	2,620,342,756
3	1,563,766,899	2,152,690,369	-588,923,470	2,031,419,286
4	1,632,690,015	2,260,324,887	-627,634,872	1,403,784,414
5	1,704,807,525	2,373,341,132	-668,533,607	735,250,807

Read the cumulative column, not the annual one. A programme that front-loads construction produces a large year-zero gap and a series of small surpluses afterwards; whether that is financeable is a question about borrowing and phasing, not about the education sector.

25. Cross-sector linkage

25.1 The schema, and the gate

The estate carries versioned linkage payloads between education and adjacent sectors — health, labour, social protection — with the same version discipline as the macro exchange. It also carries something less common and more valuable: a gate that refuses to label a linkage causal.

THE CAUSAL-LABELLING GATE
A linkage payload may declare its interpretation as *associational* or *ecological*. Any other value is rejected by the code, not by convention. A causal cross-sector claim would need an identification strategy that is not implemented, and the schema will not let an analyst assert one by accident.

25.2 The estate's own cross-country result, reported as null

The estate runs an ecological association of education attainment against youth not in employment, education or training across fifty countries. The slope is about −0.002, the correlation about −0.004, and the R^2 is essentially zero. It reports this as null rather than suppressing it.

Two reasons make that the expected result rather than a failure. The regressor mixes attainment and literacy proxies across countries, so the construct is not consistent. And national NEET is multi-causal in ways that a single education aggregate cannot capture. The estate's own conclusion is that individual-level microdata is required for a credible linkage.

25.3 Which is exactly what chapter 9 supplies

At the individual level, in one country, on the household panel, the relationship is visible: NEET among 15–24 year olds is 15.0 % overall and varies by completed education, sex and residence in patterns that a national aggregate erases entirely.

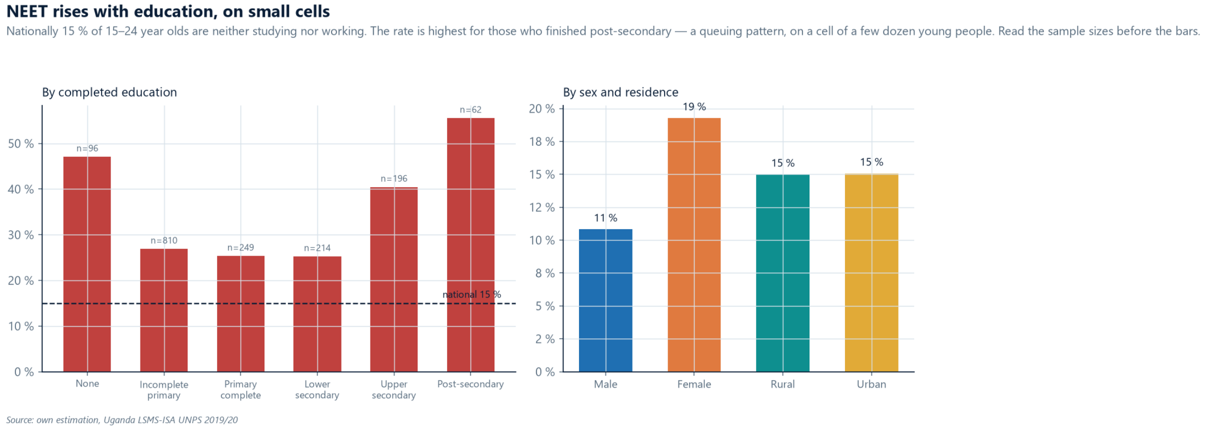


Figure 41. Youth not in education, employment or training. — Uganda LSMS-ISA UNPS 2019/20, own estimation

THE ECOLOGICAL FALLACY, IN ONE SENTENCE
A relationship between country averages is not a relationship between people, and it is not even guaranteed to have the same sign. This is why the linkage schema forces the word *ecological* onto a cross-country payload, and why the individual-level estimates in family E exist.

25.4 The education–health link the schema is built for

The canonical cross-sector linkage in this direction is a mother's education against her child's health outcomes. The schema and the gradient estimator are ready for it. What it needs is the survey's child and women's recodes rather than the household roster this course uses, which is a data request, not a modelling problem — and it is the same shape of request that chapter 8 makes for learning.

26. The reconciliation problem

Four sources describe primary education in the same country. They disagree. This chapter is about why, because a planner who cannot explain the disagreement cannot use any of the four.

Figure 42. Four sources, one country, four numbers that do not match. — Own reconciliation across DHS, LSMS, the school register and the committed run

26.1 The four numbers

Primary-age school participation

Source A	Uganda 2016 DHS, net attendance ratio
Value A	0.7881
Source B	Uganda LSMS 2019/20, currently attending, ages 6-12
Value B	0.9218
Reconciles?	**no**

Different definitions and different years. The DHS ratio counts children attending PRIMARY; the LSMS figure counts children attending ANY level, so pre-primary and secondary attenders of primary age are inside it. The three-year gap adds to the difference. Neither is wrong and they must not be quoted as a trend.

Primary pupil-teacher ratio

Source A	Uganda school register (EMIS), aggregate
Value A	49.8346
Source B	Committed model run, primary_ptr
Value B	42.6591
Reconciles?	**no**

The register divides recorded pupils by recorded teachers across all schools including the quarter with no pupil count. The run's ratio comes from its own enrolment projection and teacher requirement, which is a modelled quantity, not a measurement.

Primary enrolment

Source A	Uganda school register, sum of recorded pupils
Value A	7,094,061.0000
Source B	Committed model run, base enrolment
Value B	8,840,589.0000
Reconciles?	**no**

A quarter of the register's schools record zero pupils. The register undercounts by construction, and the run's base enrolment comes from the country bundle rather than from the register.

Household education spending

Source A	Per student, this analysis
Value A	374,678.4676
Source B	Per household with students, this analysis
Value B	1,030,549.4337
Reconciles?	yes

Both are correct and they differ by the mean number of students in a household. They reconcile exactly; the failure mode is quoting one under the other's label.

26.2 The four kinds of disagreement

Kind	**Example here**	**What to do**
Definitional	A net attendance ratio counts children attending *primary*; an 'attending school' question counts any level	Restate both under one definition, or report both with their definitions attached and refuse the comparison
Temporal	One survey is three years older than the other	Never present the difference as a trend unless the definitions match exactly
Coverage	A quarter of the register's schools record no pupils	Quote the coverage with the number, and prefer the source whose coverage suits the decision
Denominator	Per student against per household with students	This one reconciles exactly once the denominators are named

26.3 What not to do

- **Do not average them.** The mean of a definitional disagreement is a number that describes nothing.
- **Do not pick the one that suits the argument** and drop the others silently. If the choice is defensible, defending it in one sentence costs nothing.

- **Do not describe a definitional gap as data quality.** The sources are usually all correct about what they measure.
- **Do not present a cross-source difference as validation.** Agreement between two sources with the same definition is reproducibility. It is not evidence that either matches reality.

26.4 The reconciliation note

Every deliverable in this course that quotes a figure with a counterpart elsewhere carries a reconciliation note: what the other source says, why it differs, and which one this deliverable uses and why. It is three sentences and it forecloses the entire class of argument that begins 'but the ministry's own figure is...'.

Part VI — Practice and governance
The worked examples, the cases, the rules that survive contact with a minister, the laboratories, the errors, and what may be published.

27. Four worked examples

Each is reproduced from the shipped data in the laboratories. Work them by hand once before running the script.

27.1 Recovering a design effect from two standard errors

Step	Value
Estimate	0.788055
Design standard error	0.004739
Unweighted n	20,573
Simple-random-sample standard error	0.002849
Design effect = (design ÷ SRS)^2	2.7663
Reported design effect	2.7663
Effective sample size	7,437

The identity closes only with the **unweighted** n in the simple-random-sample formula. Using the weighted denominator is the single most common error in this calculation.

27.2 Three scales for one coefficient

Scale	Value	What it says
Coefficient	0.1506	Log-odds. Unusable
Odds ratio	1.162	The odds multiply by this
Implied risk ratio	1.031	Starting from 78.8 %
Average marginal effect	+1.52 pp	What a programme note carries

The three diverge and the divergence grows with the baseline. At a prevalence near eighty per cent the odds ratio is a poor guide to anything a reader will infer.

27.3 The denominator identity

Quantity	Value
Per household with students	1.03 m UGX
Mean students per household	2.6947
Implied per student	382k UGX
Per student, computed directly	375k UGX

The two per-student figures agree to within the difference between a mean of ratios and a ratio of means. That agreement is the check; the point of the exercise is that the first row and the last row differ by a factor of 2.75.

27.4 The capacity gap as a function of the norm

At the declared norm the national deficit is 67,233 classrooms. Recompute it at 35, 40, 50, 55 and 60 and plot the curve. The point of the exercise is not the number at any one norm; it is the steepness.

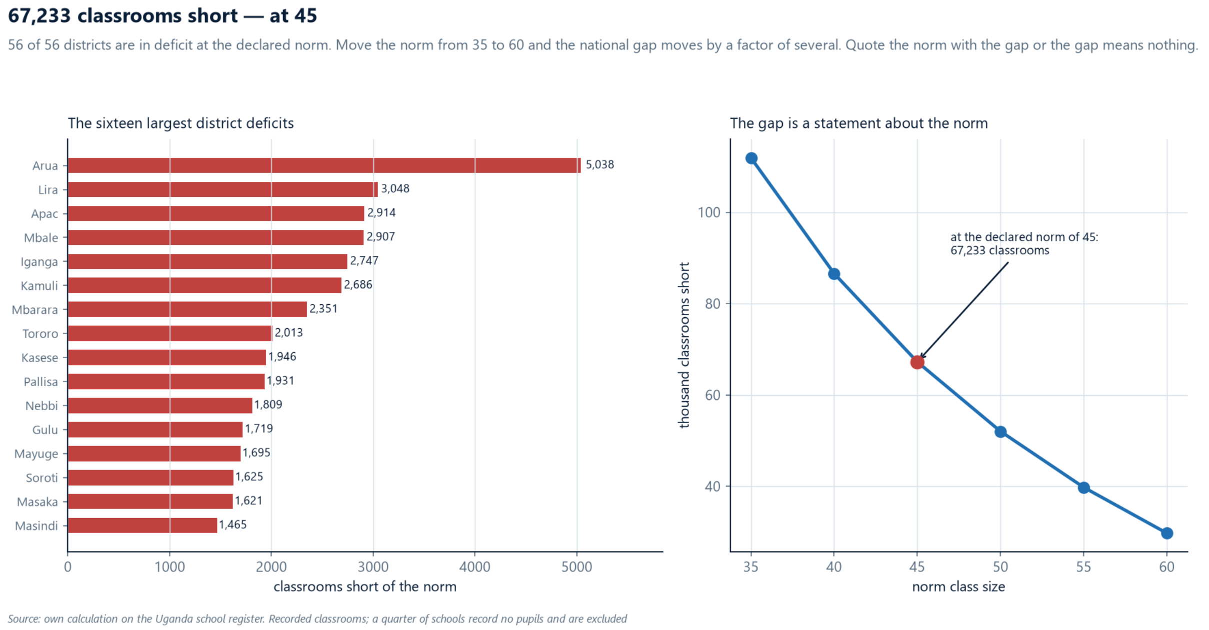


Figure 43. Classroom deficit by district at a norm of 45 pupils. — Uganda school register, own calculation

27.5 A concentration index by hand

The measure is a covariance between the outcome and the fractional rank in the socioeconomic distribution, scaled by the mean. Doing it once by hand is the only way to see why the sign depends on the outcome rather than on the inequality.

E-0-13

$$r_i = \text{cumulative weight below } i + \tfrac{1}{2}w_i, \qquad r_i \in (0, 1)$$

$$\mathrm{CI} = \frac{2}{\mu}\,\mathrm{Cov}_w\!\left(y_i,\ r_i\right)$$

the concentration index

(97)

Step	What you do
1	Sort by the ranking variable — here the wealth quintile — poorest first.
2	Form the weighted cumulative share and take its mid-point for each observation. That is the fractional rank.
3	Take the weighted covariance of the outcome with the fractional rank.
4	Multiply by two and divide by the weighted mean of the outcome.
5	Read the sign **against the outcome type**. Negative means concentrated among the poor, which is the good direction for a bad outcome and the bad direction for a good one.

Applied to this course's data: primary-age attendance gives 0.0463 and out-of-school status gives -0.3949. Opposite signs, one message.

27.6 Checking a decomposition closes

Quantity	Value
Mean, group A	73.9538 %

Quantity	Value
Mean, group B	70.9275 %
Gap	+3.0262 pp
Endowments	-0.5441 pp
Coefficients	+3.5066 pp
Interaction	+0.0638 pp
Sum of the three	+3.0262 pp

The sum equals the gap to machine precision. That identity is the only thing the decomposition guarantees, and it is the check: if it does not close, a term has been dropped and it is usually the interaction. Nothing about the identity closing makes the coefficient term a mechanism.

27.7 A coverage curve, both ways

Threshold	Settlements within	Settlement population within	Difference
1 km	54.29 %	91.03 %	+36.74 pp
2 km	86.23 %	96.45 %	+10.23 pp
3 km	95.03 %	99.77 %	+4.74 pp
5 km	98.69 %	99.99 %	+1.30 pp
10 km	99.78 %	100.00 %	+0.22 pp

The third column is above the second at every threshold, and the gap is largest at the tightest one. That is not a discrepancy: the settlements beyond the threshold are disproportionately the small ones. Siting a school is a settlement question and counting unreached children is a people question, so the two columns answer two different decisions — and the population field sums well below the national total, which the third column's label has to say.

28. Case studies

Three, each built on something in this manual. The institutional settings are composed; the arithmetic is not, and each case says which is which. They ship separately in ‘10_Case_Studies/‘ and are summarised here.

Case	Turns on	The failure it teaches
The costing that was three times too large	Chapter 6's denominator finding	A correct number under the wrong denominator label
The map that placed half the schools	Chapter 18's coordinate problem	A choropleth drawn on a non-random subset, published without saying so
The selection correction that tripled the return	Chapter 9's Heckman comparison	A large, significant, unidentified result

28.1 Case 1 — the costing that was three times too large

A ministry costs a fee-abolition proposal using a figure from its own sector's published data audit: about one million shillings per student per year. The arithmetic is a single multiplication and nobody checks the source, because there is no reason to.

Re-deriving from the same file gives 1.03 m UGX per household with students and 375k UGX per student, and the mean such household has 2.69 students. The audit's arithmetic is right and its label names the wrong denominator, so the costing is out by a factor of about 2.75. Nobody miscounted and no reviewer would have caught it from the output.

28.2 Case 2 — the map that placed half the schools

A planning unit publishes a district map of pupil-teacher ratios from the national school register and uses it to prioritise a teacher-deployment round. Six months later a district shaded green complains that its schools are among the most crowded in the country.

The estimation was correct and the join was correct. Only 53.5 % of pupil-bearing schools could be placed inside a district polygon, and the schools that could not are not a random subset — their exclusion moves the aggregate ratio from 49.84 to 48.21. The failure was entirely in what the publication carried alongside the numbers.

28.3 Case 3 — the correction that tripled the return

A ministry economist estimates the return to schooling from a household panel, applies a selection correction because the wage sample is obviously selected, and obtains a return three times the uncorrected figure with a decisively significant selection term. Applying the correction was the textbook response and selection is a real problem on this data.

With the same regressors in both equations the correction is identified by functional form alone: the return rises from 11.8 % to 34.3 % and the inverse Mills coefficient is 5.90. Add household composition as an exclusion restriction and the correction collapses to nothing. The t-statistic gave no warning, because it was testing the wrong thing.

28.4 What the three have in common

- In every one, the arithmetic is correct throughout.
- In every one, a competent reviewer looking only at the output would have seen nothing wrong.
- In every one, what was missing was a single sentence: the denominator, the placement share, or what was excluded from the wage equation.
- In every one, the correction is cheap if made early and expensive once a decision has been taken on the number.

29. Interpretation rules

Twelve rules. Each appears in a chapter, on a slide and in the marking guide. A learner who can apply all twelve has the course.

Rule	Why	Where it is enforced
An estimate is not publishable without its interval, its denominator and its disclosure status	All three come from the design. A number without them implies a precision the sample does not have.	Ch. 4; Laboratory 02; automatic fail 1
Attendance, enrolment and completion are three different quantities	They have different universes and different denominators, and the recode may not support the one you want.	Ch. 3, 5; Laboratory 03; automatic fail 6
Per student and per household are different numbers	They differ by the mean number of students in a household. Say which one you mean, every time.	Ch. 6; Laboratory 04; automatic fail 2
An odds ratio is not a risk ratio and neither is a marginal effect	Report the average marginal effect alongside any coefficient a policy reader will see.	Ch. 4; Laboratory 02; automatic fail 3
A decomposition is an accounting identity	The unexplained share is what the covariates did not account for. It is not a measured cause.	Ch. 7; Laboratory 05
A return to schooling estimated on wage employees is a return within paid employment	Quote the share of workers it covers in the same sentence.	Ch. 9; Laboratory 07
A selection correction needs an exclusion restriction	Without one, identification rests on functional form and the correction can be large, significant and meaningless.	Ch. 9; Laboratory 07
A capacity gap is a statement about the norm	Quote the norm class size with the gap or the gap has no meaning.	Ch. 10, 21; Laboratory 06; automatic fail 4
An optimisation result carries its objective or it is a political choice	Five objectives site schools five ways. Name yours before the recommendation.	Ch. 21; Laboratory 08; automatic fail 4
Straight-line distance is not travel time	It ignores roads, rivers and terrain, and it always flatters access.	Ch. 19; Laboratory 06
A microsimulation counterfactual is associational under the fitted model	It is not a treatment effect and it is not a costed programme.	Ch. 12, 23; Laboratory 08; automatic fail 3

A model that is specified is not a model that is estimated	Say which of the two you are showing. This course does both and labels them.	Ch. 8; Laboratory 01; automatic fail 6

29.1 Which rules are automatic fails, and why those

Seven of the twelve rules have an automatic fail attached. They are not the most sophisticated rules; they are the ones whose breach is invisible in the output. A missing interval looks like a clean number. A per-household figure under a per-student label looks like a per-student figure. A specified model presented as estimated looks exactly like an estimate.

9. Quoting an estimate without its interval, its denominator and its disclosure status.
10. Quoting a per-household figure under a per-student label, or the reverse.
11. Describing an adjusted association as an effect, or presenting a microsimulation counterfactual as a treatment effect.
12. Reporting a capacity gap without the norm class size that generated it, or an optimisation result without its objective.
13. Publishing a map built on the school register without stating the share of schools it could place.
14. Presenting a model this course specifies but does not estimate as though it had been estimated.
15. Reporting any figure from a survey cluster coordinate, or distributing microdata of any kind.

Every one of those is a failure a competent reviewer would not catch from the output alone, which is why the discipline has to sit with the person producing it.

30. The laboratories

Eight laboratories, all running on the estimates and derived tables shipped in ‘08_Data/‘. No laboratory opens microdata, and none needs a network.

#	Laboratory	Chapters	Hours
01	The estate and the catalogue	1–3, 17	2.5
02	Design-based estimation	4	3.0
03	Participation and progression	5, 13	2.5
04	Cost, affordability and poverty	6, 23	3.0
05	Inequality and the gender gap	7	2.5
06	Geography and accessibility	10, 18, 19	3.0
07	Returns, selection and the labour market	9	3.0
08	Projection, optimisation and the handover	11, 12, 20–22, 26	3.0

The laboratory manual carries the instructions; this chapter exists so that a reader of the manual alone knows what the practical component covers.

30.1 What each laboratory establishes

Laboratory 01 — The estate and the catalogue

That a specification is a contract with the data, and that a model with no outcome variable is a procurement statement rather than a failure. The deliverable is an estate note and a three-way triage of all seventy-two cards.

Laboratory 02 — Design-based estimation

That weights fix the middle and only a design-based variance fixes the edges. Learners recover the design effect from two standard errors and quantify what the cluster correction cost every coefficient in a fitted model.

Laboratory 03 — Participation and progression

That a variable means what its code list says, not what its name suggests. Learners read the recode's own codes before using the variable, then compute the gross-net gap and the over-age rate that explain it.

Laboratory 04 — Cost, affordability and poverty

That the denominator decides the number. Five defensible ways to divide the same total give four identical answers and one that is nearly three times larger, and the impoverishing effect follows.

Laboratory 05 — Inequality and the gender gap

That a decomposition is an accounting identity and closes exactly, and that its residual is unexplained rather than explained by anything in particular.

Laboratory 06 — Geography and accessibility

That a coverage curve depends on its denominator, a capacity gap on its norm, and a map on the share of records it could place — three parameters that are usually invisible.

Laboratory 07 — Returns, selection and the labour market

That a return is a return for the people it was estimated on, and that a correction without an exclusion restriction can be large, decisive-looking and meaningless.

Laboratory 08 — Projection, optimisation and the handover

That an objective changes the plan, a contract has fields the run does not fill, and four sources describing one country disagree for four different reasons.

30.2 What every laboratory has in common

- It runs on the shipped estimates and derived tables, never on microdata.
- It has checkpoints that must be true before the learner goes on, and a troubleshooting table for the failures that actually happen.
- It ends with the same four sentences, and they are marked.
- It can be completed without a network, and without any software beyond Julia and the estate's own project environment.

31. Common errors

Error	Where it comes from	What it does
Weighted estimate, unweighted standard error	Applying weights and stopping	Intervals too narrow by the square root of the design effect — up to a factor of three here
Per household quoted per student	A quick description of a correct calculation	Costings out by the mean number of students per household
Attendance read as current enrolment	Not reading the recode's own code list	Overstates current participation by an unbounded amount
Odds ratio read as risk ratio	Reporting the model's native scale	Overstates the programme-relevant change, worst at high prevalence
Capacity gap without the norm	Treating a policy parameter as a fact	A capital programme sized by an unstated assumption
Optimisation result without its objective	Presenting the solver's answer as the answer	A political choice presented as a technical finding
Map drawn on the placeable subset	Not checking what failed to geocode	A choropleth of a biased sample, indistinguishable from a good one
Selection correction without an exclusion restriction	Following a recipe	A large, significant and meaningless correction
National average used for a local decision	The number that was to hand	Sends the intervention to the wrong district
Cross-source difference described as data quality	Assuming one source must be wrong	Discards a source that was correct about something else

31.1 The three that are invisible in the output

Most errors in the table above announce themselves eventually — a number looks wrong, a colleague queries it, a check fails. Three do not, and those are the ones with automatic fails attached.

31.1.1 The missing denominator

A per-household figure under a per-student label produces a number of exactly the right order of magnitude, in the right currency, with the right sign. There is nothing to notice. The only defence is asking the question before the multiplication, which is why it is the first interpretation rule and why two of the seven automatic fails concern it.

31.1.2 The specified model presented as estimated

A slide showing a learning production function's coefficients looks exactly like a slide showing an estimated one. Nothing on the page distinguishes an assumed parameter from a fitted one, and a reader has no way to tell. The defence is the label: this course marks every quantity as estimated, register, projected or specified, and refuses a deliverable that omits it.

31.1.3 The map drawn on the placeable subset

A choropleth of half a register is as clear, as attractive and as authoritative-looking as a choropleth of all of it. The colours are real, the boundaries are right and the quantile classes are sensible. The defence is a line on the map itself giving the placement share and the direction of the bias — not a footnote, and not an annex.

31.2 The order errors get caught in

Caught by	Which errors	How late
The script	Scale errors, missing weights, failed joins	Immediately
A checkpoint	Identities that do not close, sign errors	Within the laboratory
A colleague	Odd magnitudes, missing intervals	Within the week
A reviewer	Overstated claims, missing caveats	Before publication
Nobody	**The three above — until a decision has been taken on the number**	**Months**

32. Governance, disclosure and reproducibility

32.1 What may leave the platform

Artefact	May be published?	Condition
An estimate with its interval and denominator	Yes	Cell at or above the disclosure threshold
A cell below the threshold	**No**	Suppress, and suppress a second cell if the first can be recovered by subtraction
A school point from the register	Yes	It is infrastructure. Attribution as the licence requires
A survey cluster coordinate	**No**	In any form, including a buffer. It stands for households
A microdata record	**No**	Never, under any condition
A district map from a model	Yes	With the share of records it could place, and the direction of the bias from those it could not

32.2 The disclosure threshold in this course

25 unweighted observations. Of 648 estimate cells produced, 0 fall below it and are marked undisclosed. They are computed and stored and not reported, and the decision travels with the cell rather than being applied at publication time — which is what stops it being forgotten.

32.3 Licences that travel with the data

Source	Condition
Survey microdata	Held under the terms of the collecting agency. Not distributed with this course and not required by it
School register	National administrative data, via the spatial infrastructure
OpenStreetMap	Open Database Licence. Attribution required, and it appears on every map in this manual
geoBoundaries	CC BY 4.0. Attribution required
Indicator databases	Open, with the originating agency named

32.4 Reproducibility, defined

A result in this course is reproducible when someone else, given the estate and the same inputs, can regenerate it by running four commands. The digests on the cover are how they check they had the same inputs.

The whole chain

```
cd D:/EDIP/Training/EDIS_Academy/__build
python education_extract.py # reads the estate
python education_analysis.py # runs the estimation
python education_figures.py # redraws every figure
python build_edu501.py # rebuilds the course
```

Extract digest '540f7c8b5d9e3695'. If the estate's source or outputs change, step one changes and every number in the course changes with it. That is the design: the course cannot drift from the model, because it is generated from it.

32.5 What reproducibility is not

REPRODUCIBLE IS NOT VALIDATED
Two implementations of one definition agreeing tells you the arithmetic is right. It does not tell you the definition matches a national report, and this course has no external benchmark. Chapter 26 is the honest version of that conversation: four sources, four numbers, four reasons they differ, and no claim that any one of them is the truth.

Appendices
Glossary; the full card register with every field of all 72 models; the mathematical specification; notation; the module index; commands; the estimate register; references; and the limitations register.

Appendix A. Glossary

Term	Meaning
Attendance ratio, net	Children of the official age for a level who are attending that level, over all children of that age. Bounded by one.
Attendance ratio, gross	Attenders of a level at any age, over the official-age population for that level. Not bounded by one, and its excess over the net ratio is the over-age problem.
Average marginal effect	The average change in the outcome probability for a unit change in a covariate, in the units the outcome is measured in. The scale a programme note carries.
Capacity, effective	The smallest of seat, teacher and desk capacity, scaled by an infrastructure index.
Catastrophic education expenditure	Education spending above a stated share of household consumption. A payment indicator: blind to households that pay nothing because the child is not enrolled.
Cluster	A primary sampling unit. The unit that carries independent information in a multi-stage design, and the unit the variance is summed over.

Term	Meaning
Concentration index	Rank-weighted inequality of an outcome ordered by socioeconomic status. Negative means concentrated among the poor; whether that is good depends entirely on whether the outcome is good.
Design effect	The ratio of the design-based variance to the variance a simple random sample of the same size would have given.
Effective sample size	Unweighted n divided by the design effect. The simple random sample that would have carried the same information.
Exclusion restriction	A variable in the selection equation and not in the outcome equation. Without one, a selection correction is identified only by functional form.
Gini coefficient	Inequality of a distribution. Used here on years of schooling, where it is driven mostly by the share with none at all.
Inverse Mills ratio	The hazard term added to a wage equation to correct for selection into the wage sample.
Mixed-integer programme	An optimisation with at least one variable that must be a whole number. A school is built or it is not.
Norm class size	The pupils-per-classroom a ministry plans to. A policy parameter, and the denominator of every classroom requirement.
Oaxaca–Blinder decomposition	Splitting a group gap into a part explained by differing characteristics and a part explained by differing returns to them. An accounting identity, not a causal attribution.
Optimality gap	The distance between the best solution an integer solver found and the bound it could not close. Must travel with the result.
Over-age for grade	Older than the official age for the grade attended, by a stated margin. The observable consequence of late entry and repetition.
Pupil-teacher ratio	Pupils divided by teachers. An administrative ratio with a coverage problem, not a survey estimate.
Selection	Being observed in a sample for reasons related to the outcome. The central problem of any wage equation in a mostly self-employed economy.
Stratum	A group of clusters sampled independently. Variance is summed within strata and added across them.
Taylor linearisation	Approximating the variance of a ratio by linearising around the estimate. The estimator behind every interval in this manual.
Two-part model	A participation equation for whether any spending occurs, then a level equation for how much. Needed whenever zeros are structural rather than small values.

Appendix B. The card register

All 72 model cards with all eight fields, in family order. Nothing here is written for the course: every field is read from *EDIS Education Modelling Technical Manual, EDIS, 2026* by ‘education_extract.py‘.

HOW TO READ A CARD
Key inputs is the field that decides whether you can run the model at all — a card runs when every variable it names is present. **Core specification** is the estimating equation. **Integration** names the downstream module that consumes the output, which is what makes a card part of an estate rather than a standalone analysis. **Status** is what this course did with it.

B.A — Participation, Progression and Completion

Specification chapter 5; 14 cards; 4 estimated in this course.

Model 1. School enrolment probability

Purpose	Estimate the probability that a school-age child is currently enrolled and identify household, individual and access constraints.
Dependent variable	Binary current enrolment status.
Core specification	Pr(Enroll_i=1)=F(β0 + X_iβ + H_iγ + A_iδ + μ_r)
Key inputs	age and age squared; sex; household consumption/poverty; parental education; household size and dependency ratio; urban/rural and region; distance or travel time to school; education costs and school type
Method	Survey-weighted logit/probit; multilevel logit when cluster or district random effects are required.
Diagnostics	ROC/AUC, Brier score, calibration plot, marginal effects, design-based standard errors, multicollinearity and sensitivity to age band.
Primary outputs	Predicted enrolment probabilities, marginal effects, enrolment-risk maps and policy elasticities.
Integration	Feeds education demand projections, OOSC model, cash-transfer simulation and spatial targeting.
Status in this course	**Estimated**

Model 2. Ever-attendance / school-entry model

Purpose	Model whether children ever enter formal schooling and distinguish delayed entry from permanent exclusion.
Dependent variable	Binary ever-attended indicator among relevant ages.
Core specification	Pr(EverAttend_i=1)=F(X_iβ + H_iγ + Access_iδ)
Key inputs	age; sex; birth order; poverty; parental schooling; distance/ travel time; disability; region
Method	Binary response model with age-cohort interactions.
Diagnostics	Age-profile consistency, cohort checks, weighted goodness-of-fit and predicted-versus-observed entry rates.
Primary outputs	Entry probability by age and subgroup; late-entry prevalence.
Integration	Initializes cohort-flow models and identifies barriers before grade 1.
Status in this course	Specified, not estimated

Model 3. Age-at-entry model

Purpose	Estimate determinants of school starting age and delayed entry.
Dependent variable	Age at first school entry.
Core specification	EntryAge_i = α + X_iβ + H_iγ + ε_i
Key inputs	sex; poverty; parental education; preschool attendance; distance; siblings; rurality
Method	OLS/quantile regression; survival formulation when right-censoring matters.
Diagnostics	Residual distribution, quantile stability, censoring assessment.
Primary outputs	Expected entry age and probability of late entry.
Integration	Improves cohort-flow and school-age population forecasts.
Status in this course	Specified, not estimated

Model 4. Attendance intensity model

Purpose	Explain regularity of attendance among enrolled students.
Dependent variable	Share of scheduled days attended or days absent.
Core specification	E(AttendanceRate_i|X)=g^{-1}(X_iβ)
Key inputs	poverty; school costs; distance; child work; illness proxy; gender; grade; season/agricultural calendar
Method	Fractional logit for proportions; Poisson/negative binomial for absence counts.
Diagnostics	Overdispersion, zero inflation, calibration by grade and season.

Purpose	Explain regularity of attendance among enrolled students.
Primary outputs	Predicted attendance and expected days lost.
Integration	Connects participation to learning, child labour and school feeding simulations.
Status in this course	**Estimated**

Model 5. Chronic absenteeism model

Purpose	Identify students with persistently high absence.
Dependent variable	Binary indicator above a defined absenteeism threshold.
Core specification	Pr(ChronicAbsent_i=1)=F(X_iβ+S_sγ)
Key inputs	attendance history where available; travel time; poverty; household shocks; school calendar; school quality proxies
Method	Logit or gradient boosting for prediction; multilevel model for school effects.
Diagnostics	Precision/recall, calibration, school-level residuals.
Primary outputs	Risk scores and high-risk school/settlement lists.
Integration	Supports targeted attendance interventions.
Status in this course	Specified, not estimated

Model 6. Out-of-school children model

Purpose	Predict which school-age children are not participating in education and quantify geographic concentration.
Dependent variable	OOSC status.
Core specification	Pr(OOSC_i=1)=F(Poverty_i, Access_i, Gender_i, ParentEdu_i, Work_i, Region_i)
Key inputs	age; sex; poverty; disability; child work; parent education; school distance; urban/rural
Method	Survey-weighted logit plus small-area/spatial prediction when census covariates are available.
Diagnostics	Classification metrics, district calibration, uncertainty intervals.
Primary outputs	OOSC count/rate by district and grid cell.
Integration	One of the principal targeting layers in the Education Decision System.
Status in this course	**Estimated**

Model 7. Grade repetition model

Purpose	Estimate the probability that a learner repeats the current grade.
Dependent variable	Repeat indicator.
Core specification	Pr(Repeat_i=1)=F(Student_iβ + Household_iγ + School_ iδ)
Key inputs	age-for-grade; attendance; poverty; previous performance; class size; teacher availability; school type
Method	Logit/probit; multilevel logistic model.
Diagnostics	Rare-event checks, school random-effect variance, sensitivity to definition.
Primary outputs	Repeat risk and expected repeaters by grade.
Integration	Feeds cohort-flow model and cost-per-graduate estimates.
Status in this course	Partly — over-age

Model 8. Promotion model

Purpose	Estimate probability of successful grade promotion.
Dependent variable	Promoted to next grade.
Core specification	Pr(Promote_i=1)=F(X_iβ + S_sγ)
Key inputs	attendance; age-for-grade; learning score; household welfare; school quality; teacher inputs

Purpose	Estimate probability of successful grade promotion.
Method	Binary/multilevel response model.
Diagnostics	Calibration by grade, school and sex; transition consistency.
Primary outputs	Promotion rates and intervention impacts.
Integration	Core transition parameter for system projection.
Status in this course	Specified, not estimated

Model 9. Dropout hazard model

Purpose	Estimate when and why students leave school before completion.
Dependent variable	Time/grade until dropout.
Core specification	h_i(t)=h_0(t) exp(X_iβ)
Key inputs	poverty; costs; distance; age-for-grade; child work; household shocks; gender; marriage risk proxies; school quality
Method	Discrete-time hazard or Cox model; survey weights where supported.
Diagnostics	Proportional-hazards test, competing-risks sensitivity, survival calibration.
Primary outputs	Grade-specific dropout hazards and survival curves.
Integration	Feeds cohort projections and dropout policy simulation.
Status in this course	Specified, not estimated

Model 10. Competing-risks school-exit model

Purpose	Distinguish different destinations after leaving education.
Dependent variable	Exit to work, inactivity, marriage/household work, migration or other state.
Core specification	Pr(J_i=j|exit)=exp(X_iβ_j)/Σ_k exp(X_iβ_k)
Key inputs	age; sex; education level; household status; local labour market; marital status
Method	Multinomial logit or competing-risk survival model.
Diagnostics	IIA tests where relevant, confusion matrix, destination shares.
Primary outputs	Probability of each post-school destination.
Integration	Links education to labour and demographic modules.
Status in this course	Specified, not estimated

Model 11. Primary completion model

Purpose	Estimate determinants of completing primary education.
Dependent variable	Primary completion indicator.
Core specification	Pr(CompletePrimary_i=1)=F(X_iβ)
Key inputs	entry age; sex; poverty; parents education; distance; repetition history; region
Method	Logit/probit with cohort controls.
Diagnostics	Cohort sensitivity, age censoring, weighted calibration.
Primary outputs	Completion probabilities and subgroup gaps.
Integration	Supports SDG-style completion projections and policy targeting.
Status in this course	Specified, not estimated

Model 12. Secondary completion model

Purpose	Estimate determinants of lower/upper secondary completion.
Dependent variable	Secondary completion indicator.
Core specification	Pr(CompleteSecondary_i=1)=F(X_iβ)
Key inputs	primary completion; gender; household resources; school access; school costs; labour demand; region
Method	Logit/probit; sequential models by education stage.

Purpose	Estimate determinants of lower/upper secondary completion.
Diagnostics	Selection and censoring checks, subgroup calibration.
Primary outputs	Secondary completion risk and marginal effects.
Integration	Feeds labour skill supply and CGE factor transition.
Status in this course	Specified, not estimated

Model 13. Tertiary entry model

Purpose	Estimate transition into university/college after secondary school.
Dependent variable	Tertiary participation/entry.
Core specification	Pr(Tertiary_i=1)=F(X_iβ + Price_iγ + Access_iδ)
Key inputs	secondary completion; household income; fees; distance; sex; field availability; urban/rural
Method	Logit/probit or nested choice model.
Diagnostics	Selection on secondary completers, calibration by income quintile.
Primary outputs	Entry probability and distributional incidence.
Integration	Supports tertiary financing and subsidy analysis.
Status in this course	Specified, not estimated

Model 14. Educational attainment ordered model

Purpose	Model attained education as an ordered outcome.
Dependent variable	Ordered categories from no schooling through tertiary.
Core specification	Edu_i*=X_iβ+ε_i; Edu_i=j if κ_{j-1}<Edu_i*≤κ_j
Key inputs	age/cohort; sex; poverty; parents education; location; access
Method	Ordered logit/probit; generalized ordered model if proportional-odds fails.
Diagnostics	Parallel-lines test, threshold ordering, marginal effects.
Primary outputs	Probability of each attainment category.
Integration	Constructs human-capital distributions for household and CGE modules.
Status in this course	**Estimated**

B.B — Household Education Demand, Costs and Affordability

Specification chapter 6; 10 cards; 5 estimated in this course.

Model 15. Years-of-schooling model

Purpose	Estimate completed years of schooling as a compact human-capital outcome.
Dependent variable	Completed schooling years.
Core specification	SchoolYears_i=α+X_iβ+ε_i
Key inputs	cohort; sex; parent education; wealth; rurality; distance; disability
Method	OLS, Poisson/negative binomial or hurdle/count model depending distribution.
Diagnostics	Residuals, overdispersion, heaping and top-coding checks.
Primary outputs	Expected years of schooling and inequality decomposition inputs.
Integration	Feeds Mincer models and human-capital stock.
Status in this course	**Estimated**

Model 16. School-type choice model

Purpose	Model household choice among public, private, religious/community and no school.
Dependent variable	Discrete school-type choice.
Core specification	Pr(Y_i=j)=exp(X_iβ_j)/Σ_k exp(X_iβ_k)

Purpose	Model household choice among public, private, religious/ community and no school.
Key inputs	household income; fees; distance by school type; quality proxies; religion only where ethically and legally appropriate; urban/rural
Method	Multinomial or nested logit.
Diagnostics	IIA/nesting tests, predicted choice shares, price elasticities.
Primary outputs	Choice probabilities and demand response to public/private expansion.
Integration	Supports provider-mix and school-siting scenarios.
Status in this course	Specified, not estimated

Model 17. Household education expenditure two-part model

Purpose	Estimate whether households spend on education and how much they spend conditional on positive expenditure.
Dependent variable	Positive-spending indicator and positive expenditure amount.
Core specification	Pr(E_h>0)=F(Z_hγ); ln(E_h|E_h>0)=X_hβ+ε_h
Key inputs	income/consumption; number of students; school levels; school type; urban/rural; fees and transport
Method	Two-part model; hurdle GLM; survey-weighted estimation.
Diagnostics	Smearing/retransformation, heteroskedasticity, prediction by quintile.
Primary outputs	Expected household education spending and cost incidence.
Integration	Feeds affordability, poverty and fee-reform simulations.
Status in this course	**Estimated**

Model 18. Education budget-share model

Purpose	Estimate the share of household expenditure allocated to education.
Dependent variable	Education expenditure share.
Core specification	w_edu=α+β ln(X/P)+γZ+ε
Key inputs	real total expenditure; prices/fees; number and age of children; school type; location
Method	Working-Leser or fractional regression.
Diagnostics	Bounds, heteroskedasticity, Engel-curve shape.
Primary outputs	Budget share and expenditure elasticity.
Integration	Quantifies household burden and demand response.
Status in this course	**Estimated**

Model 19. Almost Ideal Demand System for education

Purpose	Estimate expenditure and price substitution between education and other household consumption categories.
Dependent variable	Budget shares for multiple commodity groups.
Core specification	w_i=α_i+Σ_j γ_ij ln p_j+β_i ln(X/P)+ε_i
Key inputs	commodity-group prices; real household expenditure; demographics
Method	AIDS/QUAIDS with adding-up, homogeneity and symmetry restrictions.
Diagnostics	Restriction tests, price/expenditure elasticities, system residuals.
Primary outputs	Compensated and uncompensated demand elasticities.
Integration	Supports cost-of-schooling and subsidy policy simulations.
Status in this course	Specified, not estimated

Model 20. Education affordability model

Purpose	Measure financial stress caused by education costs.
Dependent variable	Education burden ratio and high-burden indicator.

Purpose	Measure financial stress caused by education costs.
Core specification	Burden__h=EducationExpenditure__h/TotalConsumption__h
Key inputs	total education costs; household consumption; poverty line; household composition
Method	Descriptive distribution plus logit for high-burden risk.
Diagnostics	Sensitivity to thresholds and denominator definition.
Primary outputs	Burden percentiles, high-burden incidence and poverty-group profiles.
Integration	Identifies households for bursaries/cash transfers.
Status in this course	**Estimated**

Model 21. Catastrophic education expenditure model

Purpose	Estimate incidence of education spending that pushes household non-education resources below a welfare threshold.
Dependent variable	Catastrophic expenditure indicator.
Core specification	Cat__h=1[(C__h-EduExp__h)<z__h]
Key inputs	consumption aggregate; education expenditure; poverty threshold
Method	Microsimulation and survey-weighted incidence.
Diagnostics	Threshold sensitivity and counterfactual checks.
Primary outputs	Incidence, depth and distribution of catastrophic education spending.
Integration	Links education financing to poverty protection.
Status in this course	**Estimated**

Model 22. Fee elasticity of enrolment

Purpose	Estimate responsiveness of participation to direct schooling costs.
Dependent variable	Enrolment or attendance.
Core specification	ln(Q__i)=α+η ln(Fee__i)+X__iβ+ε__i
Key inputs	fees; other costs; income; school quality/access; grade
Method	Logit marginal effects, IV where price endogeneity is credible, or quasi-experimental design.
Diagnostics	Instrument strength where IV is used, placebo and robustness checks.
Primary outputs	Fee elasticity and predicted participation response.
Integration	Parameterizes fee-removal scenarios.
Status in this course	Specified, not estimated

Model 23. Private tutoring demand model

Purpose	Estimate determinants and distribution of household spending on tutoring.
Dependent variable	Tutoring participation and expenditure.
Core specification	Pr(Tutor__i=1)=F(X__iβ); ln(Cost__i|Tutor)=Z__iγ+ε__i
Key inputs	income; exam grade; school type; learning score; urban/rural
Method	Two-part model.
Diagnostics	Selection and expenditure-tail diagnostics.
Primary outputs	Tutoring use and cost by income group.
Integration	Tracks hidden education costs and inequality.
Status in this course	Specified, not estimated

Model 24. School feeding participation model

Purpose	Estimate association/effect of school feeding on attendance and enrolment where program data permit.
Dependent variable	Attendance/enrolment and feeding participation.
Core specification	Y__i=α+τ Feed__i+X__iβ+ε__i

Purpose	Estimate association/effect of school feeding on attendance and enrolment where program data permit.
Key inputs	program exposure; poverty; distance; school fixed effects where possible
Method	Difference-in-differences, matching or multilevel regression depending design.
Diagnostics	Pre-trends if panel/DiD, balance tests, spillover sensitivity.
Primary outputs	Estimated attendance/enrolment response and beneficiary profile.
Integration	Supports costing and targeting of school-feeding programs.
Status in this course	Specified, not estimated

B.C — Child Labour, Social Inclusion and Education Inequality

Specification chapter 7; 8 cards; 2 estimated in this course.

Model 25. Child labour versus schooling multinomial model

Purpose	Model mutually exclusive child activity states.
Dependent variable	School only, work only, school+work, neither.
Core specification	Pr(Y_i=j)=exp(X_iβ_j)/Σ_k exp(X_iβ_k)
Key inputs	age; sex; poverty; farm ownership; season; household labour demand; distance; parent education
Method	Multinomial logit/probit.
Diagnostics	IIA tests, marginal effects, seasonal sensitivity.
Primary outputs	Probability of each activity state.
Integration	Links education to agriculture, household labour and social protection.
Status in this course	Specified, not estimated

Model 26. Child work-hours model

Purpose	Estimate intensive margin of child labour and its schooling trade-off.
Dependent variable	Weekly child work hours.
Core specification	E(Hours_i|X)=exp(X_iβ)
Key inputs	enrolment; farm season; household enterprise; poverty; siblings; gender
Method	Poisson/negative binomial or two-part model.
Diagnostics	Overdispersion, zero mass and seasonality.
Primary outputs	Expected work hours and schooling conflict indicator.
Integration	Inputs to attendance and learning models.
Status in this course	Specified, not estimated

Model 27. Household shock and schooling resilience model

Purpose	Estimate how drought, illness, job loss, price shocks or conflict affect education participation.
Dependent variable	Enrolment, attendance, dropout or expenditure.
Core specification	Y_it=α_i+λ_t+β Shock_it+X_itγ+ε_it
Key inputs	shock exposure; baseline poverty; assets; social protection; location
Method	Panel fixed effects or difference-in-differences when LSMS panel data exist.
Diagnostics	Pre-trends, attrition, shock timing and heterogeneous effects.
Primary outputs	Shock elasticity of schooling and resilience indicators.
Integration	Links climate/agriculture shocks to human-capital loss.
Status in this course	Specified, not estimated

Model 28. Disability and education exclusion model

Purpose	Quantify education gaps associated with disability status.
Dependent variable	Enrolment, attendance, attainment or completion.
Core specification	Y_i=$\alpha+\beta$ Disability_i+X_i$\gamma+\varepsilon$_i
Key inputs	disability domains; age; sex; poverty; access; region
Method	Survey regression with interactions and decomposition.
Diagnostics	Small-sample precision, functional-difficulty definitions.
Primary outputs	Participation and attainment gaps with confidence intervals.
Integration	Supports inclusive-education investment planning.
Status in this course	Specified, not estimated

Model 29. Gender education gap decomposition

Purpose	Separate observed gender gaps into differences in characteristics and coefficients.
Dependent variable	Education outcome gap by sex.
Core specification	ΔY=($\bar{X}$_g-$\bar{X}$_b)β* + $\bar{X}$_g(β_g-β*) + $\bar{X}$_b(β*-β_b)
Key inputs	household characteristics; access; cost; parent education; location
Method	Oaxaca-Blinder for continuous outcomes; nonlinear decomposition for binary outcomes.
Diagnostics	Reference-coefficient sensitivity and reweighting.
Primary outputs	Explained/unexplained portions of gender gaps.
Integration	Feeds gender-targeted policy design.
Status in this course	**Estimated**

Model 30. Rural-urban education gap decomposition

Purpose	Quantify how access, poverty and family characteristics explain rural-urban differences.
Dependent variable	Enrolment/completion/attainment gap.
Core specification	ΔY=Explained+Unexplained
Key inputs	distance; wealth; parent education; school supply; region
Method	Oaxaca-Blinder/reweighting/decomposition.
Diagnostics	Common-support checks and specification sensitivity.
Primary outputs	Drivers of rural disadvantage.
Integration	Connects spatial investments with household outcomes.
Status in this course	**Estimated**

Model 31. Intergenerational education mobility model

Purpose	Estimate persistence of educational attainment across generations.
Dependent variable	Child education years/level.
Core specification	EduChild_i=$\alpha+\beta$_m EduMother_i+β_f EduFather_i+X_i$\gamma+\varepsilon$_i
Key inputs	mother education; father education; child cohort; sex; location; wealth
Method	OLS/ordered models; rank-rank mobility measures.
Diagnostics	Cohort sensitivity and measurement error.
Primary outputs	Intergenerational elasticity and mobility ranks.
Integration	Creates EDIS Education Mobility Index.
Status in this course	Specified, not estimated

Model 32. Sibling competition / resource dilution model

Purpose	Estimate how household child composition influences schooling investments.
Dependent variable	Education expenditure, enrolment or attainment.
Core specification	Y_ih=$\alpha+\beta$ NChildren_h+γ BirthOrder_i+X_i$\delta+\varepsilon$_i
Key inputs	number of children; birth order; age spacing; income; parent education

Purpose	Estimate how household child composition influences schooling investments.
Method	Household fixed effects where sibling-level data permit.
Diagnostics	Within-household identification and age controls.
Primary outputs	Resource-dilution effects.
Integration	Improves household education-demand simulation.
Status in this course	Partly — sibling count

B.D — Learning Outcomes and School Quality

Specification chapter 8; 10 cards; 0 estimated in this course.

Model 33. Learning production function

Purpose	Estimate how student, household, teacher and school inputs combine to produce learning outcomes.
Dependent variable	Standardized test score.
Core specification	Score_is=α+X_iβ+S_sγ+u_s+ε_is
Key inputs	prior achievement if available; household wealth; attendance; teacher qualifications; class size; books; electricity; water/ sanitation
Method	Multilevel linear model; value-added specification with lagged score when panel tests exist.
Diagnostics	ICC, residuals, school-effect stability, omitted-variable sensitivity.
Primary outputs	Input elasticities and predicted learning.
Integration	Links infrastructure and teacher policy to learning outcomes.
Status in this course	Specified, not estimated

Model 34. Foundational literacy probability model

Purpose	Estimate probability of meeting a literacy proficiency threshold.
Dependent variable	Binary literacy proficiency.
Core specification	Pr(Literate_i=1)=F(X_iβ+S_sγ)
Key inputs	age; grade; attendance; home language; books at home; teacher/school inputs
Method	Logit/multilevel logit.
Diagnostics	Classification calibration, threshold sensitivity.
Primary outputs	Literacy risk scores and district prevalence.
Integration	Supports foundational-learning targeting.
Status in this course	Specified, not estimated

Model 35. Foundational numeracy probability model

Purpose	Estimate probability of meeting a numeracy proficiency threshold.
Dependent variable	Binary numeracy proficiency.
Core specification	Pr(Numerate_i=1)=F(X_iβ+S_sγ)
Key inputs	grade; attendance; household resources; teacher maths training; class size
Method	Logit/multilevel logit.
Diagnostics	Calibration and domain reliability.
Primary outputs	Numeracy proficiency profiles.
Integration	Supports learning recovery and teacher training.
Status in this course	Specified, not estimated

Model 36. Learning inequality decomposition

Purpose	Measure dispersion in learning and attribute it to between- and within-group components.
Dependent variable	Test score distribution.
Core specification	Var(Y)=Var(E[Y|G])+E[Var(Y|G)]
Key inputs	scores; school; district; wealth quintile; gender
Method	Variance decomposition, Theil-type indices, quantile gaps.
Diagnostics	Sampling uncertainty and test-scale comparability.
Primary outputs	Within/between school and socioeconomic learning gaps.
Integration	Prioritizes quality interventions.
Status in this course	Specified, not estimated

Model 37. Teacher effectiveness/value-added model

Purpose	Estimate teacher contribution to student learning where repeated scores and teacher links exist.
Dependent variable	Achievement gain.
Core specification	Score_it=α+ρ Score_i,t-1+θ_teacher+X_iβ+ε_it
Key inputs	lagged score; teacher id; student covariates; school fixed effects
Method	Teacher value-added model with shrinkage/empirical Bayes.
Diagnostics	Stability across years, classroom sorting tests, minimum sample sizes.
Primary outputs	Teacher-effect distributions with uncertainty.
Integration	Supports teacher development, not punitive ranking.
Status in this course	Specified, not estimated

Model 38. Class-size effect model

Purpose	Estimate association or causal effect of class size on learning.
Dependent variable	Test score.
Core specification	Score_is=α+β ClassSize_s+X_iγ+ε_is
Key inputs	class size; grade; teacher quality; school resources; student background
Method	Multilevel regression; IV/RD only when a credible rule creates exogenous class-size variation.
Diagnostics	Endogeneity discussion, nonlinearities, robustness.
Primary outputs	Marginal learning impact of class size.
Integration	Feeds classroom investment planning.
Status in this course	Specified, not estimated

Model 39. Textbook and learning-material effect model

Purpose	Estimate learning differences associated with access to textbooks/materials.
Dependent variable	Test score/literacy/numeracy.
Core specification	Y_i=α+β Materials_i+X_iγ+S_sδ+ε_i
Key inputs	textbook access; student-book ratio; grade; household and school inputs
Method	Multilevel regression; program evaluation design if applicable.
Diagnostics	Confounding and school fixed-effect checks.
Primary outputs	Expected learning gains by resource level.
Integration	Supports procurement optimization.
Status in this course	Specified, not estimated

Model 40. School infrastructure quality index

Purpose	Construct a multidimensional measure of school infrastructure.
Dependent variable	Composite school infrastructure score.
Core specification	Infra_s=Σ_k w_k z_{ks}

Purpose	Construct a multidimensional measure of school infrastructure.
Key inputs	classroom condition; WASH; electricity; internet; laboratory; library; furniture
Method	PCA/factor analysis or transparent policy weights.
Diagnostics	Internal consistency, loading stability, missingness.
Primary outputs	Infrastructure index and component gaps.
Integration	Used in learning, access and capital-budget models.
Status in this course	Specified, not estimated

Model 41. School quality latent-factor model

Purpose	Estimate a latent school-quality measure from multiple observed outcomes and inputs.
Dependent variable	Latent quality factor.
Core specification	y_{ks}=λ_k Q_s+ε_{ks}
Key inputs	learning; attendance; completion; teacher presence; infrastructure
Method	Confirmatory factor analysis / item-response style latent model.
Diagnostics	Fit indices, factor reliability and invariance.
Primary outputs	Comparable school-quality score.
Integration	Provides quality weights for accessibility and school choice models.
Status in this course	Specified, not estimated

Model 42. Learning-adjusted years of schooling

Purpose	Combine schooling quantity and learning quality into a human-capital metric.
Dependent variable	Learning-adjusted schooling years.
Core specification	LAYS=ExpectedYearsSchooling × (LearningIndex/ Benchmark)
Key inputs	expected schooling years; standardized learning score; benchmark
Method	Deterministic index with uncertainty propagation.
Diagnostics	Benchmark sensitivity and score comparability.
Primary outputs	Learning-adjusted schooling by cohort/region.
Integration	Stronger human-capital input for growth/CGE than years alone.
Status in this course	Specified, not estimated

B.E — Education, Employment, Earnings and Skills

Specification chapter 9; 10 cards; 7 estimated in this course.

Model 43. Mincer returns-to-schooling model

Purpose	Estimate wage returns to each additional year of schooling.
Dependent variable	Log hourly or monthly wage.
Core specification	ln(w_i)=α+β SchoolYears_i+γ Exp_i+δ Exp_i^{2}+X_ iθ+ε_i
Key inputs	schooling years; potential/actual experience; sex; region; occupation; industry; formal status
Method	Survey-weighted OLS; quantile regression for distributional heterogeneity.
Diagnostics	Heteroskedasticity, functional form, wage outliers and sample selection.
Primary outputs	Return per year of schooling and subgroup returns.
Integration	Converts education outcomes into household income and productivity effects.
Status in this course	**Estimated**

Model 44. Returns by education level model

Purpose	Estimate wage premia for primary, secondary, TVET and tertiary education.
Dependent variable	Log wage.
Core specification	ln(w_i)=α+Σ_e β_e D_{ie}+X_iγ+ε_i
Key inputs	education-level dummies; experience; sex; occupation; industry
Method	OLS/quantile regression.
Diagnostics	Reference-category sensitivity and sample selection.
Primary outputs	Education-level wage premia.
Integration	Maps skill categories to CGE factor wages.
Status in this course	**Estimated**

Model 45. Heckman wage-selection model

Purpose	Correct wage equations for non-random selection into observed wage employment.
Dependent variable	Employment selection and conditional log wage.
Core specification	Pr(Employed_i=1)=Φ(Z_iγ); ln(w_i)=X_iβ+ρλ_i+ε_i
Key inputs	education; experience; household composition; non-labour income; local labour conditions
Method	Heckman two-step or full-information ML.
Diagnostics	Exclusion restriction, inverse-Mills significance, sensitivity.
Primary outputs	Selection-corrected returns to schooling.
Integration	Improves labour-income microsimulation.
Status in this course	**Estimated**

Model 46. Employment probability model

Purpose	Estimate how education changes probability of employment.
Dependent variable	Employment status.
Core specification	Pr(Employed_i=1)=F(Edu_iβ+X_iγ)
Key inputs	education; age; sex; region; marital status; household context
Method	Logit/probit with survey design.
Diagnostics	Calibration, marginal effects and subgroup heterogeneity.
Primary outputs	Employment probability by education level.
Integration	Links skill supply to labour-market absorption.
Status in this course	**Estimated**

Model 47. Formal-employment probability model

Purpose	Estimate probability that workers enter formal employment.
Dependent variable	Formal versus informal employment.
Core specification	Pr(Formal_i=1)=F(Edu_iβ+X_iγ)
Key inputs	education level; occupation; industry; urban/rural; sex; age
Method	Logit/probit.
Diagnostics	Definition sensitivity and sector interactions.
Primary outputs	Formalization premium associated with education.
Integration	Supports tax-base and productivity channels in CGE/fiscal models.
Status in this course	**Estimated**

Model 48. Education-to-occupation multinomial model

Purpose	Estimate occupational destination conditional on education and demographics.
Dependent variable	Occupation group.
Core specification	Pr(Occupation_i=j)=exp(X_iβ_j)/Σ_k exp(X_iβ_k)
Key inputs	education level/field; age; sex; region; experience

Purpose	Estimate occupational destination conditional on education and demographics.
Method	Multinomial logit or nested occupational choice.
Diagnostics	IIA/nesting tests and confusion matrix.
Primary outputs	Occupation probabilities by education.
Integration	Transforms graduate projections into occupation supply.
Status in this course	**Estimated**

Model 49. Education-to-industry allocation model

Purpose	Estimate distribution of workers across economic sectors by education.
Dependent variable	Industry of employment.
Core specification	$\Pr(\text{Sector}_i=s)=\exp(X_i\beta_s)/\Sigma_k \exp(X_i\beta_k)$
Key inputs	education; occupation; region; sex; age
Method	Multinomial model or empirical transition matrix.
Diagnostics	Predicted-versus-observed sector shares.
Primary outputs	Education-by-sector labour supply matrix.
Integration	Direct input to SAM/CGE factor allocation.
Status in this course	Specified, not estimated

Model 50. Skills mismatch model

Purpose	Measure overeducation, undereducation and match relative to occupational requirements.
Dependent variable	Mismatch category.
Core specification	$\text{Mismatch}_i = \text{compare}(\text{Edu}_i, \text{RequiredEdu}_{\text{occupation}})$
Key inputs	education level; occupation; wage; industry
Method	Realized-matches approach, modal education or external occupation-skill mapping.
Diagnostics	Alternative required-education definitions.
Primary outputs	Mismatch rates and wage penalties.
Integration	Supports TVET/tertiary planning and productivity analysis.
Status in this course	Specified, not estimated

Model 51. Youth NEET model

Purpose	Estimate probability that youth are not in employment, education or training.
Dependent variable	NEET indicator.
Core specification	$\Pr(\text{NEET}_i=1)=F(\text{Edu}_i\beta+\text{Household}_i\gamma+\text{Location}_i\delta)$
Key inputs	age; sex; education; wealth; marital status; disability; local labour conditions
Method	Survey-weighted logit/probit.
Diagnostics	Calibration and age-band sensitivity.
Primary outputs	NEET risk and geographic concentration.
Integration	Connects education, youth employment and social policy.
Status in this course	**Estimated**

Model 52. School-to-work transition duration model

Purpose	Estimate time between leaving education and first stable employment.
Dependent variable	Duration to employment.
Core specification	$h_i(t)=h_0(t)\exp(\text{Edu}_i\beta+X_i\gamma)$
Key inputs	education; sex; location; field of study; household resources
Method	Survival model; competing risks for formal/informal employment.
Diagnostics	Censoring and proportionality checks.
Primary outputs	Expected transition duration by education level.
Integration	Supports graduate absorption and labour-market scenarios.

Purpose	Estimate time between leaving education and first stable employment.
Status in this course	Specified, not estimated

B.F — Spatial Education Planning and Optimization

Specification chapter 10; 9 cards; 3 estimated in this course.

Model 53. Nearest-school accessibility model

Purpose	Measure travel distance/time from settlements or households to nearest school by level.
Dependent variable	Network travel time/distance.
Core specification	Access_i=min_j TravelTime(i,j)
Key inputs	road/path network; school coordinates; settlement/ population grid; terrain where available
Method	GIS network analysis / least-cost path.
Diagnostics	Coordinate validation, road-network completeness and spot checks.
Primary outputs	Travel time to nearest primary/secondary/TVET school.
Integration	Core spatial determinant used in enrolment and siting models.
Status in this course	**Estimated**

Model 54. Gravity-based education accessibility index

Purpose	Measure access to multiple schools weighted by capacity and travel impedance.
Dependent variable	Continuous accessibility score.
Core specification	A_i=Σ_j Capacity_j exp(-λ T_ij)
Key inputs	school capacity; travel time matrix; decay parameter
Method	Spatial gravity model with calibrated decay.
Diagnostics	Sensitivity to λ and capacity definitions.
Primary outputs	Accessibility surfaces and population-weighted district scores.
Integration	Improves on nearest-facility indicators.
Status in this course	Specified, not estimated

Model 55. School catchment demand model

Purpose	Estimate school-age population potentially served by each school.
Dependent variable	Potential students in catchment.
Core specification	Demand_s=Σ_{i∈C_s} PopSchoolAge_i
Key inputs	population grid by age; travel-time catchments; school level
Method	GIS catchment overlay or Voronoi/network catchments.
Diagnostics	Overlap rules, boundary effects, enrolment comparison.
Primary outputs	Catchment population and demand pressure.
Integration	Feeds capacity-gap and siting optimization.
Status in this course	**Estimated**

Model 56. School capacity-gap model

Purpose	Compare potential or actual demand with seats/classroom capacity.
Dependent variable	Capacity surplus/deficit.
Core specification	Gap_s=Demand_s-Capacity_s
Key inputs	enrolment; classrooms; target class size; catchment population
Method	Deterministic calculation with scenario thresholds.
Diagnostics	Capacity-definition sensitivity.
Primary outputs	Seat/classroom deficits by school/district.

Purpose	Compare potential or actual demand with seats/classroom capacity.
Integration	Prioritizes capital investments.
Status in this course	**Estimated**

Model 57. Optimal new-school location model

Purpose	Choose school sites that maximize coverage/minimize travel subject to a capital budget.
Dependent variable	Binary site decisions and assignment.
Core specification	min Σ_i Pop_i d_{i,j(i)} subject to Σ_j Cost_j x_j≤B
Key inputs	candidate sites; population demand; travel-time matrix; site costs; existing schools
Method	p-median, maximal covering or capacitated facility-location MILP.
Diagnostics	Optimality gap, alternative budgets, site robustness.
Primary outputs	Recommended school locations, beneficiaries and cost.
Integration	Flagship EDIS infrastructure optimization application.
Status in this course	Illustrated

Model 58. Equity-constrained school siting model

Purpose	Ensure new facilities improve access for poor/remote groups, not only total coverage.
Dependent variable	Site decisions with equity constraints.
Core specification	min Cost/Travel subject to Coverage_poor≥τ and Coverage_remote≥κ
Key inputs	poverty maps; remote-area flags; population; candidate sites
Method	Multi-objective MILP.
Diagnostics	Pareto frontier and equity-efficiency trade-off.
Primary outputs	Equity-adjusted investment plan.
Integration	Links poverty targeting to infrastructure planning.
Status in this course	Specified, not estimated

Model 59. Teacher allocation optimization

Purpose	Allocate teachers across schools to reduce pupil-teacher disparities.
Dependent variable	Teacher assignments.
Core specification	min Σ_s |Students_s/T_s - PTR*| subject to Σ_s T_s=T_available
Key inputs	enrolment; existing teachers; subject specialization; minimum staffing; transfer constraints
Method	Linear/MILP approximation of staffing objective.
Diagnostics	Feasibility, transfer limits, alternative target PTRs.
Primary outputs	Optimal teacher distribution and shortage list.
Integration	Supports staffing budgets and teacher deployment.
Status in this course	Specified, not estimated

Model 60. Subject-teacher allocation model

Purpose	Allocate secondary teachers by subject to meet timetable and curriculum requirements.
Dependent variable	Teacher-subject-school assignment.
Core specification	min Σ_{s,k} Shortage_{sk} subject to teacher qualification and workload constraints
Key inputs	teacher qualifications; subjects; classes; period requirements; locations
Method	Assignment/MILP.
Diagnostics	Constraint violations and workload balance.
Primary outputs	Subject-specific staffing gaps.

Purpose	Allocate secondary teachers by subject to meet timetable and curriculum requirements.
Integration	Supports STEM/TVET staffing strategy.
Status in this course	Specified, not estimated

Model 61. School consolidation/expansion model

Purpose	Choose whether to expand, rehabilitate, consolidate or close facilities under demographic change.
Dependent variable	Facility investment decisions.
Core specification	min TotalCost + λ TravelBurden subject to capacity and access standards
Key inputs	projected population; school condition; operating costs; travel times; expansion costs
Method	Multi-period facility planning optimization.
Diagnostics	Scenario sensitivity and transition costs.
Primary outputs	Long-run school network plan.
Integration	Integrates demographic projections and capital budgeting.
Status in this course	Specified, not estimated

B.G — Education System Projection, Costing and Public Finance

Specification chapter 11; 7 cards; 1 estimated in this course.

Model 62. Education cohort-flow projection model

Purpose	Project enrolment by grade using promotion, repetition, dropout and entry rates.
Dependent variable	Enrolment by grade and year.
Core specification	E_{g,t+1}=Promoted_{g-1,t}+Repeaters_{g,t}+NewEntrants_{g,t+1}
Key inputs	base enrolment; entry; promotion; repetition; dropout; population
Method	Deterministic cohort-component simulation with stochastic uncertainty option.
Diagnostics	Identity checks, historical backcast and transition consistency.
Primary outputs	Enrolment, entrants, repeaters, dropouts and graduates by year.
Integration	Core national education system projection engine.
Status in this course	From the run

Model 63. Demographic education-demand model

Purpose	Translate population projections into potential demand for each education level.
Dependent variable	School-age population and potential enrolment.
Core specification	PotentialEnroll_{l,t}=Pop_{age(l),t} × ParticipationRate_{l,t}
Key inputs	UN/national population projections; age-specific participation; entry-age rules
Method	Cohort-component demographic projection interface.
Diagnostics	Population-source reconciliation and age-band sensitivity.
Primary outputs	Future student demand by district and level.
Integration	Drives teacher, classroom and financing requirements.
Status in this course	From the run

Model 64. Teacher requirement projection

Purpose	Project teachers needed to meet target pupil-teacher ratios and subject requirements.
Dependent variable	Required teachers.
Core specification	TeachersReq_{l,t}=Enroll_{l,t}/PTR_l
Key inputs	projected enrolment; target PTR; teacher attrition; training pipeline
Method	Stock-flow projection.
Diagnostics	Attrition and recruitment sensitivity.
Primary outputs	Annual teacher recruitment need and wage-bill implications.
Integration	Links education planning to fiscal module.
Status in this course	From the run

Model 65. Classroom requirement projection

Purpose	Estimate future classroom requirements by level and geography.
Dependent variable	Required classrooms.
Core specification	ClassroomsReq_{l,t}=Enroll_{l,t}/TargetClassSize_l
Key inputs	enrolment projection; existing classrooms; condition/rehab status
Method	Stock-flow capital needs model.
Diagnostics	Utilization and double-shift sensitivity.
Primary outputs	Construction/rehabilitation requirements.
Integration	Feeds public investment planning.
Status in this course	**Estimated**

Model 66. Education unit-cost model

Purpose	Estimate recurrent and capital cost per student and per graduate.
Dependent variable	Unit cost and cost per graduate.
Core specification	UnitCost_l=Spending_l/Students_l; CostGraduate_ l=Spending_l/Graduates_l
Key inputs	budget/expenditure; students; graduates; teachers; capital spending
Method	Accounting model with inflation/price normalization.
Diagnostics	Classification reconciliation and denominator sensitivity.
Primary outputs	Unit costs by level and cost-efficiency ratios.
Integration	Inputs to budget projection and optimization.
Status in this course	From the run

Model 67. Education expenditure projection model

Purpose	Project public education budget required under policy targets.
Dependent variable	Total education spending by level/year.
Core specification	Budget_t=Σ_l Students_{l,t}×UnitCost_ {l,t}+CapitalNeed_t
Key inputs	student projections; unit costs; teacher wages; capital plans; inflation
Method	Medium-term fiscal simulation.
Diagnostics	Budget identity and price-base checks.
Primary outputs	Required spending, financing gaps and share of GDP/ revenue.
Integration	Connects education system targets to macro-fiscal framework.
Status in this course	From the run

Model 68. Benefit incidence of public education spending

Purpose	Allocate public education subsidies across household welfare groups.
Dependent variable	Per-household/quantile imputed education subsidy.
Core specification	Benefit_h=Σ_l Enrolment_{hl}×UnitSubsidy_l
Key inputs	household enrolment; public/private status; unit subsidy by level; weights
Method	Benefit-incidence analysis with concentration curves.
Diagnostics	Incidence sensitivity to unit-subsidy assumptions.
Primary outputs	Share of subsidy accruing to each quintile and concentration index.
Integration	Evaluates progressivity of education budgets.
Status in this course	Specified, not estimated

B.H — Policy Microsimulation, Human Capital and CGE Integration

Specification chapter 12; 4 cards; 2 estimated in this course.

Model 69. Cash-transfer education microsimulation

Purpose	Simulate effect of household transfers on enrolment, attendance and poverty.
Dependent variable	Counterfactual education and welfare outcomes.
Core specification	ΔP_i=P(Enroll|Income+Transfer)-P(Enroll|Income)
Key inputs	estimated enrolment model; transfer design; household welfare; eligibility rules
Method	Static microsimulation with probabilistic or expected-value response.
Diagnostics	Baseline replication, take-up sensitivity and confidence intervals.
Primary outputs	Additional students, fiscal cost, poverty and targeting metrics.
Integration	Links social protection, education and poverty modules.
Status in this course	**Estimated**

Model 70. Fee-removal and bursary simulation

Purpose	Simulate changes in participation and household welfare when school fees/costs are reduced.
Dependent variable	Counterfactual enrolment, education spending and consumption.
Core specification	P_i^1=F(Fee_i-Subsidy_i,X_i); Welfare_h^1=C_h+Savings_h-ResidualCosts_h
Key inputs	fee elasticity/model; baseline fees; subsidy design; household expenditure
Method	Microsimulation with behavioral response.
Diagnostics	Elasticity scenarios and fiscal-cost reconciliation.
Primary outputs	Participation gains, beneficiaries, distributional and fiscal effects.
Integration	Supports free-education and targeted bursary evaluation.
Status in this course	Specified, not estimated

Model 71. Education-human-capital dynamic stock model

Purpose	Translate education flows into labour-force skill stocks over time.
Dependent variable	Labour force by education/skill category.
Core specification	L_{e,t+1}=L_{e,t}+Graduates_{e,t}-Retirements_{e,t}-NetMigration_{e,t}
Key inputs	graduates; age-specific labour participation; retirement; migration; skill mapping
Method	Dynamic stock-flow simulation.
Diagnostics	Mass-balance checks and labour-force reconciliation.
Primary outputs	Future skill supply by category and cohort.

Purpose	Translate education flows into labour-force skill stocks over time.
Integration	Principal bridge from education projections to dynamic CGE factors.
Status in this course	Specified, not estimated

Model 72. Integrated education-CGE-poverty-budget model

Purpose	Evaluate economy-wide and distributional effects of education spending and reallocation.
Dependent variable	GDP, sector output, factor wages, employment, fiscal balance, household welfare and poverty.
Core specification	EduSpend_t → Capacity_t → Graduates_{e,t} → L_{e,t} & A_{s,t} → CGE equilibrium → HouseholdIncome_h → Poverty_h ; Share_{h,e} = LabourIncome_{h,e} / Σ_e LabourIncome_{h,e} ; Y_h^1 = Σ_e (1+Δw_e)Y_{h,e}^0 + Transfers_h^1 + OtherIncome_h^1 ; RealConsumption_h^1 = NominalResources_h^1 / HouseholdPriceIndex_h^1 ; EVI_i = w1 Poverty_i + w2 AccessDeficit_i + w3 OOSCRisk_i + w4 DropoutRisk_i + w5 CapacityGap_i + w6 LearningDeficit_i ; EDIS_EDUCATION/ data_raw/ data_harmonized/ metadata/ src/ harmonization/ participation/ household_demand/ learning/ labour/ spatial/ optimization/ projections/ microsimulation/ cge_link/ diagnostics/ config/ scenarios/ results/ tests/ docs/
Key inputs	education budget scenario; cohort-flow outputs; skill supply; returns/productivity parameters; SAM/CGE data; household factor-income shares
Method	Recursive dynamic CGE linked iteratively or sequentially to household microsimulation; scenario-based fiscal closure.
Diagnostics	SAM balance, CGE replication, closure report, labour-factor reconciliation, microsimulation income matching, scenario decomposition and sensitivity.
Primary outputs	Economy-wide growth, sector effects, skill wages, household welfare, poverty, inequality, tax revenue and fiscal returns to education.
Integration	Top-level EDIS policy engine for comparing education investments with alternative public spending uses.
Status in this course	**Estimated**

Appendix C. The mathematical specification

The estate's consolidated specification, block by block, as it is written in the source. Chapter references show where each block is taught.

C.Population

Taught in chapter 13.

Population
N[r,a+1,t+1] = N[r,a,t](1-m[r,a,t]) + M[r,a,t]
(98)

C.Access and enrolment

Taught in chapter 13.

Access and enrolment
P_access = logistic(alpha0 + alpha1 income - alpha2 travel_time + policy - poverty and disability penalties)
Demand = school_age_population * P_access
Enrolment = min(Demand, Capacity)
(99)

C.Grade flow

Taught in chapter 13.

Grade flow
Promotion + Repetition + Dropout = 1
Completers = final_grade_enrolment * promotion_rate
Transitions are bounded by next-level capacity.
(100)

C.Supply

Taught in chapter 14.

Supply
SeatCapacity = Classrooms * NormClassSize * Shifts
TeacherCapacity = Teachers * PTRTarget
EffectiveCapacity = min(SeatCapacity, TeacherCapacity, DeskCapacity) * InfrastructureIndex
(101)

C.Teachers

Taught in chapter 14.

Teachers
Teachers[t+1] = Teachers[t] + Graduates + Recruits - Attrition + Migration
(102)

C.Learning

Taught in chapter 14 (and 8).

Learning
Score[t+1] = Score[t] + beta1 instructional_time + beta2 attendance + beta3 teacher_quality + beta4 materials + beta5 digital_access - beta6 class_size
(103)

C.Finance

Taught in chapter 15.

Finance
Spend = UnitCost*Enrolment + Wage*Teachers + CapitalSpending
Spend <= Revenue + DonorFunds + HouseholdSpending + Borrowing
(104)

C.Labour market

Taught in chapter 9 and 22.

Labour market
SkillSupply = Completers * EmploymentRate
SkillGap = max(0, Demand - Supply)
(105)

C.School location MILP

Taught in chapter 20, 21.

School location MILP
min fixed construction cost + weighted student travel time
subject to student assignment, school capacities, opening decisions and maximum travel time.
(106)

C.Teacher deployment MILP

Taught in chapter 20, 21.

Teacher deployment MILP
min weighted shortages
subject to available teachers, regional-subject demand, salary budget and equity floors.
(107)

C.Budget allocation LP

Taught in chapter 20.

Budget allocation LP
max learning + completion + equity benefits
subject to total budget, programme scale limits and regional equity floors.
(108)

C.Student transport LP

Taught in chapter 20.

Student transport LP
min transport cost
subject to all students being assigned and school capacities.
(109)

C.Transition planning LP

Taught in chapter 20.

Transition planning LP
min unmet transition targets
subject to completers, existing capacity, expansion and capital budget.
(110)

C.Multi-objective policy LP

Taught in chapter 20, 21.

Multi-objective policy LP max weighted learning, completion, equity and employment gains subject to the national education budget. (111)	

C. The specification's own equations

The specification document carries 78 displayed equations, one on most model cards. They appear on the cards in Appendix B rather than being repeated here.

Appendix D. Notation

Symbol	Meaning
w__{hij}	Sampling weight of person j in cluster i of stratum h
h, i, j	Stratum, primary sampling unit, person
n__h	Number of sampling units in stratum h
$\hat{R}$	Weighted ratio estimator
z__{hij}	Linearised value used in the variance
deff	Design effect
n__eff	Effective sample size
β	Regression coefficient, log-odds scale for a logit
OR	Odds ratio, exp(β)
AME	Average marginal effect
λ	Inverse Mills ratio
N[r,a,t]	Population in region r, age a, year t
m[r,a,t]	Mortality rate
M[r,a,t]	Net migration
C__eff	Effective capacity
PTR	Pupil-teacher ratio
PCR	Pupil-classroom ratio
τ	Distance or travel-time threshold
γ	Distance-decay parameter in a gravity model

Appendix E. The estate, module by module

90 Julia modules across 30 layers, 11,183 lines. This is the map from a number to the file that produced it.

Layer	Module	Lines	What the layer owns
Data	‘SchoolHarmonizer.jl‘	240	Loaders and harmonisers: DHS, LSMS, schools, catalogues
	‘Schemas.jl‘	200	
	‘HouseholdHarmonizer.jl‘	189	
	‘SchoolLoader.jl‘	182	
	‘HouseholdLoader.jl‘	175	
	‘HouseholdSurveyCatalog.jl‘	165	
	‘UgandaLSMS.jl‘	162	
	‘LSMSHarmonizer.jl‘	156	
	‘DataValidation.jl‘	117	
	‘LSMSLoader.jl‘	64	
	‘Loader.jl‘	4	
Services	‘Services.jl‘	957	Job service, scenario store, runner
	‘JobService.jl‘	82	
	‘ScenarioStore.jl‘	67	
	‘AuditLog.jl‘	54	
Web	‘Service.jl‘	1,011	The Genie application
	‘CanonicalRoutes.jl‘	51	
econometrics	‘DiscreteChoice.jl‘	250	The discrete-choice and regression engine behind the population-education models
	‘Equity.jl‘	217	

Layer	Module	Lines	What the layer owns
	‘ExpenditureModel.jl‘	148	
	‘LinearModels.jl‘	111	
	‘AttainmentModel.jl‘	105	
	‘EnrolmentModel.jl‘	66	
	‘AttendanceModel.jl‘	58	
integration	‘SDIBridge.jl‘	197	CGE, MTEF, BOE, SDI, unit-cost and cross-sector links
	‘MTEFBridge.jl‘	193	
	‘CGEExchange.jl‘	185	
	‘CrossSector.jl‘	146	
	‘UnitCostBridge.jl‘	108	
	‘BOEBridge.jl‘	73	
Core	‘Indicators.jl‘	226	Configuration, types, validation, scenarios, run context
	‘Config.jl‘	141	
	‘Auth.jl‘	112	
	‘Validation.jl‘	93	
	‘RunContext.jl‘	82	
	‘Scenario.jl‘	70	
	‘RunLog.jl‘	53	
	‘Paths.jl‘	36	
	‘Types.jl‘	25	
Models	‘CohortFlow.jl‘	67	The education system engine: demography, cohort flow, enrolment, supply, teachers, finance, learning, labour
	‘Demography.jl‘	63	
	‘PopulationProjection.jl‘	62	
	‘Enrolment.jl‘	60	
	‘Learning.jl‘	57	
	‘Supply.jl‘	53	
	‘GradeFlow.jl‘	52	
	‘Finance.jl‘	50	
	‘Teachers.jl‘	50	
	‘LabourMarket.jl‘	49	
	‘Impacts.jl‘	48	
	‘EducationIndicators.jl‘	44	
	‘Requirements.jl‘	40	
GIS	‘SpatialData.jl‘	276	Spatial data, road network and climate risk
	‘Network.jl‘	81	
	‘ClimateRisk.jl‘	28	
finance	‘BudgetLoader.jl‘	147	Costing, budget and MTEF
	‘Costing.jl‘	122	
	‘MTEF.jl‘	93	
Optimization	‘FacilityLocation.jl‘	145	Mathematical programmes: school siting, teacher deployment, transport, budget, portfolio, transition
	‘Diagnostics.jl‘	71	
	‘InvestmentPortfolio.jl‘	37	
	‘SchoolLocation.jl‘	14	
	‘TeacherDeployment.jl‘	12	
	‘TransitionPlanning.jl‘	11	
	‘BudgetAllocation.jl‘	10	
	‘StudentTransport.jl‘	9	
	‘MultiObjectivePolicy.jl‘	8	
accessibility	‘RoadNetwork.jl‘	134	Road-network routing and catchment access
	‘Accessibility.jl‘	122	
enrolment	‘ProgressionModel.jl‘	217	Enrolment forecasting
demography	‘Population.jl‘	197	Population projection
learning	‘LearningModel.jl‘	172	Learning models

Layer	Module	Lines	What the layer owns
dropout	‘DropoutModel.jl‘	149	Dropout and progression
costbenefit	‘CostBenefit.jl‘	144	Cost-benefit and rate of return
teachers	‘TeacherDemand.jl‘	137	Teacher demand and deployment
CDCGE	‘Bridge.jl‘	135	The dynamic CGE bridge
scenarios	‘ScenarioEngine.jl‘	122	Scenario engine
prioritization	‘MCDA.jl‘	121	Multi-criteria decision analysis
diagnostics	‘DataQuality.jl‘	108	Solver and model diagnostics
forecasting	‘EnrolmentForecast.jl‘	105	Forecast engine
reporting	‘Report.jl‘	105	Report generation
microsim	‘Microsimulation.jl‘	95	Household policy microsimulation
pipeline	‘Pipeline.jl‘	93	The end-to-end pipeline
tertiary	‘PostBasic.jl‘	89	Post-basic and tertiary
perf	‘Benchmark.jl‘	73	Performance harness
Simulation	‘Engine.jl‘	71	The run engine
acceptance	‘Acceptance.jl‘	60	Acceptance checks
API	‘app.jl‘	4	REST routes
(top level)	‘EDISEducation.jl‘	318	Package entry points
(top level)	‘IntegratedEducationSystem Model.jl‘	82	Package entry points

E.2 The documentation set

Document	Title	Command
‘ACCEPTANCE.md‘	End-to-End Multi-Country Acceptance Run	33
‘ACCESSIBILITY.md‘	School Accessibility & Catchment Areas	14
‘ARCHITECTURE.md‘	EDIS Education — Architecture	01 — Canonical architecture (additive)
‘ATTAINMENT_ EXPENDITURE.md‘	Attainment & Household Education-Expenditure Models	09
‘BASELINE_TEST_STATUS.md‘	Baseline Test Status	00 — baseline capture (no code changed)
‘CGE_EXCHANGE.md‘	Macro / CDCGE Exchange Schema	23
‘COSTING.md‘	Education Costing & Unit Costs	18
‘COST_BENEFIT.md‘	Cost-Effectiveness & Cost-Benefit Analysis	20
‘CROSS_SECTOR.md‘	Cross-Sector Linkages (Education ↔ Health / Labour)	24
‘DATA_DICTIONARY.md‘	Data Dictionary — Canonical Schemas	03 (regenerated)
‘DATA_QUALITY.md‘	Diagnostics & Data-Quality Dashboard	31
‘DATA_USAGE_AUDIT.md‘	Data Usage Audit — SDInfrastracture, D:\EDIP\Data, LSMS	
‘DEMOGRAPHY.md‘	Population & Demography — School-Age Cohorts	13
‘DROPOUT_PROGRESSION.md‘	Dropout & Progression Models	08
‘ECONOMETRIC_MODELS.md‘	Econometric Models	07 (attendance & enrolment)
‘ENROLMENT_FORECAST.md‘	Enrolment / Cohort Forecasting	17
‘EQUITY_AND_DEPRIVATION.md‘	Education Equity & Deprivation	10
‘GIS_MANUAL.md‘	GIS / Spatial Foundation	12
‘HOUSEHOLD_ HARMONIZATION.md‘	Household Survey Harmonization	04
‘INDEX.md‘	EDIS Education Planning, Modelling & Decision-Support System — Documentation Index	
‘INDICATOR_ENGINE.md‘	Education Indicator Engine	06
‘LEARNING_MODELS.md‘	Learning Outcome & Education Production Models	11
‘MATHEMATICAL_ SPECIFICATION.md‘	Mathematical Specification	
‘MCDA.md‘	Prioritization — Multi-Criteria Decision Analysis	26

Document	Title	Command
‘MICROSIMULATION.md‘	Household Policy Microsimulation	22
‘MTEF.md‘	MTEF — Medium-Term Expenditure Framework	19
‘MTEF_LINK.md‘	Direct Link — EducationModelling ↔ MTEFModelling (FINEX)	
‘OPTIMIZATION.md‘	Facility-Location Optimization (School Siting)	15
‘PERFORMANCE.md‘	Performance & Benchmarking	35
‘PIPELINE.md‘	End-to-End Pipeline Orchestration	27
‘POST_BASIC.md‘	TVET & Tertiary Sub-Sector Models	21
‘REGRESSION_SUITE.md‘	Consolidated Test Suite / Regression Runner	32
‘REPORTING.md‘	Reporting Layer	30
‘REPOSITORY_AUDIT.md‘	Repository Audit — EDIS Integrated Education System Model	00 — Audit the Existing Repository (no code changed)
‘SCENARIO_ENGINE.md‘	Scenario Engine	25
‘SCHOOL_HARMONIZATION.md‘	EMIS / School Data Harmonization	05
‘SERVICES.md‘	API / Services Layer	28
‘TEACHER_DEMAND.md‘	Teacher Demand & Allocation	16
‘TRAVEL_TIME.md‘	Network Travel-Time Accessibility	
‘WEB_INTEGRATION.md‘	Genie UI Integration	29

Appendix F. Commands

F.1 Rebuilding this course

The full chain

```
cd D:/EDIP/Training/EDIS_Academy/_build
python education_extract.py
python education_analysis.py
python education_figures.py
python build_edu501.py
python render_check.py --json \
"../05_Sector_Planning/Education/EDU_501_Education_Sector_Modelling"
python build_edu501.py # second pass picks up counts
```

F.2 Running a laboratory script

Every laboratory uses this form

```
cd 09_Code/Julia
julia --pkgimages=no --project=D:/EDIP/models/EducationModelling ex01_estate.jl
```

--PKGIMAGES=NO IS NOT OPTIONAL

A machine-level Application Control policy on this host blocks freshly compiled, unsigned Julia package-cache libraries. Without the flag, Julia fails at load with a message about a blocked file. It is the most common first-hour failure in this course.

F.3 Running the estate itself

The estate's own test suite and application

```
cd D:/EDIP/models/EducationModelling
julia --pkgimages=no --project=. test/runtests.jl
julia --pkgimages=no --project=. app.jl
```

F.4 Configuration files worth reading before you run anything

File	What it sets
‘config/cdcge.yml‘	The macro exchange configuration
‘config/defaults.toml‘	Global run defaults
‘config/mappings/household/dhs_pr.toml‘	Survey recode to canonical variables
‘config/mappings/household/lsms_uga_unps.toml‘	Household panel to canonical variables
‘config/models/attendance_logit.toml‘	The attendance model specification
‘config/models/edi_weights.toml‘	Deprivation index weights, which must sum to one
‘config/models/enrolment_logit.toml‘	The enrolment model specification
‘config/models/learning_production.toml‘	The learning production function
‘config/scenario.yml‘	Scenario engine settings
‘config/scenarios/baseline.toml‘	The baseline scenario

Appendix G. The estimate register

Every design-based estimate this course produced: 648 cells across two surveys, five dimensions and eight indicators. The chapters quote these a few rows at a time; this appendix carries them whole, so any claim in the course can be checked without opening a spreadsheet.

READ THE LAST TWO COLUMNS FIRST
deff is the design effect: divide n by it for the effective sample size. **Rel.** is the disclosure decision — ‘NO‘ means the cell is below the threshold and must not be reported, or reconstructed by subtraction from cells that were.

G. Uganda 2016 DHS (UG2016DHS) — 192 cells

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
Attending, ages 6–18	national		83.08 %	82.28–83.88 %	3.86	32,904	yes
Attending, ages 6–18	region	Acholi	79.29 %	75.75–82.83 %	4.44	2,240	yes
Attending, ages 6–18	region	Ankole	86.64 %	84.62–88.66 %	2.09	2,281	yes
Attending, ages 6–18	region	Bugisu	88.74 %	86.70–90.77 %	2.06	1,911	yes
Attending, ages 6–18	region	Bukedi	88.68 %	87.12–90.23 %	1.53	2,437	yes
Attending, ages 6–18	region	Bunyoro	77.40 %	74.13–80.66 %	3.15	1,986	yes
Attending, ages 6–18	region	Busoga	90.52 %	88.98–92.06 %	2.24	3,131	yes
Attending, ages 6–18	region	Kampala	83.58 %	80.01–87.14 %	2.86	1,187	yes
Attending, ages 6–18	region	Karamoja	35.38 %	23.84–46.92 %	25.22	1,664	yes
Attending, ages 6–18	region	Lango	76.98 %	74.18–79.79 %	2.90	2,502	yes
Attending, ages 6–18	region	North Buganda	85.94 %	84.01–87.87 %	1.96	2,443	yes
Attending, ages 6–18	region	South Buganda	84.62 %	82.54–86.71 %	1.86	2,132	yes
Attending, ages 6–18	region	Teso	86.46 %	84.53–88.39 %	2.08	2,511	yes
Attending, ages 6–18	region	Tooro	81.66 %	79.05–84.27 %	2.82	2,377	yes
Attending, ages 6–18	region	West Nile	77.57 %	73.29–81.84 %	6.93	2,531	yes
Attending, ages 6–18	residence	Rural	82.83 %	81.98–83.68 %	3.60	27,243	yes
Attending, ages 6–18	residence	Urban	84.14 %	82.03–86.24 %	4.90	5,661	yes
Attending, ages 6–18	sex	Female	82.67 %	81.72–83.61 %	2.67	16,435	yes
Attending, ages 6–18	sex	Male	83.49 %	82.57–84.42 %	2.66	16,469	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
Attending, ages 6–18	wealth	Fourth	87.51 %	86.29–88.73 %	2.18	6,173	yes
Attending, ages 6–18	wealth	Highest	86.61 %	85.20–88.02 %	2.12	4,719	yes
Attending, ages 6–18	wealth	Lowest	72.22 %	70.15–74.29 %	4.48	8,052	yes
Attending, ages 6–18	wealth	Middle	86.41 %	85.38–87.44 %	1.60	6,814	yes
Attending, ages 6–18	wealth	Second	83.36 %	82.14–84.58 %	1.99	7,146	yes
Gross attendance ratio, primary	national		114.43 %	112.86–116.01 %	—	20,573	yes
Gross attendance ratio, primary	region	Acholi	121.14 %	113.79–128.49 %	—	1,389	yes
Gross attendance ratio, primary	region	Ankole	104.97 %	100.55–109.39 %	—	1,504	yes
Gross attendance ratio, primary	region	Bugisu	126.18 %	119.48–132.88 %	—	1,197	yes
Gross attendance ratio, primary	region	Bukedi	131.32 %	126.59–136.05 %	—	1,508	yes
Gross attendance ratio, primary	region	Bunyoro	107.94 %	104.11–111.76 %	—	1,240	yes
Gross attendance ratio, primary	region	Busoga	120.77 %	116.80–124.74 %	—	1,993	yes
Gross attendance ratio, primary	region	Kampala	103.65 %	98.89–108.41 %	—	695	yes
Gross attendance ratio, primary	region	Karamoja	52.38 %	33.33–71.42 %	43.26	1,143	yes
Gross attendance ratio, primary	region	Lango	119.88 %	113.77–126.00 %	—	1,499	yes
Gross attendance ratio, primary	region	North Buganda	110.79 %	105.40–116.18 %	—	1,507	yes
Gross attendance ratio, primary	region	South Buganda	106.12 %	101.01–111.23 %	—	1,293	yes
Gross attendance ratio, primary	region	Teso	130.30 %	125.74–134.87 %	—	1,533	yes
Gross attendance ratio, primary	region	Tooro	106.76 %	101.85–111.67 %	—	1,492	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
Gross attendance ratio, primary	region	West Nile	119.73 %	112.27–127.19 %	—	1,571	yes
Gross attendance ratio, primary	residence	Rural	115.49 %	113.78–117.21 %	—	17,141	yes
Gross attendance ratio, primary	residence	Urban	109.78 %	105.89–113.68 %	—	3,432	yes
Gross attendance ratio, primary	sex	Female	113.59 %	111.63–115.54 %	—	10,309	yes
Gross attendance ratio, primary	sex	Male	115.29 %	113.32–117.25 %	—	10,264	yes
Gross attendance ratio, primary	wealth	Fourth	116.46 %	113.60–119.32 %	—	3,757	yes
Gross attendance ratio, primary	wealth	Highest	112.39 %	109.11–115.68 %	—	2,818	yes
Gross attendance ratio, primary	wealth	Lowest	106.07 %	102.37–109.78 %	—	5,259	yes
Gross attendance ratio, primary	wealth	Middle	117.62 %	114.63–120.61 %	—	4,269	yes
Gross attendance ratio, primary	wealth	Second	119.39 %	116.17–122.62 %	—	4,470	yes
Net attendance, primary	national		78.81 %	77.88–79.73 %	2.77	20,573	yes
Net attendance, primary	region	Acholi	78.44 %	74.42–82.45 %	3.45	1,389	yes
Net attendance, primary	region	Ankole	74.59 %	71.28–77.89 %	2.25	1,504	yes
Net attendance, primary	region	Bugisu	83.25 %	79.82–86.67 %	2.62	1,197	yes
Net attendance, primary	region	Bukedi	84.75 %	82.59–86.92 %	1.43	1,508	yes
Net attendance, primary	region	Bunyoro	76.41 %	73.25–79.56 %	1.78	1,240	yes
Net attendance, primary	region	Busoga	84.59 %	82.80–86.38 %	1.27	1,993	yes
Net attendance, primary	region	Kampala	87.18 %	84.88–89.49 %	0.86	695	yes
Net attendance, primary	region	Karamoja	33.37 %	22.26–44.49 %	16.54	1,143	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
Net attendance, primary	region	Lango	74.97 %	72.32–77.62 %	1.46	1,499	yes
Net attendance, primary	region	North Buganda	78.97 %	76.38–81.56 %	1.58	1,507	yes
Net attendance, primary	region	South Buganda	79.86 %	76.29–83.43 %	2.67	1,293	yes
Net attendance, primary	region	Teso	86.08 %	83.84–88.31 %	1.66	1,533	yes
Net attendance, primary	region	Tooro	75.98 %	72.93–79.03 %	1.98	1,492	yes
Net attendance, primary	region	West Nile	75.79 %	71.11–80.46 %	4.87	1,571	yes
Net attendance, primary	residence	Rural	77.94 %	76.93–78.94 %	2.63	17,141	yes
Net attendance, primary	residence	Urban	82.62 %	80.26–84.99 %	3.48	3,432	yes
Net attendance, primary	sex	Female	79.45 %	78.33–80.57 %	2.07	10,309	yes
Net attendance, primary	sex	Male	78.16 %	77.03–79.29 %	1.99	10,264	yes
Net attendance, primary	wealth	Fourth	80.76 %	79.21–82.31 %	1.51	3,757	yes
Net attendance, primary	wealth	Highest	87.25 %	85.84–88.66 %	1.31	2,818	yes
Net attendance, primary	wealth	Lowest	70.38 %	68.13–72.62 %	3.31	5,259	yes
Net attendance, primary	wealth	Middle	80.04 %	78.63–81.46 %	1.39	4,269	yes
Net attendance, primary	wealth	Second	78.05 %	76.41–79.68 %	1.81	4,470	yes
Net attendance, secondary	national		17.81 %	16.56–19.05 %	3.41	12,331	yes
Net attendance, secondary	region	Acholi	9.13 %	6.08–12.17 %	2.48	851	yes
Net attendance, secondary	region	Ankole	14.69 %	11.74–17.64 %	1.40	777	yes
Net attendance, secondary	region	Bugisu	18.25 %	13.79–22.71 %	2.48	714	yes
Net attendance, secondary	region	Bukedi	15.63 %	12.79–18.46 %	1.48	929	yes
Net attendance, secondary	region	Bunyoro	13.44 %	10.03–16.85 %	1.94	746	yes
Net attendance, secondary	region	Busoga	23.19 %	17.83–28.55 %	4.78	1,138	yes
Net attendance, secondary	region	Kampala	41.69 %	36.19–47.18 %	1.59	492	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
Net attendance, secondary	region	Karamoja	2.58 %	1.09–4.07 %	1.20	521	yes
Net attendance, secondary	region	Lango	3.44 %	2.13–4.76 %	1.36	1,003	yes
Net attendance, secondary	region	North Buganda	24.07 %	19.31–28.83 %	3.02	936	yes
Net attendance, secondary	region	South Buganda	29.21 %	23.55–34.88 %	3.39	839	yes
Net attendance, secondary	region	Teso	15.66 %	9.75–21.56 %	6.72	978	yes
Net attendance, secondary	region	Tooro	18.69 %	14.63–22.74 %	2.49	885	yes
Net attendance, secondary	region	West Nile	7.36 %	5.53–9.19 %	1.23	960	yes
Net attendance, secondary	residence	Rural	14.35 %	13.00–15.69 %	3.86	10,102	yes
Net attendance, secondary	residence	Urban	31.74 %	28.93–34.55 %	2.11	2,229	yes
Net attendance, secondary	sex	Female	18.05 %	16.45–19.64 %	2.73	6,126	yes
Net attendance, secondary	sex	Male	17.57 %	16.13–19.01 %	2.31	6,205	yes
Net attendance, secondary	wealth	Fourth	24.15 %	21.67–26.64 %	2.11	2,416	yes
Net attendance, secondary	wealth	Highest	35.65 %	32.72–38.58 %	1.85	1,901	yes
Net attendance, secondary	wealth	Lowest	5.20 %	4.13–6.26 %	1.69	2,793	yes
Net attendance, secondary	wealth	Middle	16.52 %	14.51–18.53 %	1.94	2,545	yes
Net attendance, secondary	wealth	Second	9.02 %	7.66–10.38 %	1.56	2,676	yes
No education, 15+	national		13.31 %	12.62–14.01 %	4.96	45,775	yes
No education, 15+	region	Acholi	12.66 %	10.58–14.74 %	2.71	2,655	yes
No education, 15+	region	Ankole	14.34 %	12.40–16.28 %	2.52	3,162	yes
No education, 15+	region	Bugisu	10.02 %	7.91–12.13 %	3.26	2,537	yes
No education, 15+	region	Bukedi	12.15 %	10.63–13.66 %	1.69	3,021	yes
No education, 15+	region	Bunyoro	14.67 %	11.82–17.51 %	4.74	2,819	yes
No education, 15+	region	Busoga	13.33 %	11.28–15.38 %	3.59	3,798	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
No education, 15+	region	Kampala	4.22 %	2.91–5.53 %	3.22	2,910	yes
No education, 15+	region	Karamoja	65.50 %	54.65–76.34 %	28.11	2,074	yes
No education, 15+	region	Lango	13.13 %	11.32–14.94 %	2.35	3,161	yes
No education, 15+	region	North Buganda	11.34 %	8.96–13.72 %	5.21	3,551	yes
No education, 15+	region	South Buganda	8.16 %	6.19–10.13 %	5.31	3,943	yes
No education, 15+	region	Teso	10.55 %	8.91–12.18 %	2.37	3,233	yes
No education, 15+	region	Tooro	16.19 %	13.69–18.68 %	3.90	3,272	yes
No education, 15+	region	West Nile	14.42 %	11.83–17.01 %	4.42	3,124	yes
No education, 15+	residence	Rural	15.19 %	14.43–15.94 %	4.14	35,612	yes
No education, 15+	residence	Urban	7.56 %	6.03–9.09 %	8.87	10,163	yes
No education, 15+	sex	Female	18.21 %	17.34–19.08 %	3.20	24,227	yes
No education, 15+	sex	Male	7.78 %	7.08–8.48 %	3.82	21,548	yes
No education, 15+	wealth	Fourth	9.34 %	8.53–10.15 %	1.74	8,584	yes
No education, 15+	wealth	Highest	4.83 %	3.90–5.76 %	4.57	9,389	yes
No education, 15+	wealth	Lowest	25.10 %	23.21–26.98 %	4.91	9,942	yes
No education, 15+	wealth	Middle	12.72 %	11.73–13.72 %	2.02	8,694	yes
No education, 15+	wealth	Second	17.17 %	16.20–18.15 %	1.60	9,166	yes
Out of school, primary age	national		10.49 %	9.64–11.34 %	4.08	20,573	yes
Out of school, primary age	region	Acholi	18.18 %	14.14–22.23 %	3.98	1,389	yes
Out of school, primary age	region	Ankole	5.77 %	3.53–8.01 %	3.62	1,504	yes
Out of school, primary age	region	Bugisu	8.31 %	6.53–10.08 %	1.28	1,197	yes
Out of school, primary age	region	Bukedi	7.95 %	6.20–9.71 %	1.65	1,508	yes
Out of school, primary age	region	Bunyoro	13.41 %	9.32–17.49 %	4.65	1,240	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
Out of school, primary age	region	Busoga	5.16 %	3.76–6.57 %	2.09	1,993	yes
Out of school, primary age	region	Kampala	3.22 %	1.63–4.81 %	1.47	695	yes
Out of school, primary age	region	Karamoja	64.96 %	54.12–75.79 %	15.35	1,143	yes
Out of school, primary age	region	Lango	17.63 %	14.79–20.46 %	2.17	1,499	yes
Out of school, primary age	region	North Buganda	3.53 %	2.15–4.90 %	2.19	1,507	yes
Out of school, primary age	region	South Buganda	4.46 %	2.48–6.44 %	3.09	1,293	yes
Out of school, primary age	region	Teso	8.59 %	6.45–10.73 %	2.32	1,533	yes
Out of school, primary age	region	Tooro	9.98 %	7.23–12.72 %	3.25	1,492	yes
Out of school, primary age	region	West Nile	19.59 %	14.84–24.34 %	5.85	1,571	yes
Out of school, primary age	residence	Rural	11.27 %	10.38–12.16 %	3.56	17,141	yes
Out of school, primary age	residence	Urban	7.06 %	4.71–9.41 %	7.53	3,432	yes
Out of school, primary age	sex	Female	10.28 %	9.30–11.27 %	2.82	10,309	yes
Out of school, primary age	sex	Male	10.70 %	9.74–11.66 %	2.58	10,264	yes
Out of school, primary age	wealth	Fourth	4.98 %	4.00–5.96 %	1.99	3,757	yes
Out of school, primary age	wealth	Highest	2.61 %	1.90–3.33 %	1.47	2,818	yes
Out of school, primary age	wealth	Lowest	24.41 %	22.15–26.67 %	3.78	5,259	yes
Out of school, primary age	wealth	Middle	6.91 %	5.99–7.83 %	1.45	4,269	yes
Out of school, primary age	wealth	Second	10.95 %	9.71–12.18 %	1.82	4,470	yes
Over-age for grade	national		62.01 %	60.92–63.09 %	3.03	23,295	yes
Over-age for grade	region	Acholi	65.90 %	63.01–68.79 %	1.61	1,670	yes
Over-age for grade	region	Ankole	66.60 %	63.07–70.13 %	2.33	1,592	yes
Over-age for grade	region	Bugisu	61.73 %	56.96–66.50 %	3.78	1,510	yes
Over-age for grade	region	Bukedi	64.94 %	61.77–68.10 %	2.26	1,974	yes
Over-age for grade	region	Bunyoro	63.56 %	57.09–70.02 %	6.30	1,340	yes
Over-age for grade	region	Busoga	57.20 %	54.09–60.32 %	2.48	2,408	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
Over-age for grade	region	Kampala	27.37 %	23.71–31.02 %	1.25	715	yes
Over-age for grade	region	Karamoja	70.11 %	67.41–72.81 %	0.52	577	yes
Over-age for grade	region	Lango	75.57 %	73.38–77.75 %	1.22	1,812	yes
Over-age for grade	region	North Buganda	58.80 %	54.87–62.72 %	2.74	1,650	yes
Over-age for grade	region	South Buganda	48.25 %	42.55–53.96 %	4.69	1,383	yes
Over-age for grade	region	Teso	63.24 %	59.96–66.52 %	2.41	2,002	yes
Over-age for grade	region	Tooro	65.18 %	61.22–69.13 %	2.86	1,590	yes
Over-age for grade	region	West Nile	74.68 %	72.00–77.36 %	1.84	1,855	yes
Over-age for grade	residence	Rural	65.58 %	64.41–66.76 %	3.09	19,526	yes
Over-age for grade	residence	Urban	45.47 %	42.58–48.36 %	3.30	3,769	yes
Over-age for grade	sex	Female	58.62 %	57.35–59.90 %	2.01	11,528	yes
Over-age for grade	sex	Male	65.36 %	64.04–66.68 %	2.36	11,767	yes
Over-age for grade	wealth	Fourth	58.56 %	56.60–60.51 %	1.82	4,465	yes
Over-age for grade	wealth	Highest	38.53 %	35.94–41.12 %	2.34	3,176	yes
Over-age for grade	wealth	Lowest	70.88 %	69.21–72.55 %	1.85	5,252	yes
Over-age for grade	wealth	Middle	65.94 %	64.00–67.88 %	2.20	5,030	yes
Over-age for grade	wealth	Second	69.71 %	67.91–71.50 %	2.12	5,372	yes
Years of schooling, 20–24	national		7.954	7.80–8.11	—	7,271	yes
Years of schooling, 20–24	region	Acholi	8.170	7.63–8.71	—	423	yes
Years of schooling, 20–24	region	Ankole	7.531	6.89–8.17	—	444	yes
Years of schooling, 20–24	region	Bugisu	8.164	7.62–8.71	—	368	yes
Years of schooling, 20–24	region	Bukedi	7.182	6.77–7.60	—	490	yes
Years of schooling, 20–24	region	Bunyoro	6.547	6.01–7.08	—	466	yes
Years of schooling, 20–24	region	Busoga	8.073	7.52–8.63	—	589	yes
Years of schooling, 20–24	region	Kampala	10.770	10.30–11.24	—	661	yes
Years of schooling, 20–24	region	Karamoja	3.332	2.05–4.61	—	231	yes
Years of schooling, 20–24	region	Lango	7.039	6.75–7.33	—	482	yes
Years of schooling, 20–24	region	North Buganda	7.932	7.41–8.45	—	549	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
Years of schooling, 20–24	region	South Buganda	9.768	9.25–10.29	—	677	yes
Years of schooling, 20–24	region	Teso	7.423	7.11–7.73	—	539	yes
Years of schooling, 20–24	region	Tooro	7.090	6.66–7.52	—	543	yes
Years of schooling, 20–24	region	West Nile	6.708	5.99–7.42	—	453	yes
Years of schooling, 20–24	residence	Rural	7.190	7.00–7.38	—	5,357	yes
Years of schooling, 20–24	residence	Urban	9.870	9.61–10.13	—	1,914	yes
Years of schooling, 20–24	sex	Female	7.812	7.64–7.99	—	4,041	yes
Years of schooling, 20–24	sex	Male	8.131	7.93–8.33	—	3,230	yes
Years of schooling, 20–24	wealth	Fourth	8.388	8.16–8.61	—	1,326	yes
Years of schooling, 20–24	wealth	Highest	10.578	10.32–10.84	—	1,798	yes
Years of schooling, 20–24	wealth	Lowest	5.623	5.40–5.85	—	1,411	yes
Years of schooling, 20–24	wealth	Middle	7.121	6.90–7.34	—	1,306	yes
Years of schooling, 20–24	wealth	Second	6.459	6.27–6.65	—	1,430	yes

G. Kenya 2022 DHS (KE2022DHS) — 456 cells

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
Attending, ages 6–18	national		93.20 %	92.45–93.96 %	13.03	55,506	yes
Attending, ages 6–18	region	1	94.86 %	93.53–96.20 %	0.71	752	yes
Attending, ages 6–18	region	10	76.63 %	67.71–85.55 %	15.07	1,303	yes
Attending, ages 6–18	region	11	86.41 %	81.28–91.55 %	8.09	1,384	yes
Attending, ages 6–18	region	12	93.95 %	92.02–95.88 %	1.47	865	yes
Attending, ages 6–18	region	13	94.37 %	91.65–97.10 %	2.91	797	yes
Attending, ages 6–18	region	14	96.58 %	94.67–98.49 %	2.05	714	yes
Attending, ages 6–18	region	15	96.58 %	95.35–97.81 %	1.27	1,067	yes
Attending, ages 6–18	region	16	98.00 %	96.78–99.22 %	1.73	873	yes
Attending, ages 6–18	region	17	96.80 %	95.81–97.78 %	0.74	904	yes
Attending, ages 6–18	region	18	97.49 %	96.42–98.56 %	1.01	829	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
Attending, ages 6–18	region	19	95.68 %	92.29–99.07 %	4.07	564	yes
Attending, ages 6–18	region	2	90.37 %	86.95–93.80 %	4.32	1,232	yes
Attending, ages 6–18	region	20	96.96 %	95.26–98.67 %	1.67	650	yes
Attending, ages 6–18	region	21	97.81 %	96.49–99.12 %	1.59	757	yes
Attending, ages 6–18	region	22	97.44 %	95.59–99.29 %	2.61	731	yes
Attending, ages 6–18	region	23	57.64 %	47.17–68.12 %	16.51	1,412	yes
Attending, ages 6–18	region	24	77.94 %	63.34–92.55 %	54.85	1,698	yes
Attending, ages 6–18	region	25	65.23 %	55.34–75.12 %	15.73	1,402	yes
Attending, ages 6–18	region	26	96.68 %	95.25–98.10 %	2.21	1,349	yes
Attending, ages 6–18	region	27	96.96 %	95.61–98.32 %	1.74	1,074	yes
Attending, ages 6–18	region	28	98.10 %	97.01–99.19 %	1.91	1,154	yes
Attending, ages 6–18	region	29	96.62 %	95.36–97.88 %	1.44	1,140	yes
Attending, ages 6–18	region	3	94.90 %	93.20–96.60 %	1.93	1,239	yes
Attending, ages 6–18	region	30	92.66 %	85.77–99.56 %	23.96	1,316	yes
Attending, ages 6–18	region	31	97.01 %	95.96–98.07 %	0.87	872	yes
Attending, ages 6–18	region	32	96.87 %	95.91–97.84 %	0.84	1,043	yes
Attending, ages 6–18	region	33	92.15 %	89.71–94.60 %	3.24	1,505	yes
Attending, ages 6–18	region	34	90.04 %	85.72–94.36 %	5.22	962	yes
Attending, ages 6–18	region	35	96.39 %	95.23–97.55 %	1.16	1,158	yes
Attending, ages 6–18	region	36	98.49 %	97.65–99.33 %	1.53	1,232	yes
Attending, ages 6–18	region	37	96.99 %	96.14–97.83 %	0.94	1,486	yes
Attending, ages 6–18	region	38	97.20 %	96.11–98.29 %	1.50	1,322	yes
Attending, ages 6–18	region	39	97.16 %	96.28–98.05 %	1.09	1,488	yes
Attending, ages 6–18	region	4	69.16 %	60.68–77.64 %	12.48	1,423	yes
Attending, ages 6–18	region	40	96.66 %	95.56–97.76 %	1.43	1,479	yes
Attending, ages 6–18	region	41	96.35 %	95.18–97.52 %	1.23	1,212	yes
Attending, ages 6–18	region	42	97.19 %	96.16–98.23 %	1.22	1,189	yes
Attending, ages 6–18	region	43	96.87 %	95.92–97.82 %	1.15	1,483	yes
Attending, ages 6–18	region	44	95.76 %	94.92–96.60 %	0.71	1,583	yes
Attending, ages 6–18	region	45	96.82 %	95.65–98.00 %	1.48	1,269	yes
Attending, ages 6–18	region	46	97.23 %	96.11–98.35 %	1.48	1,217	yes
Attending, ages 6–18	region	47	93.35 %	85.52–100.00 %	21.52	838	yes
Attending, ages 6–18	region	5	91.55 %	88.86–94.24 %	2.80	1,151	yes
Attending, ages 6–18	region	6	96.98 %	95.83–98.13 %	0.90	766	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
Attending, ages 6–18	region	7	54.29 %	47.97–60.61 %	6.80	1,621	yes
Attending, ages 6–18	region	8	68.54 %	60.05–77.03 %	17.49	2,010	yes
Attending, ages 6–18	region	9	61.34 %	58.27–64.42 %	2.06	1,991	yes
Attending, ages 6–18	residence	Rural	92.98 %	92.33–93.62 %	6.58	39,999	yes
Attending, ages 6–18	residence	Urban	93.82 %	91.61–96.03 %	34.00	15,507	yes
Attending, ages 6–18	sex	Female	93.55 %	92.81–94.29 %	6.42	27,435	yes
Attending, ages 6–18	sex	Male	92.86 %	92.01–93.70 %	7.87	28,071	yes
Attending, ages 6–18	wealth	Fourth	96.14 %	95.50–96.79 %	2.68	9,236	yes
Attending, ages 6–18	wealth	Highest	95.63 %	91.80–99.46 %	52.78	5,774	yes
Attending, ages 6–18	wealth	Lowest	84.41 %	82.66–86.15 %	10.68	17,732	yes
Attending, ages 6–18	wealth	Middle	96.16 %	95.65–96.66 %	1.97	11,125	yes
Attending, ages 6–18	wealth	Second	95.85 %	95.40–96.30 %	1.55	11,639	yes
Gross attendance ratio, primary	national		123.76 %	122.08–125.43 %	—	31,035	yes
Gross attendance ratio, primary	region	1	120.03 %	112.07–127.99 %	—	430	yes
Gross attendance ratio, primary	region	10	96.71 %	79.90–113.52 %	182.27	788	yes
Gross attendance ratio, primary	region	11	104.97 %	96.17–113.77 %	—	853	yes
Gross attendance ratio, primary	region	12	123.42 %	112.50–134.34 %	—	447	yes
Gross attendance ratio, primary	region	13	134.95 %	122.02–147.87 %	—	419	yes
Gross attendance ratio, primary	region	14	128.63 %	118.56–138.70 %	—	391	yes
Gross attendance ratio, primary	region	15	138.67 %	132.53–144.81 %	—	567	yes
Gross attendance ratio, primary	region	16	136.21 %	124.25–148.17 %	—	432	yes
Gross attendance ratio, primary	region	17	144.23 %	136.26–152.21 %	—	436	yes
Gross attendance ratio, primary	region	18	129.18 %	118.76–139.60 %	—	426	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
Gross attendance ratio, primary	region	19	124.10 %	115.05–133.15 %	—	297	yes
Gross attendance ratio, primary	region	2	127.93 %	117.16–138.71 %	—	683	yes
Gross attendance ratio, primary	region	20	118.23 %	108.72–127.74 %	—	351	yes
Gross attendance ratio, primary	region	21	126.51 %	115.89–137.13 %	—	390	yes
Gross attendance ratio, primary	region	22	114.36 %	107.52–121.21 %	—	419	yes
Gross attendance ratio, primary	region	23	66.75 %	51.56–81.95 %	24.39	901	yes
Gross attendance ratio, primary	region	24	90.84 %	68.93–112.74 %	156.29	1,042	yes
Gross attendance ratio, primary	region	25	79.25 %	66.49–92.01 %	21.83	847	yes
Gross attendance ratio, primary	region	26	125.35 %	118.55–132.14 %	—	767	yes
Gross attendance ratio, primary	region	27	130.79 %	123.33–138.25 %	—	549	yes
Gross attendance ratio, primary	region	28	131.07 %	123.62–138.52 %	—	644	yes
Gross attendance ratio, primary	region	29	134.59 %	123.50–145.68 %	—	610	yes
Gross attendance ratio, primary	region	3	139.11 %	130.60–147.61 %	—	714	yes
Gross attendance ratio, primary	region	30	120.11 %	108.70–131.52 %	—	758	yes
Gross attendance ratio, primary	region	31	122.25 %	114.81–129.69 %	—	463	yes
Gross attendance ratio, primary	region	32	127.59 %	119.93–135.24 %	—	570	yes
Gross attendance ratio, primary	region	33	124.37 %	115.95–132.79 %	—	853	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
Gross attendance ratio, primary	region	34	106.02 %	98.77–113.27 %	—	600	yes
Gross attendance ratio, primary	region	35	128.11 %	122.73–133.49 %	—	628	yes
Gross attendance ratio, primary	region	36	139.72 %	131.02–148.42 %	—	647	yes
Gross attendance ratio, primary	region	37	141.41 %	134.50–148.33 %	—	755	yes
Gross attendance ratio, primary	region	38	148.16 %	138.36–157.97 %	—	656	yes
Gross attendance ratio, primary	region	39	132.00 %	125.75–138.25 %	—	804	yes
Gross attendance ratio, primary	region	4	82.80 %	69.27–96.32 %	29.62	886	yes
Gross attendance ratio, primary	region	40	136.19 %	128.70–143.67 %	—	842	yes
Gross attendance ratio, primary	region	41	135.09 %	128.38–141.79 %	—	652	yes
Gross attendance ratio, primary	region	42	128.30 %	122.32–134.29 %	—	653	yes
Gross attendance ratio, primary	region	43	126.21 %	118.76–133.65 %	—	782	yes
Gross attendance ratio, primary	region	44	127.50 %	120.10–134.91 %	—	877	yes
Gross attendance ratio, primary	region	45	136.94 %	129.99–143.90 %	—	660	yes
Gross attendance ratio, primary	region	46	141.52 %	134.46–148.59 %	—	618	yes
Gross attendance ratio, primary	region	47	119.19 %	107.70–130.67 %	—	478	yes
Gross attendance ratio, primary	region	5	122.17 %	113.90–130.44 %	—	637	yes
Gross attendance ratio, primary	region	6	117.35 %	108.49–126.20 %	—	447	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
Gross attendance ratio, primary	region	7	70.75 %	61.01–80.49 %	11.72	982	yes
Gross attendance ratio, primary	region	8	83.18 %	66.82–99.54 %	56.88	1,142	yes
Gross attendance ratio, primary	region	9	79.31 %	72.63–85.99 %	8.79	1,242	yes
Gross attendance ratio, primary	residence	Rural	126.23 %	124.39–128.07 %	—	22,314	yes
Gross attendance ratio, primary	residence	Urban	117.34 %	113.71–120.98 %	—	8,721	yes
Gross attendance ratio, primary	sex	Female	123.13 %	120.92–125.34 %	—	15,306	yes
Gross attendance ratio, primary	sex	Male	124.38 %	122.17–126.59 %	—	15,729	yes
Gross attendance ratio, primary	wealth	Fourth	122.76 %	119.59–125.93 %	—	4,984	yes
Gross attendance ratio, primary	wealth	Highest	115.17 %	109.60–120.75 %	—	3,273	yes
Gross attendance ratio, primary	wealth	Lowest	113.08 %	109.29–116.86 %	—	10,747	yes
Gross attendance ratio, primary	wealth	Middle	132.85 %	130.24–135.47 %	—	5,812	yes
Gross attendance ratio, primary	wealth	Second	135.08 %	132.24–137.91 %	—	6,219	yes
Net attendance, primary	national		78.16 %	77.25–79.08 %	3.98	31,035	yes
Net attendance, primary	region	1	77.65 %	73.73–81.56 %	0.99	430	yes
Net attendance, primary	region	10	65.68 %	55.56–75.80 %	9.32	788	yes
Net attendance, primary	region	11	71.31 %	65.61–77.02 %	3.53	853	yes
Net attendance, primary	region	12	77.09 %	73.22–80.96 %	0.99	447	yes
Net attendance, primary	region	13	82.31 %	77.89–86.73 %	1.46	419	yes
Net attendance, primary	region	14	87.13 %	83.19–91.07 %	1.41	391	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
Net attendance, primary	region	15	85.30 %	82.69–87.91 %	0.80	567	yes
Net attendance, primary	region	16	90.68 %	87.47–93.89 %	1.37	432	yes
Net attendance, primary	region	17	86.44 %	82.71–90.18 %	1.35	436	yes
Net attendance, primary	region	18	84.57 %	81.31–87.83 %	0.90	426	yes
Net attendance, primary	region	19	85.41 %	81.01–89.81 %	1.20	297	yes
Net attendance, primary	region	2	68.33 %	63.00–73.65 %	2.33	683	yes
Net attendance, primary	region	20	83.85 %	78.49–89.20 %	1.93	351	yes
Net attendance, primary	region	21	83.34 %	78.88–87.80 %	1.46	390	yes
Net attendance, primary	region	22	87.06 %	83.53–90.59 %	1.21	419	yes
Net attendance, primary	region	23	40.01 %	30.84–49.17 %	8.21	901	yes
Net attendance, primary	region	24	57.02 %	43.96–70.08 %	18.89	1,042	yes
Net attendance, primary	region	25	51.22 %	42.55–59.88 %	6.63	847	yes
Net attendance, primary	region	26	81.63 %	77.88–85.37 %	1.87	767	yes
Net attendance, primary	region	27	79.95 %	76.81–83.09 %	0.88	549	yes
Net attendance, primary	region	28	84.89 %	80.71–89.07 %	2.28	644	yes
Net attendance, primary	region	29	82.32 %	79.51–85.13 %	0.86	610	yes
Net attendance, primary	region	3	79.29 %	75.69–82.89 %	1.46	714	yes
Net attendance, primary	region	30	76.73 %	66.95–86.51 %	10.57	758	yes
Net attendance, primary	region	31	82.62 %	76.74–88.50 %	2.90	463	yes
Net attendance, primary	region	32	80.83 %	77.58–84.08 %	1.01	570	yes
Net attendance, primary	region	33	69.94 %	64.64–75.25 %	2.97	853	yes
Net attendance, primary	region	34	73.36 %	67.63–79.08 %	2.62	600	yes
Net attendance, primary	region	35	85.61 %	82.47–88.76 %	1.31	628	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
Net attendance, primary	region	36	84.39 %	80.81–87.97 %	1.64	647	yes
Net attendance, primary	region	37	82.42 %	79.34–85.49 %	1.28	755	yes
Net attendance, primary	region	38	85.11 %	81.87–88.35 %	1.41	656	yes
Net attendance, primary	region	39	80.61 %	77.71–83.50 %	1.12	804	yes
Net attendance, primary	region	4	47.08 %	39.18–54.97 %	5.77	886	yes
Net attendance, primary	region	40	83.62 %	80.48–86.76 %	1.58	842	yes
Net attendance, primary	region	41	86.87 %	83.37–90.37 %	1.83	652	yes
Net attendance, primary	region	42	84.16 %	81.81–86.51 %	0.70	653	yes
Net attendance, primary	region	43	76.44 %	72.66–80.21 %	1.61	782	yes
Net attendance, primary	region	44	75.98 %	72.63–79.32 %	1.40	877	yes
Net attendance, primary	region	45	84.49 %	81.50–87.48 %	1.17	660	yes
Net attendance, primary	region	46	85.03 %	82.07–87.98 %	1.10	618	yes
Net attendance, primary	region	47	83.52 %	76.69–90.35 %	4.22	478	yes
Net attendance, primary	region	5	75.38 %	69.93–80.83 %	2.65	637	yes
Net attendance, primary	region	6	81.47 %	78.09–84.86 %	0.88	447	yes
Net attendance, primary	region	7	37.40 %	31.78–43.02 %	3.45	982	yes
Net attendance, primary	region	8	46.25 %	37.06–55.44 %	10.10	1,142	yes
Net attendance, primary	region	9	45.38 %	40.34–50.42 %	3.31	1,242	yes
Net attendance, primary	residence	Rural	77.31 %	76.35–78.28 %	3.06	22,314	yes
Net attendance, primary	residence	Urban	80.37 %	78.23–82.51 %	6.60	8,721	yes
Net attendance, primary	sex	Female	79.46 %	78.47–80.46 %	2.41	15,306	yes
Net attendance, primary	sex	Male	76.86 %	75.65–78.08 %	3.39	15,729	yes
Net attendance, primary	wealth	Fourth	81.05 %	79.52–82.58 %	1.98	4,984	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
Net attendance, primary	wealth	Highest	84.41 %	81.04–87.78 %	7.36	3,273	yes
Net attendance, primary	wealth	Lowest	65.81 %	63.76–67.86 %	5.22	10,747	yes
Net attendance, primary	wealth	Middle	83.61 %	82.46–84.76 %	1.47	5,812	yes
Net attendance, primary	wealth	Second	80.86 %	79.63–82.08 %	1.57	6,219	yes
Net attendance, secondary	national		38.72 %	37.79–39.64 %	2.31	24,471	yes
Net attendance, secondary	region	1	37.14 %	32.63–41.65 %	0.73	322	yes
Net attendance, secondary	region	10	28.52 %	20.28–36.76 %	4.46	515	yes
Net attendance, secondary	region	11	34.54 %	30.77–38.30 %	0.87	531	yes
Net attendance, secondary	region	12	40.75 %	33.12–48.37 %	2.62	418	yes
Net attendance, secondary	region	13	39.39 %	32.25–46.54 %	2.10	378	yes
Net attendance, secondary	region	14	47.26 %	40.67–53.85 %	1.46	323	yes
Net attendance, secondary	region	15	35.69 %	30.70–40.69 %	1.41	500	yes
Net attendance, secondary	region	16	54.86 %	48.43–61.28 %	1.92	441	yes
Net attendance, secondary	region	17	43.62 %	38.25–48.99 %	1.43	468	yes
Net attendance, secondary	region	18	51.94 %	44.68–59.19 %	2.21	403	yes
Net attendance, secondary	region	19	55.49 %	49.61–61.37 %	0.97	267	yes
Net attendance, secondary	region	2	17.72 %	13.49–21.95 %	1.76	549	yes
Net attendance, secondary	region	20	55.62 %	48.28–62.96 %	1.70	299	yes
Net attendance, secondary	region	21	51.34 %	44.04–58.64 %	2.04	367	yes
Net attendance, secondary	region	22	57.81 %	52.17–63.46 %	1.06	312	yes
Net attendance, secondary	region	23	15.61 %	11.37–19.85 %	1.82	511	yes
Net attendance, secondary	region	24	27.88 %	21.32–34.44 %	3.65	656	yes
Net attendance, secondary	region	25	23.67 %	17.09–30.24 %	3.46	555	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
Net attendance, secondary	region	26	39.89 %	35.93–43.85 %	0.99	582	yes
Net attendance, secondary	region	27	43.15 %	38.39–47.91 %	1.26	525	yes
Net attendance, secondary	region	28	43.34 %	37.95–48.74 %	1.57	510	yes
Net attendance, secondary	region	29	34.63 %	30.41–38.84 %	1.08	530	yes
Net attendance, secondary	region	3	16.40 %	12.77–20.02 %	1.31	525	yes
Net attendance, secondary	region	30	36.04 %	29.44–42.65 %	2.75	558	yes
Net attendance, secondary	region	31	51.00 %	42.73–59.28 %	2.92	409	yes
Net attendance, secondary	region	32	41.34 %	37.07–45.61 %	0.93	473	yes
Net attendance, secondary	region	33	21.51 %	16.66–26.36 %	2.37	652	yes
Net attendance, secondary	region	34	42.63 %	34.96–50.29 %	2.26	362	yes
Net attendance, secondary	region	35	47.20 %	42.34–52.05 %	1.31	530	yes
Net attendance, secondary	region	36	38.44 %	33.34–43.53 %	1.67	585	yes
Net attendance, secondary	region	37	35.51 %	31.25–39.77 %	1.51	731	yes
Net attendance, secondary	region	38	36.25 %	32.14–40.37 %	1.27	666	yes
Net attendance, secondary	region	39	36.98 %	32.49–41.47 %	1.54	684	yes
Net attendance, secondary	region	4	12.78 %	8.49–17.07 %	2.30	537	yes
Net attendance, secondary	region	40	31.90 %	28.03–35.78 %	1.15	637	yes
Net attendance, secondary	region	41	40.25 %	36.32–44.18 %	0.94	560	yes
Net attendance, secondary	region	42	42.52 %	37.86–47.17 %	1.24	536	yes
Net attendance, secondary	region	43	40.13 %	35.70–44.56 %	1.49	701	yes
Net attendance, secondary	region	44	30.78 %	26.42–35.14 %	1.64	706	yes
Net attendance, secondary	region	45	39.10 %	35.00–43.20 %	1.12	609	yes
Net attendance, secondary	region	46	40.16 %	36.80–43.52 %	0.73	599	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
Net attendance, secondary	region	47	45.35 %	39.39–51.30 %	1.34	360	yes
Net attendance, secondary	region	5	30.16 %	24.26–36.06 %	2.21	514	yes
Net attendance, secondary	region	6	47.22 %	41.40–53.04 %	1.13	319	yes
Net attendance, secondary	region	7	16.67 %	13.90–19.44 %	0.92	639	yes
Net attendance, secondary	region	8	28.77 %	24.04–33.50 %	2.47	868	yes
Net attendance, secondary	region	9	19.23 %	15.63–22.82 %	1.62	749	yes
Net attendance, secondary	residence	Rural	36.41 %	35.41–37.41 %	1.98	17,685	yes
Net attendance, secondary	residence	Urban	45.33 %	43.22–47.44 %	3.16	6,786	yes
Net attendance, secondary	sex	Female	40.66 %	39.46–41.86 %	1.89	12,129	yes
Net attendance, secondary	sex	Male	36.79 %	35.53–38.05 %	2.20	12,342	yes
Net attendance, secondary	wealth	Fourth	48.07 %	46.04–50.10 %	1.82	4,252	yes
Net attendance, secondary	wealth	Highest	52.36 %	48.87–55.86 %	3.19	2,501	yes
Net attendance, secondary	wealth	Lowest	20.72 %	19.25–22.18 %	2.37	6,985	yes
Net attendance, secondary	wealth	Middle	42.64 %	41.02–44.26 %	1.48	5,313	yes
Net attendance, secondary	wealth	Second	36.43 %	34.90–37.96 %	1.42	5,420	yes
No education, 15+	national		8.34 %	7.92–8.76 %	5.44	90,454	yes
No education, 15+	region	1	7.03 %	5.76–8.29 %	1.19	1,874	yes
No education, 15+	region	10	55.10 %	47.57–62.64 %	9.52	1,593	yes
No education, 15+	region	11	30.25 %	25.78–34.72 %	4.40	1,781	yes
No education, 15+	region	12	8.51 %	6.56–10.46 %	2.41	1,895	yes
No education, 15+	region	13	4.95 %	3.65–6.24 %	1.63	1,760	yes
No education, 15+	region	14	5.00 %	3.48–6.51 %	2.12	1,693	yes
No education, 15+	region	15	10.25 %	7.93–12.57 %	2.90	1,897	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
No education, 15+	region	16	2.93 %	2.18–3.68 %	1.04	2,032	yes
No education, 15+	region	17	5.53 %	4.33–6.74 %	1.45	2,007	yes
No education, 15+	region	18	4.46 %	3.25–5.66 %	1.50	1,690	yes
No education, 15+	region	19	2.60 %	1.55–3.65 %	1.80	1,587	yes
No education, 15+	region	2	19.54 %	15.24–23.83 %	5.80	1,898	yes
No education, 15+	region	20	4.68 %	3.63–5.73 %	1.10	1,703	yes
No education, 15+	region	21	4.92 %	3.66–6.18 %	1.59	1,811	yes
No education, 15+	region	22	3.13 %	2.08–4.18 %	1.78	1,863	yes
No education, 15+	region	23	57.49 %	48.79–66.18 %	14.45	1,795	yes
No education, 15+	region	24	32.13 %	21.02–43.23 %	29.75	2,021	yes
No education, 15+	region	25	47.38 %	37.89–56.86 %	15.84	1,687	yes
No education, 15+	region	26	5.35 %	3.87–6.82 %	2.19	1,958	yes
No education, 15+	region	27	2.74 %	1.89–3.59 %	1.50	2,126	yes
No education, 15+	region	28	5.73 %	3.89–7.58 %	3.06	1,867	yes
No education, 15+	region	29	4.26 %	2.79–5.74 %	2.71	1,948	yes
No education, 15+	region	3	15.86 %	13.56–18.16 %	2.12	2,049	yes
No education, 15+	region	30	12.24 %	7.09–17.38 %	11.97	1,866	yes
No education, 15+	region	31	7.38 %	4.49–10.27 %	5.28	1,659	yes
No education, 15+	region	32	4.30 %	3.50–5.11 %	0.84	2,038	yes
No education, 15+	region	33	19.22 %	13.93–24.52 %	9.37	1,996	yes
No education, 15+	region	34	13.25 %	8.67–17.82 %	8.36	1,762	yes
No education, 15+	region	35	3.97 %	2.96–4.99 %	1.51	2,141	yes
No education, 15+	region	36	4.27 %	2.90–5.64 %	2.45	2,049	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
No education, 15+	region	37	6.25 %	4.76–7.74 %	2.17	2,194	yes
No education, 15+	region	38	4.35 %	3.24–5.47 %	1.67	2,148	yes
No education, 15+	region	39	2.56 %	1.87–3.24 %	1.06	2,157	yes
No education, 15+	region	4	44.34 %	34.58–54.10 %	17.87	1,779	yes
No education, 15+	region	40	7.01 %	5.55–8.47 %	1.83	2,143	yes
No education, 15+	region	41	5.64 %	3.55–7.72 %	4.03	1,896	yes
No education, 15+	region	42	3.16 %	2.19–4.14 %	1.67	2,076	yes
No education, 15+	region	43	5.29 %	4.10–6.47 %	1.51	2,060	yes
No education, 15+	region	44	4.96 %	3.78–6.13 %	1.55	2,028	yes
No education, 15+	region	45	3.62 %	2.52–4.72 %	1.82	2,014	yes
No education, 15+	region	46	4.75 %	3.54–5.96 %	1.63	1,947	yes
No education, 15+	region	47	2.27 %	0.24–4.29 %	11.64	2,413	yes
No education, 15+	region	5	15.96 %	11.64–20.29 %	6.95	1,913	yes
No education, 15+	region	6	6.12 %	4.52–7.72 %	1.83	1,578	yes
No education, 15+	region	7	47.07 %	34.92–59.21 %	28.50	1,849	yes
No education, 15+	region	8	55.73 %	49.86–61.60 %	8.07	2,218	yes
No education, 15+	region	9	62.17 %	57.25–67.09 %	5.35	1,995	yes
No education, 15+	residence	Rural	10.44 %	9.92–10.96 %	4.43	58,941	yes
No education, 15+	residence	Urban	4.66 %	3.95–5.37 %	9.37	31,513	yes
No education, 15+	sex	Female	10.85 %	10.32–11.37 %	3.46	47,110	yes
No education, 15+	sex	Male	5.61 %	5.21–6.01 %	3.41	43,344	yes
No education, 15+	wealth	Fourth	3.53 %	3.09–3.97 %	2.83	19,331	yes
No education, 15+	wealth	Highest	2.34 %	1.28–3.40 %	18.18	14,230	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
No education, 15+	wealth	Lowest	25.95 %	24.46–27.44 %	6.32	20,929	yes
No education, 15+	wealth	Middle	5.86 %	5.39–6.33 %	1.93	18,824	yes
No education, 15+	wealth	Second	7.96 %	7.38–8.54 %	2.06	17,140	yes
Out of school, primary age	national		5.78 %	4.96–6.59 %	9.76	31,035	yes
Out of school, primary age	region	1	2.81 %	1.31–4.30 %	0.92	430	yes
Out of school, primary age	region	10	21.89 %	11.90–31.89 %	11.99	788	yes
Out of school, primary age	region	11	11.54 %	5.27–17.81 %	8.55	853	yes
Out of school, primary age	region	12	0.69 %	0.00–1.60 %	1.42	447	yes
Out of school, primary age	region	13	4.32 %	1.48–7.17 %	2.14	419	yes
Out of school, primary age	region	14	0.27 %	0.00–0.80 %	1.04	391	yes
Out of school, primary age	region	15	1.93 %	0.95–2.91 %	0.75	567	yes
Out of school, primary age	region	16	0.00 %	0.00–0.00 %	—	432	yes
Out of school, primary age	region	17	1.29 %	0.27–2.30 %	0.92	436	yes
Out of school, primary age	region	18	0.40 %	0.00–1.00 %	1.01	426	yes
Out of school, primary age	region	19	3.20 %	0.00–7.09 %	3.79	297	yes
Out of school, primary age	region	2	7.84 %	4.30–11.38 %	3.09	683	yes
Out of school, primary age	region	20	0.00 %	0.00–0.00 %	—	351	yes
Out of school, primary age	region	21	1.03 %	0.09–1.97 %	0.89	390	yes
Out of school, primary age	region	22	0.70 %	0.00–2.07 %	2.91	419	yes
Out of school, primary age	region	23	43.45 %	32.53–54.38 %	11.39	901	yes
Out of school, primary age	region	24	24.82 %	7.73–41.91 %	42.46	1,042	yes
Out of school, primary age	region	25	32.86 %	21.70–44.02 %	12.45	847	yes
Out of school, primary age	region	26	2.02 %	0.95–3.10 %	1.16	767	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
Out of school, primary age	region	27	1.70 %	0.70–2.71 %	0.86	549	yes
Out of school, primary age	region	28	1.10 %	0.09–2.11 %	1.57	644	yes
Out of school, primary age	region	29	1.56 %	0.60–2.52 %	0.95	610	yes
Out of school, primary age	region	3	3.29 %	2.08–4.50 %	0.85	714	yes
Out of school, primary age	region	30	6.38 %	0.00–14.50 %	21.79	758	yes
Out of school, primary age	region	31	1.57 %	0.50–2.63 %	0.89	463	yes
Out of school, primary age	region	32	0.89 %	0.13–1.66 %	0.98	570	yes
Out of school, primary age	region	33	5.89 %	2.75–9.03 %	3.95	853	yes
Out of school, primary age	region	34	9.81 %	4.59–15.03 %	4.81	600	yes
Out of school, primary age	region	35	1.57 %	0.48–2.67 %	1.27	628	yes
Out of school, primary age	region	36	0.39 %	0.00–0.96 %	1.36	647	yes
Out of school, primary age	region	37	1.61 %	0.77–2.44 %	0.87	755	yes
Out of school, primary age	region	38	0.41 %	0.00–0.90 %	1.00	656	yes
Out of school, primary age	region	39	1.06 %	0.27–1.84 %	1.24	804	yes
Out of school, primary age	region	4	30.54 %	20.79–40.29 %	10.34	886	yes
Out of school, primary age	region	40	1.36 %	0.56–2.16 %	1.04	842	yes
Out of school, primary age	region	41	1.30 %	0.30–2.30 %	1.32	652	yes
Out of school, primary age	region	42	0.84 %	0.19–1.50 %	0.87	653	yes
Out of school, primary age	region	43	0.84 %	0.08–1.60 %	1.40	782	yes
Out of school, primary age	region	44	1.95 %	1.02–2.87 %	1.02	877	yes
Out of school, primary age	region	45	0.99 %	0.14–1.85 %	1.28	660	yes
Out of school, primary age	region	46	0.36 %	0.00–0.86 %	1.14	618	yes
Out of school, primary age	region	47	4.52 %	0.00–11.14 %	12.66	478	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
Out of school, primary age	region	5	5.04 %	2.49–7.59 %	2.25	637	yes
Out of school, primary age	region	6	2.41 %	0.51–4.30 %	1.78	447	yes
Out of school, primary age	region	7	54.09 %	47.53–60.65 %	4.43	982	yes
Out of school, primary age	region	8	38.65 %	29.01–48.28 %	11.64	1,142	yes
Out of school, primary age	region	9	45.76 %	40.75–50.78 %	3.27	1,242	yes
Out of school, primary age	residence	Rural	6.13 %	5.29–6.98 %	7.22	22,314	yes
Out of school, primary age	residence	Urban	4.85 %	2.92–6.78 %	18.35	8,721	yes
Out of school, primary age	sex	Female	5.49 %	4.74–6.23 %	4.28	15,306	yes
Out of school, primary age	sex	Male	6.07 %	5.09–7.04 %	6.80	15,729	yes
Out of school, primary age	wealth	Fourth	2.65 %	1.82–3.47 %	3.41	4,984	yes
Out of school, primary age	wealth	Highest	2.84 %	0.00–6.03 %	31.25	3,273	yes
Out of school, primary age	wealth	Lowest	15.47 %	13.33–17.60 %	9.74	10,747	yes
Out of school, primary age	wealth	Middle	2.11 %	1.64–2.58 %	1.61	5,812	yes
Out of school, primary age	wealth	Second	2.33 %	1.90–2.77 %	1.36	6,219	yes
Over-age for grade	national		50.56 %	49.53–51.59 %	4.00	36,048	yes
Over-age for grade	region	1	50.71 %	45.77–55.65 %	1.32	518	yes
Over-age for grade	region	10	46.69 %	41.32–52.07 %	2.38	788	yes
Over-age for grade	region	11	46.71 %	40.78–52.65 %	3.18	864	yes
Over-age for grade	region	12	47.28 %	39.02–55.53 %	3.94	554	yes
Over-age for grade	region	13	51.30 %	43.58–59.02 %	3.32	535	yes
Over-age for grade	region	14	31.42 %	25.44–37.39 %	2.12	492	yes
Over-age for grade	region	15	46.36 %	41.91–50.81 %	1.63	784	yes
Over-age for grade	region	16	28.50 %	23.75–33.24 %	1.67	580	yes
Over-age for grade	region	17	44.75 %	39.76–49.73 %	1.65	632	yes
Over-age for grade	region	18	40.83 %	35.24–46.41 %	1.79	533	yes
Over-age for grade	region	19	32.79 %	26.15–39.42 %	1.91	368	yes
Over-age for grade	region	2	76.67 %	72.64–80.71 %	2.07	874	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
Over-age for grade	region	20	32.71 %	27.68–37.74 %	1.23	412	yes
Over-age for grade	region	21	34.20 %	28.92–39.48 %	1.58	491	yes
Over-age for grade	region	22	28.60 %	23.42–33.77 %	1.65	485	yes
Over-age for grade	region	23	67.92 %	63.01–72.83 %	1.84	639	yes
Over-age for grade	region	24	62.52 %	56.95–68.10 %	3.49	1,013	yes
Over-age for grade	region	25	58.09 %	50.95–65.24 %	3.54	647	yes
Over-age for grade	region	26	55.32 %	51.51–59.13 %	1.46	954	yes
Over-age for grade	region	27	51.24 %	45.35–57.13 %	2.62	725	yes
Over-age for grade	region	28	43.93 %	38.83–49.04 %	2.33	846	yes
Over-age for grade	region	29	55.63 %	50.79–60.47 %	2.04	824	yes
Over-age for grade	region	3	73.48 %	69.59–77.37 %	1.98	981	yes
Over-age for grade	region	30	44.68 %	37.33–52.03 %	5.10	896	yes
Over-age for grade	region	31	42.88 %	32.98–52.78 %	5.87	564	yes
Over-age for grade	region	32	46.67 %	41.33–52.02 %	2.16	722	yes
Over-age for grade	region	33	70.52 %	65.77–75.26 %	2.97	1,053	yes
Over-age for grade	region	34	36.03 %	30.04–42.02 %	2.48	612	yes
Over-age for grade	region	35	33.93 %	29.48–38.39 %	1.83	793	yes
Over-age for grade	region	36	51.89 %	46.75–57.03 %	2.48	900	yes
Over-age for grade	region	37	60.00 %	56.02–63.98 %	1.85	1,073	yes
Over-age for grade	region	38	55.17 %	50.91–59.43 %	1.80	943	yes
Over-age for grade	region	39	61.06 %	56.96–65.15 %	1.94	1,057	yes
Over-age for grade	region	4	74.02 %	69.68–78.36 %	1.87	731	yes
Over-age for grade	region	40	59.65 %	55.82–63.48 %	1.76	1,109	yes
Over-age for grade	region	41	42.51 %	38.04–46.98 %	1.84	864	yes
Over-age for grade	region	42	47.80 %	43.07–52.53 %	1.95	834	yes
Over-age for grade	region	43	62.47 %	57.96–66.98 %	2.24	990	yes
Over-age for grade	region	44	63.61 %	59.34–67.88 %	2.26	1,100	yes
Over-age for grade	region	45	55.42 %	51.25–59.59 %	1.64	895	yes
Over-age for grade	region	46	54.84 %	49.26–60.43 %	2.83	864	yes
Over-age for grade	region	47	31.09 %	25.73–36.46 %	2.02	577	yes
Over-age for grade	region	5	64.98 %	59.08–70.87 %	3.11	783	yes
Over-age for grade	region	6	42.82 %	37.54–48.10 %	1.55	523	yes
Over-age for grade	region	7	71.50 %	62.36–80.65 %	6.54	612	yes
Over-age for grade	region	8	73.42 %	69.40–77.43 %	2.21	1,024	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
Over-age for grade	region	9	75.91 %	70.39–81.44 %	4.30	990	yes
Over-age for grade	residence	Rural	55.05 %	54.04–56.06 %	2.79	26,001	yes
Over-age for grade	residence	Urban	38.05 %	35.88–40.21 %	5.20	10,047	yes
Over-age for grade	sex	Female	46.23 %	44.96–47.51 %	3.02	17,748	yes
Over-age for grade	sex	Male	54.86 %	53.67–56.05 %	2.71	18,300	yes
Over-age for grade	wealth	Fourth	40.50 %	38.36–42.63 %	2.97	6,047	yes
Over-age for grade	wealth	Highest	27.24 %	24.94–29.54 %	2.66	3,834	yes
Over-age for grade	wealth	Lowest	68.55 %	67.03–70.07 %	2.91	10,467	yes
Over-age for grade	wealth	Middle	48.22 %	46.63–49.81 %	1.99	7,556	yes
Over-age for grade	wealth	Second	56.70 %	55.06–58.33 %	2.31	8,144	yes
Years of schooling, 20–24	national		10.745	10.63–10.86	—	12,470	yes
Years of schooling, 20–24	region	1	10.962	10.45–11.48	—	322	yes
Years of schooling, 20–24	region	10	6.223	5.02–7.43	—	221	yes
Years of schooling, 20–24	region	11	9.368	8.46–10.28	—	237	yes
Years of schooling, 20–24	region	12	9.772	9.01–10.54	—	239	yes
Years of schooling, 20–24	region	13	10.870	10.29–11.45	—	201	yes
Years of schooling, 20–24	region	14	11.369	10.99–11.75	—	162	yes
Years of schooling, 20–24	region	15	10.434	9.56–11.30	—	247	yes
Years of schooling, 20–24	region	16	11.540	11.16–11.92	—	268	yes
Years of schooling, 20–24	region	17	10.985	10.63–11.34	—	271	yes
Years of schooling, 20–24	region	18	10.887	10.27–11.50	—	161	yes
Years of schooling, 20–24	region	19	12.174	11.72–12.63	—	195	yes
Years of schooling, 20–24	region	2	9.184	8.48–9.89	—	269	yes
Years of schooling, 20–24	region	20	10.988	10.54–11.44	—	176	yes
Years of schooling, 20–24	region	21	11.039	10.54–11.54	—	201	yes
Years of schooling, 20–24	region	22	11.888	11.47–12.31	—	236	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
Years of schooling, 20–24	region	23	5.950	4.73–7.17	—	262	yes
Years of schooling, 20–24	region	24	8.341	7.46–9.23	—	337	yes
Years of schooling, 20–24	region	25	6.901	5.59–8.21	—	287	yes
Years of schooling, 20–24	region	26	10.398	9.94–10.85	—	251	yes
Years of schooling, 20–24	region	27	11.806	11.30–12.31	—	372	yes
Years of schooling, 20–24	region	28	10.795	10.32–11.28	—	275	yes
Years of schooling, 20–24	region	29	11.029	10.03–12.03	—	276	yes
Years of schooling, 20–24	region	3	9.346	8.89–9.80	—	303	yes
Years of schooling, 20–24	region	30	10.054	9.18–10.93	—	261	yes
Years of schooling, 20–24	region	31	11.538	10.69–12.39	—	196	yes
Years of schooling, 20–24	region	32	11.074	10.70–11.45	—	298	yes
Years of schooling, 20–24	region	33	9.342	8.93–9.76	—	321	yes
Years of schooling, 20–24	region	34	10.871	10.13–11.61	—	270	yes
Years of schooling, 20–24	region	35	10.950	10.53–11.37	—	342	yes
Years of schooling, 20–24	region	36	10.603	10.23–10.97	—	299	yes
Years of schooling, 20–24	region	37	10.362	9.92–10.80	—	260	yes
Years of schooling, 20–24	region	38	10.312	9.97–10.65	—	258	yes
Years of schooling, 20–24	region	39	10.588	10.17–11.00	—	321	yes
Years of schooling, 20–24	region	4	6.557	5.47–7.65	—	284	yes
Years of schooling, 20–24	region	40	9.726	9.31–10.14	—	321	yes
Years of schooling, 20–24	region	41	10.141	9.58–10.70	—	221	yes
Years of schooling, 20–24	region	42	10.972	10.59–11.35	—	320	yes
Years of schooling, 20–24	region	43	10.612	10.21–11.02	—	293	yes

Indicator	Dimension	Level	Estimate	95 % interval	deff	n	Rel.
Years of schooling, 20–24	region	44	9.744	9.34–10.15	—	296	yes
Years of schooling, 20–24	region	45	11.339	10.71–11.97	—	281	yes
Years of schooling, 20–24	region	46	11.141	10.81–11.47	—	207	yes
Years of schooling, 20–24	region	47	12.143	11.71–12.58	—	429	yes
Years of schooling, 20–24	region	5	8.830	8.19–9.47	—	251	yes
Years of schooling, 20–24	region	6	10.557	10.12–10.99	—	189	yes
Years of schooling, 20–24	region	7	8.040	6.93–9.16	—	244	yes
Years of schooling, 20–24	region	8	6.909	5.26–8.56	—	297	yes
Years of schooling, 20–24	region	9	6.688	5.73–7.65	—	242	yes
Years of schooling, 20–24	residence	Rural	9.982	9.86–10.11	—	7,394	yes
Years of schooling, 20–24	residence	Urban	11.731	11.55–11.91	—	5,076	yes
Years of schooling, 20–24	sex	Female	10.752	10.61–10.90	—	6,717	yes
Years of schooling, 20–24	sex	Male	10.736	10.59–10.88	—	5,753	yes
Years of schooling, 20–24	wealth	Fourth	11.478	11.33–11.63	—	3,057	yes
Years of schooling, 20–24	wealth	Highest	12.466	12.22–12.72	—	2,072	yes
Years of schooling, 20–24	wealth	Lowest	7.893	7.68–8.11	—	2,633	yes
Years of schooling, 20–24	wealth	Middle	10.603	10.44–10.77	—	2,500	yes
Years of schooling, 20–24	wealth	Second	9.874	9.72–10.03	—	2,208	yes

G.3 The household-panel estimates

Dimension	Level	Spend per student	Budget share	Catastrophic 10 %	n	Rel.
quintile	Quintile 1	98k UGX	6.8 %	24.4 %	449	yes
quintile	Quintile 2	136k UGX	7.1 %	25.0 %	487	yes
quintile	Quintile 3	229k UGX	8.9 %	31.9 %	474	yes
quintile	Quintile 4	367k UGX	9.7 %	34.6 %	462	yes
quintile	Quintile 5	933k UGX	13.4 %	42.9 %	371	yes
residence	Rural	243k UGX	7.9 %	26.9 %	1,725	yes
residence	Urban	691k UGX	12.9 %	45.0 %	518	yes

Dimension	Level	Spend per student	Budget share	Catastrophic 10 %	n	Rel.
region	Central	502k UGX	10.0 %	37.7 %	559	yes
region	Eastern	319k UGX	9.2 %	32.7 %	563	yes
region	Northern	268k UGX	7.4 %	20.5 %	565	yes
region	Western	352k UGX	10.1 %	33.6 %	556	yes
poverty	Non-poor	414k UGX	9.7 %	33.5 %	1,917	yes
poverty	Poor	86k UGX	6.5 %	23.1 %	326	yes

G.4 The school register, district by district

58 districts, from the register's own district field. These are administrative counts, not survey estimates: they have a coverage problem rather than a sampling one, and a quarter of schools record no pupils.

DISTNAM	schools	pupils	teachers	classrooms	ptr	pcr
Pader	153	99293.0	1212.0	764.0	81.92491749174917	129.96465968586386
Kitgum	136	76987.0	1125.0	735.0	68.43288888888888	104.74421768707484
Soroti	191	136893.0	2047.0	1417.0	66.87493893502686	96.6076217360621
Nebbi	251	148113.0	2292.0	1482.0	64.6217277486911	99.9412955465587
Kamuli	390	223264.0	3554.0	2275.0	62.82048396173326	98.13802197802198
BUGIRI	173	57477.0	924.0	609.0	62.20454545454545	94.37931034482759
Pallisa	222	166841.0	2698.0	1777.0	61.838769458858415	93.88913899831176
Kyenjojo	215	101087.0	1642.0	1193.0	61.56333739342266	84.73344509639564
Lira	327	264323.0	4325.0	2826.0	61.11514450867052	93.53255484784147
Arua	364	349199.0	5884.0	2722.0	59.34721278042148	128.2876561351947
Tororo	292	172072.0	2904.0	1811.0	59.2534435261708	95.01490889011596
Kotido	89	39274.0	665.0	516.0	59.05864661654135	76.1124031007752
Gulu	251	164442.0	2794.0	1935.0	58.85540443808161	84.9829457364341
Apac	286	233323.0	3972.0	2271.0	58.741943605236656	102.74020255394099
Mayuge	177	120124.0	2110.0	974.0	56.930805687203794	123.3305954825462
Kasese	400	169097.0	3095.0	1812.0	54.63554119547658	93.32064017660043
Iganga	435	232070.0	4281.0	2410.0	54.20929689324924	96.29460580912863
Kumi	225	121731.0	2249.0	1454.0	54.12672298799466	83.72145804676754
Katakwi	173	95163.0	1765.0	1017.0	53.91671388101983	93.57227138643069
Sironko	179	100770.0	1887.0	1213.0	53.40222575516693	83.0750206100577
Ntungamo	321	116964.0	2203.0	1609.0	53.09305492510213	72.69359850839031
Hoima	220	90249.0	1700.0	1321.0	53.08764705882353	68.31869795609387
Sembabule	212	82703.0	1563.0	876.0	52.91298784388996	94.40981735159818
Masindi	239	131730.0	2504.0	1462.0	52.60782747603834	90.10259917920656
Mbale	397	240900.0	4661.0	2446.0	51.68418794250161	98.48732624693378
Bugiri	101	65712.0	1274.0	772.0	51.57927786499215	85.11917098445596

DISTNAM	schools	pupils	teachers	classrooms	ptr	pcr
Nakapiripir	64	17178.0	335.0	222.0	51.277611940298506	77.3783783783783 7837
Kaberamaido	90	41738.0	815.0	568.0	51.21226993865031	73.48239436619718
Kamwenge	156	73810.0	1442.0	983.0	51.18585298196949	75.08646998982707
Moyo	77	38013.0	746.0	568.0	50.95576407506702	66.92429577464789
Busia	136	69999.0	1382.0	894.0	50.650506512301014	78.29865771812081
Kayunga	296	105842.0	2125.0	1367.0	49.808	77.42648134601316
Kabarole	154	103924.0	2088.0	1204.0	49.772030651341	86.3156146179402
Bundibugyo	159	49321.0	993.0	636.0	49.6686807653575	77.54874213836477
Mpigi	420	134318.0	2737.0	1918.0	49.074899525027405	70.03023983315954
Bushenyi	616	212161.0	4324.0	3350.0	49.0659111933395	63.331641791044774
Kibaale	294	116668.0	2389.0	1582.0	48.83549602344077	73.74715549936789
Kisoro	143	62643.0	1286.0	1003.0	48.711508553654745	62.45563310069791
Kanungu	159	64705.0	1355.0	980.0	47.752767527675275	66.02551020408163
Moroto	46	16737.0	351.0	328.0	47.68376068376068	51.02743902439025
Adjumani	88	42308.0	888.0	741.0	47.64414414414414	57.095816464237515
Luwero	527	167611.0	3541.0	2617.0	47.334368822366564	64.04700038211693
Mubende	779	175480.0	3933.0	2524.0	44.61734045258073	69.52456418383518
Kiboga	61	20511.0	461.0	336.0	44.49240780911063	61.044642857142854
Jinja	186	104238.0	2355.0	1450.0	44.26242038216561	71.88827586206897
Kalangala	21	3082.0	71.0	113.0	43.40845070422535	27.27433628318584
Yumbe	112	70561.0	1633.0	635.0	43.20943049601959	111.11968503937008
Kapchorwa	140	55796.0	1305.0	812.0	42.75555555555555	68.71428571428571
KIBOGA	168	52723.0	1257.0	832.0	41.94351630867144	63.36899038461539
Rakai	430	135406.0	3243.0	2050.0	41.75331483194573	66.05170731707317
Masaka	712	215546.0	5194.0	3169.0	41.49903735078937	68.01704007573368
Rukungiri	224	86493.0	2088.0	1516.0	41.423850574712645	57.053430079155675
Mbarara	963	303132.0	7504.0	4385.0	40.396055437100216	69.12930444697834
Mukono	912	223472.0	5552.0	3541.0	40.2507204610951	63.10985597288 9015
Kabale	316	145676.0	3639.0	2279.0	40.03187688925529	63.92101799034664
Nakasongola	174	47002.0	1220.0	841.0	38.52622950819672	55.88822829964328
Wakiso	1103	176341.0	5282.0	3718.0	33.385270730783795	47.428994082840234
Kampala	739	161825.0	5488.0	3942.0	29.487062682215743	41.05149670218163

G.5 The classroom gap, district by district

At the declared norm. Recompute at another norm and every row changes; the norm is the policy decision.

DISTNAM	schools	pupils	teachers	classrooms	ptr	pcr	classrooms_required	classroom_gap
Pader	153	99293.0	1212.0	764.0	81.92491749174917	129.96465968586386	2206.511111111111	1442.511111111111
Kitgum	136	76987.0	1125.0	735.0	68.43288888888888	104.74421768707484	1710.8222222222223	975.8222222222223
Soroti	191	136893.0	2047.0	1417.0	66.87493893502686	96.6076217360621	3042.0666666666666	1625.0666666666666
Nebbi	251	148113.0	2292.0	1482.0	64.6217277486911	99.9412955465587	3291.4	1809.4
Kamuli	390	223264.0	3554.0	2275.0	62.82048396173326	98.13802197802198	4961.422222222222	2686.4222222222224
BUGIRI	173	57477.0	924.0	609.0	62.20454545454545	94.37931034482759	1277.2666666666667	668.2666666666667
Pallisa	222	166841.0	2698.0	1777.0	61.838769458858415	93.88913899831176	3707.5777777777776	1930.5777777777776
Kyenjojo	215	101087.0	1642.0	1193.0	61.56333739342266	84.73344509639564	2246.3777777777777	1053.3777777777777
Lira	327	264323.0	4325.0	2826.0	61.11514450867052	93.53255484784147	5873.844444444445	3047.844444444445
Arua	364	349199.0	5884.0	2722.0	59.34721278042148	128.2876561351947	7759.977777777778	5037.977777777778
Tororo	292	172072.0	2904.0	1811.0	59.2534435261708	95.01490889011596	3823.822222222222	2012.822222222222
Kotido	89	39274.0	665.0	516.0	59.05864661654135	76.1124031007752	872.7555555555556	356.7555555555556
Gulu	251	164442.0	2794.0	1935.0	58.85540443808161	84.9829457364341	3654.266666666667	1719.2666666666669
Apac	286	233323.0	3972.0	2271.0	58.741943605236656	102.74020255394099	5184.955555555555	2913.9555555555553
Mayuge	177	120124.0	2110.0	974.0	56.930805687203794	123.3305954825462	2669.4222222222224	1695.4222222222224
Kasese	400	169097.0	3095.0	1812.0	54.63554119547658	93.32064017660043	3757.711111111111	1945.7111111111112
Iganga	435	232070.0	4281.0	2410.0	54.20929689324924	96.29460580912863	5157.111111111111	2747.1111111111113
Kumi	225	121731.0	2249.0	1454.0	54.12672298799466	83.72145804676754	2705.133333333333	1251.1333333333332
Katakwi	173	95163.0	1765.0	1017.0	53.91671388101983	93.57227138643069	2114.733333333333	1097.7333333333331
Sironko	179	100770.0	1887.0	1213.0	53.40222575516693	83.0750206100577	2239.3333333333335	1026.3333333333335
Ntungamo	321	116964.0	2203.0	1609.0	53.09305492510213	72.69359850839031	2599.2	990.1999999999998
Hoima	220	90249.0	1700.0	1321.0	53.08764705882353	68.31869795609387	2005.5333333333333	684.5333333333333
Sembabule	212	82703.0	1563.0	876.0	52.91298784388996	94.40981735159818	1837.8444444444444	961.8444444444444
Masindi	239	131730.0	2504.0	1462.0	52.60782747603834	90.10259917920656	2927.3333333333335	1465.3333333333335
Mbale	397	240900.0	4661.0	2446.0	51.68418794250161	98.48732624693378	5353.333333333333	2907.333333333333
Bugiri	101	65712.0	1274.0	772.0	51.57927786499215	85.11917098445596	1460.2666666666667	688.2666666666667
Nakapiripir	64	17178.0	335.0	222.0	51.277611940298506	77.37837837837837	381.733333333333335	159.733333333333335
Kaberamaido	90	41738.0	815.0	568.0	51.21226993865031	73.48239436619718	927.5111111111111	359.511111111111106
Kamwenge	156	73810.0	1442.0	983.0	51.18585298196949	75.08646998982707	1640.2222222222222	657.2222222222222
Moyo	77	38013.0	746.0	568.0	50.95576407506702	66.92429577464789	844.7333333333333	276.733333333333335

DISTNAM	schools	pupils	teachers	classrooms	ptr	pcr	classrooms_required	classroom_gap
Busia	136	69999.0	1382.0	894.0	50.650506512301014	78.29865771812081	1555.5333333333333	661.5333333333333
Kayunga	296	105842.0	2125.0	1367.0	49.808	77.42648134601316	2352.0444444444443	985.0444444444443
Kabarole	154	103924.0	2088.0	1204.0	49.772030651341	86.3156146179402	2309.4222222222224	1105.4222222222224
Bundibugyo	159	49321.0	993.0	636.0	49.6686807653575	77.54874213836477	1096.0222222222221	460.0222222222221
Mpigi	420	134318.0	2737.0	1918.0	49.074899525027405	70.03023983315954	2984.844444444444	1066.844444444444
Bushenyi	616	212161.0	4324.0	3350.0	49.0659111933395	63.331641791044774	4714.688888888889	1364.688888888889
Kibaale	294	116668.0	2389.0	1582.0	48.83549602344077	73.74715549936789	2592.6222222222223	1010.6222222222223
Kisoro	143	62643.0	1286.0	1003.0	48.711508553654745	62.45563310069791	1392.0666666666666	389.0666666666666
Kanungu	159	64705.0	1355.0	980.0	47.752767527675275	66.02551020408163	1437.888888888889	457.888888888889
Moroto	46	16737.0	351.0	328.0	47.68376068376068	51.02743902439025	371.9333333333334	43.9333333333334
Adjumani	88	42308.0	888.0	741.0	47.64414414414414	57.095816464237515	940.1777777777778	199.1777777777778
Luwero	527	167611.0	3541.0	2617.0	47.334368822366564	64.04700038211693	3724.688888888889	1107.688888888889
Mubende	779	175480.0	3933.0	2524.0	44.61734045258073	69.52456418383518	3899.5555555555557	1375.5555555555557
Kiboga	61	20511.0	461.0	336.0	44.49240780911063	61.044642857142854	455.8	119.80000000000001
Jinja	186	104238.0	2355.0	1450.0	44.26242038216561	71.88827586206897	2316.4	866.4000000000001
Kalangala	21	3082.0	71.0	113.0	43.40845070422535	27.27433628318584	68.48888888888889	0.0
Yumbe	112	70561.0	1633.0	635.0	43.20943049601959	111.11968503937008	1568.0222222222221	933.0222222222221
Kapchorwa	140	55796.0	1305.0	812.0	42.75555555555555	68.71428571428571	1239.9111111111111	427.91111111111104
KIBOGA	168	52723.0	1257.0	832.0	41.94351630867144	63.36899038461539	1171.6222222222223	339.62222222222226
Rakai	430	135406.0	3243.0	2050.0	41.75331483194573	66.05170731707317	3009.0222222222224	959.0222222222224
Masaka	712	215546.0	5194.0	3169.0	41.49903735078937	68.01704007573368	4789.911111111115	1620.911111111115
Rukungiri	224	86493.0	2088.0	1516.0	41.423850574712645	57.053430079155675	1922.0666666666666	406.0666666666666
Mbarara	963	303132.0	7504.0	4385.0	40.396055437100216	69.12930444697834	6736.266666666666	2351.2666666666664
Mukono	912	223472.0	5552.0	3541.0	40.2507204610951	63.109855972889015	4966.0444444444445	1425.0444444444447
Kabale	316	145676.0	3639.0	2279.0	40.03187688925529	63.92101799034664	3237.2444444444445	958.2444444444445
Nakasongola	174	47002.0	1220.0	841.0	38.52622950819672	55.88822829964328	1044.488888888888	203.48888888888882
Wakiso	1103	176341.0	5282.0	3718.0	33.385270730783795	47.428994082840234	3918.688888888889	200.688888888887

DISTNAM	schools	pupils	teachers	classrooms	ptr	pcr	classrooms__required	classroom__gap
Kampala	739	161825.0	5488.0	3942.0	29.48706 26822 15743	41.05149 670218163	3596.11111 11111113	0.0

G.6 Distance to the nearest school, by county

149 counties with at least ten settlements. Straight-line distance, from the settlement register to the nearest school of any level.

admin2__name	settlements	median__km	p90__km	population
Pokot	152	3.64825	7.939400000000001	0.0
Bokora	268	3.29735	10.28245	0.0
Chekwii	155	2.4164	4.153379999999999	0.0
Jie	180	2.38705	5.7469399999999995	22900.0
Pian	64	2.29015	8.932210000000001	0.0
Ntoroko	527	2.2762	4.865360000000001	0.0
Aruu	655	2.0731	4.05276	0.0
Lamwo	384	2.0575	5.0207799999999985	0.0
Otuke	465	2.0229	3.31672	0.0
Nakaseke	400	2.0129	5.48858	0.0
Dodoth	626	1.9319000000000002	5.95125	23900.0
Matheniko	166	1.93175	6.7428	0.0
Kyamuswa	70	1.8436	6.17303	0.0
Usuk	404	1.8287	4.1312299999999995	0.0
Moroto	635	1.7175	2.85096	0.0
Amuria	307	1.6577	2.78414	0.0
Bbaale	220	1.6559	3.21493	0.0
Agago	1044	1.6405	3.15936	15000.0
Chua	616	1.6319	3.60365	19300.0
Kwania	505	1.6209	2.9231200000000004	0.0
Obongi	121	1.617	3.6086	7741.0
Kioga	483	1.5839	3.1548400000000005	0.0
Maruzi	362	1.5784500000000001	2.99853	13700.0
Bujenje	189	1.5751	2.66772	0.0
Bujjumba	73	1.5392	3.45296	0.0
Oyam	700	1.5286	2.6630599999999998	0.0
Aswa	122	1.5035	3.4943500000000003	0.0
Lwemiyaga	99	1.4985	2.83494	0.0
Kasilo	210	1.4788999999999999	2.83716	0.0
Dokolo	572	1.4727000000000001	2.60093	0.0
Bukanga	436	1.4677	4.2516	0.0
Kaberamaido	278	1.467	2.7206	0.0
Kilak	183	1.465	3.6096400000000006	0.0
East Moyo	410	1.4364	2.62426	0.0
Nyabushozi	342	1.43215	3.4716700000000027	600.0
Buruli	753	1.4018	3.1717000000000013	3500.0
Kalaki	256	1.38615	2.29795	0.0
Buwekula	655	1.3719	3.62892	700.0
Soroti	906	1.3712499999999999	2.58805	3000.0
Mwenge	601	1.3604	2.9771	0.0
Kole	459	1.3515	2.2911200000000003	0.0
Labwor	303	1.348	3.2723200000000006	0.0
Serere	229	1.3456	2.346660000000001	0.0
Buyanja	286	1.344	2.3582	7600.0
Kibanda	254	1.3336999999999999	2.6415800000000007	8800.0
Omoro	138	1.3053	2.86507	152276.0
Busiro	783	1.297	4.17416	214399.0
Kakuuto	184	1.29285	3.5954800000000002	0.0
Bugangaizi	525	1.2884	2.4163400000000004	0.0
Erute	760	1.2827	2.37583	99059.0
Buruuli	248	1.2755	3.622830000000001	94622.0
Kyaka	536	1.2467000000000001	2.2489999999999997	0.0

admin2__name	settlements	median__km	p90__km	population
Buvuma	336	1.23	2.97645	0.0
Kumi	247	1.2268	2.49362	13000.0
Kazo	251	1.219	2.4436	0.0
Budiope	385	1.1924	2.782360000000001	0.0
Buliisa	206	1.1694499999999999	2.9968	0.0
Koboko	400	1.1668	2.71098	0.0
Jonam	416	1.1644999999999999	2.83495	0.0
Nwoya	211	1.1621	4.047	22360.0
Kibale	665	1.126	1.9023000000000003	0.0
Kassanda	624	1.11635	2.1972700000000014	0.0
Gomba	305	1.0997	2.510820000000002	0.0
Kyadondo	267	1.0981	3.6424400000000015	485761.0
Madi-Okollo	477	1.0973	2.78336	18933.0
Bugahya	436	1.0925500000000001	2.416	100625.0
Kabula	221	1.0867	2.1993	0.0
Kooki	494	1.0787	2.138149999999999	0.0
Ngora	153	1.0752	2.7410400000000004	0.0
Kiboga	742	1.06695	2.6395	0.0
Budadiri	1207	1.0539	2.1886000000000005	0.0
Padyere	1322	1.0473	1.90384	3333.0
Isingiro	525	1.029	2.3046800000000007	195013.0
West Moyo	192	1.0208	2.74331	0.0
Okoro	1233	1.0203	1.8841599999999998	297491.0
Bamunanika	82	1.0163	3.0463900000000006	5000.0
Rushenyi	139	1.0124	2.01372	0.0
Bukedea	189	1.0037	2.0153800000000013	0.0
Kinkiizi	3442	0.99545	1.8395000000000004	17273.0
Maracha	418	0.9901	1.66716	885.0
Bukomansimbi	49	0.9765	1.6758000000000002	0.0
Kitagwenda	240	0.9742500000000001	1.82177	0.0
Butebo	275	0.9548	1.65452	0.0
Busiki	402	0.9492499999999999	1.72207	0.0
Buyaga	712	0.9489000000000001	1.8153599999999999	1288.0
Bulamogi	365	0.9408	1.7847000000000006	0.0
Ntenjeru	203	0.9356	1.8738400000000004	0.0
Rwampara	162	0.9296500000000001	1.7140300000000002	0.0
Kibuku	306	0.92845	1.8416	0.0
West Budama	590	0.9221	1.6512400000000003	0.0
Buhaguzi	491	0.9142	2.1287	116625.0
Aringa	1189	0.9109	1.9271000000000003	534337.0
Mawogola	239	0.903	2.07084	0.0
Mawokota	422	0.90025	2.2051	0.0
Buhweju	264	0.88765	1.6378700000000006	0.0
Bunyole	400	0.87755	1.7616200000000002	0.0
Ibanda	668	0.8763	1.5550300000000004	0.0
Buzaaya	275	0.8715	1.51044	0.0
Rubanda	610	0.86385	2.1485800000000004	0.0
Busujju	184	0.86	1.6474100000000003	0.0
Bugabula	539	0.8597	1.82318	0.0
Luuka	181	0.8596	1.5285	0.0
Budaka	256	0.859	1.7016499999999999	178900.0
Bunyangabu	4193	0.8577	1.5420400000000003	0.0
Pallisa	426	0.85345	1.79475	0.0
Kashari	295	0.8444	1.7082200000000012	0.0
Vurra	294	0.8269	2.08281	33122.0
Samia-Bugwe	907	0.8232	1.5814199999999998	14624.0
Mityana	467	0.817	1.6781800000000007	0.0
Kalungu	331	0.8042	1.6257	0.0
Ndorwa	845	0.7831	1.4720000000000002	0.0
Kyotera	282	0.78165	1.64301	32700.0
Bukonjo	736	0.7782	2.4693	0.0
Ruhaama	397	0.7738	1.4873800000000001	0.0
Rukiga	384	0.77045	1.52544	0.0
Kween	599	0.769	3.1589800000000037	0.0
Bukooli	601	0.7616	1.6187	0.0

admin2__name	settlements	median__km	p90__km	population
Bujumbura	1393	0.7613	1.48182	2300.0
Bukuto	953	0.7604	1.6744200000000005	103829.0
Katikamu	101	0.7569	1.9872	96131.0
Rubabo	556	0.7563	1.3981	0.0
Nakifuma	252	0.7522	1.59237	0.0
Burahya	1298	0.74735	1.6155499999999998	54275.0
Busongora	1162	0.74335	1.7513700000000012	101679.0
Bugweri	262	0.7354499999999999	1.49475	0.0
Kajara	497	0.7274	1.40318	0.0
Terego	602	0.7229	1.5458499999999997	68167.0
Bughendra	1198	0.7219	1.4570100000000001	0.0
Kongasis	410	0.7101500000000001	1.72028	0.0
Tororo	521	0.71	1.5771	0.0
Ruhinda	1188	0.70395	1.3375799999999998	0.0
Sheema	603	0.6923	1.3317800000000004	0.0
Ayivu	468	0.6715	1.52584	69957.0
Buikwe	510	0.6591	1.5403	0.0
Bufumbira	1843	0.6482	1.2131800000000001	249970.0
Mukono	420	0.6338	1.5788000000000004	161996.0
Igara	732	0.62935	1.27383	41063.0
Tingey	723	0.6108	1.2152800000000006	66000.0
Butambala	288	0.5998	1.3666900000000004	0.0
Bunyaruguru	320	0.5949	1.2642400000000003	2309.0
Bubulo	1623	0.5853	1.09602	0.0
Kagoma	474	0.5811999999999999	1.26812	22000.0
Kigulu	492	0.56535	1.34915	0.0
Bunya	299	0.5525	1.7594599999999994	72931.0
Bulambuli	1183	0.542	1.1126999999999998	0.0
Bungokho	876	0.5225500000000001	1.1475	96189.0
Manjiya	998	0.44465	0.8038100000000001	0.0
Butembe	346	0.34955	1.29645	0.0
Kampala	968	0.18335	0.43854000000000015	1516210.0

G.7 Pupil-teacher and pupil-classroom ratios, by county

Computed on the schools whose coordinates could be placed inside a county polygon — about half the register. The reliability flag marks counties with fewer than five mapped schools; they are left blank on the maps.

admin2__name	schools	pupils	teachers	classrooms	ptr	pcr	reliable
Agago	25	17542.0	246.0	157.0	71.30894 30894309	111.73248 407643312	True
Amuria	21	12599.0	214.0	128.0	58.87383 1775700936	98.4296875	True
Aringa	47	31448.0	632.0	243.0	49.75949 367088607	129.41563 78600823	True
Aruu	34	25763.0	357.0	201.0	72.16526 610644257	128.17412 935323384	True
Aswa	62	49864.0	870.0	607.0	57.31494 2528735635	82.14827 01812191	True
Ayivu	29	30625.0	597.0	323.0	51.29815 745393635	94.81424 148606811	True
Bamunanika	44	18133.0	383.0	292.0	47.34464 7519582246	62.09931 506849315	True
Bbaale	77	39576.0	789.0	465.0	50.15969 5817490494	85.10967 741935484	True
Bubulo	47	28519.0	507.0	288.0	56.25049 309664694	99.02430 555555556	True
Budadiri	52	29138.0	564.0	388.0	51.66312 0567375884	75.09793 81443299	True
Budaka	18	15254.0	250.0	175.0	61.016	87.16571 428571429	True
Budiope	72	49329.0	679.0	397.0	72.64948 453608247	124.25440 80604534	True

admin2__ name	schools	pupils	teachers	classrooms	ptr	pcr	reliable
Bufumbira	66	30722.0	659.0	503.0	46.61911987860395	61.07753479125248	True
Bugabula	77	55373.0	949.0	561.0	58.348788198103264	98.70409982174688	True
Bugahya	53	24814.0	574.0	414.0	43.22996515679443	59.93719806763285	True
Bugangaizi	23	7459.0	180.0	115.0	41.43888888888889	64.86086956521739	True
Bughendra	54	27503.0	600.0	353.0	45.83833333333333	77.91218130311614	True
Bugweri	55	38957.0	751.0	402.0	51.87350199733688	96.90796019900498	True
Buhaguzi	18	10632.0	186.0	135.0	57.16129032258065	78.75555555555556	True
Buhweju	8	3238.0	57.0	37.0	56.80701754385965	87.51351351351352	True
Buikwe	130	45982.0	1100.0	735.0	41.801818181818184	62.560544217687074	True
Bujenje	11	4725.0	130.0	92.0	36.34615384615385	51.358695652173914	True
Bujjumba	9	2100.0	42.0	69.0	50.0	30.434782608695652	True
Bujumbura	46	22636.0	554.0	395.0	40.85920577617328	57.30632911392405	True
Bukanga	37	12854.0	276.0	151.0	46.572463768115945	85.12582781456953	True
Bukedea	18	10392.0	190.0	118.0	54.694736842105264	88.0677966101695	True
Bukomansimbi	50	23481.0	535.0	326.0	43.88971962616822	72.02760736196319	True
Bukonjo	122	77052.0	1489.0	797.0	51.74748153122901	96.67754077791719	True
Bukooli	102	57861.0	927.0	608.0	62.41747572815534	95.16611842105263	True
Bukuto	122	53962.0	1470.0	955.0	36.708843537414964	56.504712041884815	True
Bulambuli	13	6844.0	129.0	87.0	53.054263565891475	78.66666666666667	True
Bulamogi	65	47739.0	760.0	484.0	62.814473684210526	98.63429752066116	True
Buliisa	5	3255.0	54.0	33.0	60.27777777777778	98.63636363636364	True
Bungokho	60	39701.0	826.0	470.0	48.06416464891041	84.47021276595744	True
Bunya	107	78932.0	1430.0	727.0	55.1972027972028	108.57221458046767	True
Bunyangabu	31	22791.0	406.0	222.0	56.135467980295566	102.66216216216216	True
Bunyaruguru	7	3064.0	52.0	46.0	58.92307692307692	66.6086956521739	True
Bunyole	37	22996.0	357.0	197.0	64.41456582633053	116.73096446700508	True
Burahya	37	22584.0	484.0	341.0	46.66115702479339	66.22873900293256	True
Buruli	95	31464.0	822.0	535.0	38.277372262773724	58.811214953271026	True
Buruuli	16	8904.0	203.0	124.0	43.86206896551724	71.80645161290323	True
Busiki	48	29929.0	501.0	318.0	59.73852295409181	94.11635220125787	True
Busiro	258	96884.0	3038.0	2162.0	31.890717577353524	44.81221091581869	True
Busongora	117	61788.0	1101.0	710.0	56.119891008174385	87.02535211267606	True
Busujju	51	21278.0	521.0	300.0	40.840690978886755	70.92666666666666	True
Butambala	40	16687.0	380.0	257.0	43.91315789473684	64.92996108949416	True

admin2_ name	schools	pupils	teachers	classrooms	ptr	pcr	reliable
Butebo	22	18781.0	293.0	184.0	64.09897610921502	102.07065217391305	True
Butembe	23	12349.0	313.0	193.0	39.453674121405754	63.984455958549226	True
Buvuma	13	5791.0	78.0	78.0	74.24358974358974	74.24358974358974	True
Buwekula	91	40675.0	809.0	619.0	50.278121137206426	65.7108239095315	True
Buyaga	132	65711.0	1273.0	825.0	51.61901021209741	79.64969696969698	True
Buyanja	31	10245.0	234.0	161.0	43.782051282051285	63.63354037267081	True
Buzaaya	105	67859.0	1129.0	814.0	60.10540301151462	83.36486486486487	True
Chekwii	3	1472.0	37.0	26.0	39.78378378378378	56.61538461538461	False
Chua	26	20280.0	320.0	231.0	63.375	87.79220779220779	True
Dodoth	8	3275.0	38.0	44.0	86.1842105263158	74.43181818181819	True
Dokolo	18	12715.0	200.0	136.0	63.575	93.49264705882354	True
East Moyo	29	15746.0	298.0	260.0	52.83892617449664	60.56153846153846	True
Erute	54	38238.0	621.0	342.0	61.57487922705314	111.80701754385964	True
Gomba	52	24287.0	461.0	271.0	52.68329718004338	89.619926199262	True
Ibanda	97	41319.0	939.0	478.0	44.00319488817891	86.44142259414225	True
Igara	52	21565.0	505.0	348.0	42.7029702970297	61.9683908045977	True
Isingiro	94	35560.0	973.0	578.0	36.54676258992806	61.522491349480966	True
Jie	4	1379.0	27.0	22.0	51.074074074074076	62.68181818181818	False
Jonam	14	6716.0	99.0	57.0	67.83838383838383	117.82456140350877	True
Kaberamaido	42	21043.0	398.0	257.0	52.87185929648241	81.87937743190662	True
Kabula	15	6123.0	160.0	85.0	38.26875	72.03529411764706	True
Kagoma	27	19283.0	359.0	196.0	53.713091922005574	98.38265306122449	True
Kajara	90	48656.0	926.0	675.0	52.54427645788337	72.08296296296297	True
Kakuuto	6	2671.0	62.0	48.0	43.08064516129032	55.645833333333336	True
Kalaki	44	20695.0	417.0	311.0	49.62829736211031	66.54340836012862	True
Kalungu	93	51411.0	1187.0	681.0	43.3117101937658	75.49339207048457	True
Kampala	338	133203.0	4427.0	3136.0	30.08877343573526	42.47544642857143	True
Kashari	53	20497.0	582.0	378.0	35.218213058419245	54.22486772486773	True
Kasilo	13	8454.0	104.0	67.0	81.28846153846153	126.17910447761194	True
Kassanda	71	26459.0	639.0	379.0	41.406885758998435	69.81266490765171	True
Katikamu	119	53898.0	1249.0	909.0	43.152922337870294	59.29372937293729	True
Kazo	66	21111.0	482.0	320.0	43.79875518672199	65.971875	True
Kibale	33	15241.0	298.0	183.0	51.144295302013425	83.28415300546447	True
Kibanda	26	23857.0	358.0	199.0	66.63966480446928	119.88442211055276	True

admin2__ name	schools	pupils	teachers	classrooms	ptr	pcr	reliable
Kiboga	126	46921.0	1130.0	754.0	41.52300 884955752	62.22944 2970822284	True
Kibuku	27	21644.0	336.0	215.0	64.41666 666666667	100.66976 744186047	True
Kigulu	37	22518.0	527.0	280.0	42.72865 275142315	80.42142 857142858	True
Kilak	50	36960.0	523.0	363.0	70.66921 606118547	101.81818 181818181	True
Kinkiizi	60	23727.0	539.0	361.0	44.02040 816326531	65.72576 177285319	True
Kioga	8	8080.0	118.0	73.0	68.47457 627118644	110.68493 150684931	True
Kitagwenda	17	7966.0	155.0	115.0	51.39354 838709677	69.26956 52173913	True
Koboko	21	21401.0	313.0	143.0	68.37380 19169329	149.65734 265734267	True
Kole	11	9634.0	137.0	65.0	70.32116 788321167	148.21538 46153846	True
Kongasis	27	14018.0	318.0	210.0	44.08176 1006289305	66.75238 095238095	True
Kooki	30	12222.0	311.0	165.0	39.29903 536977492	74.07272 727272728	True
Kumi	59	34326.0	628.0	412.0	54.65923 566878981	83.31553 398058253	True
Kwania	14	10588.0	224.0	116.0	47.26785 7142857146	91.27586 206896552	True
Kween	35	19580.0	430.0	296.0	45.53488 372093023	66.14864 864864865	True
Kyadondo	101	35448.0	1113.0	766.0	31.84905 6603773583	46.27676 240208877	True
Kyaka	22	8365.0	145.0	91.0	57.68965 5172413794	91.92307 692307692	True
Kyamuswa	7	982.0	29.0	44.0	33.86206 896551724	22.31818 1818181817	True
Kyotera	36	13818.0	380.0	266.0	36.36315 7894736844	51.94736 842105263	True
Lamwo	17	9957.0	122.0	68.0	81.61475 409836065	146.42647 05882353	True
Luuka	25	16160.0	257.0	166.0	62.87937 743190661	97.34939 759036145	True
Lwemiyaga	15	4864.0	81.0	39.0	60.04938 271604938	124.71794 871794872	True
Madi-Okollo	28	19255.0	319.0	142.0	60.36050 156739812	135.59859 154929578	True
Manjiya	32	19312.0	334.0	212.0	57.82035 928143713	91.09433 962264151	True
Maracha	30	32846.0	574.0	251.0	57.22299 651567944	130.86055 77689243	True
Maruzi	16	9102.0	171.0	97.0	53.22807 0175438596	93.83505 154639175	True
Matheniko	3	829.0	17.0	22.0	48.76470 588235294	37.68181 818181818	False
Mawogola	63	25526.0	491.0	266.0	51.98778 00407332	95.96240 60150376	True
Mawokota	54	22637.0	487.0	382.0	46.48254 620123203	59.25916 230366492	True
Mityana	71	26312.0	667.0	415.0	39.44827 586206897	63.40240 963855422	True
Moroto	4	2181.0	46.0	30.0	47.41304 347826087	72.7	False
Mukono	90	34468.0	888.0	554.0	38.81531 531531532	62.21660 649819495	True
Mwenge	43	20616.0	345.0	244.0	59.75652 1739130434	84.49180 327868852	True
Nakaseke	30	11594.0	256.0	201.0	45.2890625	57.68159 2039801	True
Nakifuma	71	22990.0	610.0	359.0	37.68852 459016394	64.03899 721448468	True

admin2__ name	schools	pupils	teachers	classrooms	ptr	pcr	reliable
Ndorwa	29	14032.0	361.0	242.0	38.86980 609418283	57.98347 107438016	True
Ngora	52	31534.0	607.0	371.0	51.95057 66062603	84.99730 458221025	True
Ntenjeru	101	58688.0	1179.0	779.0	49.77777 777777778	75.33761 232349165	True
Ntoroko	24	9037.0	178.0	110.0	50.76966 292134831	82.15454 545454546	True
Nwoya	56	32857.0	551.0	438.0	59.63157 894736842	75.01598 173515981	True
Nyabushozi	92	22833.0	671.0	387.0	34.02831 5946348734	59.0	True
Obongi	5	1817.0	50.0	22.0	36.34	82.59090 90909091	True
Okoro	12	7060.0	97.0	79.0	72.78350 515463917	89.36708 860759494	True
Omoro	43	31457.0	577.0	412.0	54.51819 757365685	76.35194 174757281	True
Oyam	28	20889.0	355.0	161.0	58.84225 352112676	129.74534 161490683	True
Padyere	25	13768.0	216.0	134.0	63.74074 074074074	102.74626 865671642	True
Pallisa	26	20734.0	375.0	215.0	55.29066 666666667	96.43720 930232558	True
Pian	1	537.0	1.0	0.0	537.0		False
Rubabo	30	13916.0	334.0	249.0	41.66467 065868263	55.88755 020080321	True
Rubanda	30	12785.0	284.0	164.0	45.01760 5633802816	77.95731 707317073	True
Ruhaama	57	26645.0	508.0	357.0	52.45078 74015748	74.63585 434173669	True
Ruhinda	31	14934.0	298.0	201.0	50.11409 395973154	74.29850 746268657	True
Rukiga	23	10072.0	272.0	203.0	37.02941 1764705884	49.61576 354679803	True
Rushenyi	23	12488.0	211.0	168.0	59.18483 412322275	74.33333 333333333	True
Rwampara	71	29236.0	766.0	420.0	38.16710 182767624	69.60952 380952381	True
Samia-Bugwe	79	47672.0	959.0	616.0	49.71011 470281543	77.38961 03896104	True
Serere	11	6110.0	81.0	51.0	75.43209 87654321	119.80392 156862744	True
Sheema	29	11364.0	249.0	210.0	45.63855 421686747	54.11428 5714285714	True
Soroti	42	24768.0	359.0	249.0	68.99164 3454039	99.46987 951807229	True
Terego	29	24976.0	416.0	200.0	60.03846 153846154	124.88	True
Tingey	38	19710.0	478.0	265.0	41.23430 962343096	74.37735 849056604	True
Tororo	51	38018.0	748.0	496.0	50.82620 320855615	76.64919 35483871	True
Usuk	19	6566.0	126.0	57.0	52.11111 1111111114	115.19298 245614036	True
Vurra	16	15126.0	246.0	119.0	61.48780 487804878	127.10924 369747899	True
West Budama	33	19070.0	308.0	178.0	61.91558 441558441	107.13483 146067416	True
West Moyo	15	6290.0	125.0	90.0	50.32	69.88888 888888889	True

G.8 The labour estimates

Quantity	Level	Value	n
Employment rate	Incomplete primary	81.6 %	2,791
Employment rate	Lower secondary	81.4 %	872
Employment rate	None	72.6 %	567
Employment rate	Post-secondary	79.3 %	433
Employment rate	Primary complete	85.5 %	930
Employment rate	Upper secondary	78.1 %	672
Formal, of paid employees	Incomplete primary	2.4 %	432
Formal, of paid employees	Lower secondary	27.3 %	221
Formal, of paid employees	None	2.8 %	44
Formal, of paid employees	Post-secondary	74.9 %	223
Formal, of paid employees	Primary complete	6.8 %	176
Formal, of paid employees	Upper secondary	18.7 %	184
NEET, 15–24	Incomplete primary	26.9 %	810
NEET, 15–24	Lower secondary	25.2 %	214
NEET, 15–24	None	47.1 %	96
NEET, 15–24	Post-secondary	55.5 %	62
NEET, 15–24	Primary complete	25.3 %	249
NEET, 15–24	Upper secondary	40.4 %	196

Appendix H. References and sources

Level	What it is	How to treat it
L1-MODEL	The implemented system	Source, configuration and committed outputs. Everything at this level can be re-read, and every number in this course traces to one of them.
L1-DOC	The system's own documentation	Authoritative for what the system intends. Checked against the code wherever the two could differ.
L2-LIT	Method literature	The published methods the estate implements. Cited so a learner can read the original derivation; never cited to support a number.
L3-DATA	Data sources	The surveys, registers and geographic layers. Access conditions differ and are stated.

THE ORDER MATTERS
A statement in this course is supported by an L1-MODEL source wherever one exists. An L1-DOC source supports intent. An L2-LIT source supports a derivation but never a number. An L3-DATA source is named whenever an estimate is quoted. If you cannot say which level supports a claim, you do not yet have the claim.

L1-MODEL — The implemented system

Source	What it is
EDIS Education Sector Decision Intelligence System (IntegratedEducationSystemModel)	D:/EDIP/models/EducationModelling — 90 Julia modules across 30 layers, 11,183 lines, including the system engine, nine optimisation programmes, the GIS layer and six integration bridges
The committed model runs	6 runs on Uganda and The Gambia, each with its configuration snapshot, run log and metadata
The model configuration	10 configuration files read verbatim: model specifications, scenario definitions, the CGE contract and the household mappings

L1-DOC — The system's own documentation

Source	What it is
The education modelling technical manual	EDIS Education Modelling Technical Manual, EDIS, 2026 — 72 model cards in 8 families across 30 chapters, with 78 displayed equations
The estate's own documentation	40 documents including the mathematical specification, the data usage audit, the CGE exchange contract, the cross-sector schema and the baseline test status

L2-LIT — Method literature

Source	What it is
Mincer earnings function	The log-earnings specification in schooling and experience that Family E estimates
Heckman two-step selection correction	The selection model Family E runs with and without an exclusion restriction
Oaxaca–Blinder decomposition	The endowment/coefficient split used on the gender gap
Concentration index and concentration curve	The rank-weighted inequality measure used across Families A, B and C
Two-part and selection models of expenditure	The participation-then-level structure of Family B
Taylor linearisation for complex survey designs	The variance estimator behind every interval in this course

L3-DATA — Data sources

Source	What it is
Uganda 2016 DHS (UG2016DHS)	Household-member recode: 91,167 person records, 696 clusters, 34 strata. Restricted; not distributed with this course
Kenya 2022 DHS (KE2022DHS)	Household-member recode: 156,571 person records, 1,691 clusters, 92 strata. Restricted; not distributed with this course
Uganda LSMS-ISA National Panel Survey 2019/20 (UNPS)	3,098 households and 16,056 persons with an education module, a labour module and a consumption aggregate. Restricted; not distributed with this course
Uganda school register (EMIS), via the EDIS spatial data infrastructure	16,884 primary and 3,594 secondary schools with pupils, teachers, classrooms and ownership
EDIS Africa education GIS database (AEGDB)	46,555 matched schools, 81,746 settlements, 80,160 settlement-to-school distances and geoBoundaries admin areas
The DHS library	215 household-member recodes across 40 countries
The LSMS library	81 survey rounds across 16 countries
EDIS spatial data infrastructure	D:/EDIP/SDInfrastracture — school registers by level, administrative boundaries, population grids and derived accessibility surfaces
OpenStreetMap	School and settlement points, under the Open Database Licence. Attribution required
geoBoundaries	Administrative boundaries, CC BY 4.0. Attribution required

H.5 Access conditions

Source	Condition
Survey microdata	Held by the collecting agency under registered-user terms. **Not distributed with this course and not required by it.**
Survey cluster coordinates	Displaced as a disclosure control by the survey programme. Never used or published here in any form.
The model estate	Internal. Source, configuration and committed outputs.
OpenStreetMap	Open Database Licence. Attribution required and given on every map.
geoBoundaries	CC BY 4.0. Attribution required.
School register	National administrative data, accessed through the spatial data infrastructure.

Appendix I. The limitations register

16 limitations, each with the evidence that established it and what this course does about it. Five were found by re-deriving a shipped figure rather than by reading documentation, and those are the ones worth studying: they are the kind you will have to find yourself.

ID	What it is	Evidence	How the course handles it
EG-01	The catalogue is 72 specifications; the data decide how many can be estimated	72 model cards in 8 families. This course estimates models across six families on real microdata and estimates none of the ten in family D.	Chapter 5 reads the catalogue before any estimation, and every domain chapter states which of its models this course estimates and which it only specifies.
EG-02	No pupil-level learning assessment data exists on any platform this course reaches	The model's country bundles carry one aggregate learning row per country-region-year and no pupil records; no assessment microdata is present in the education GIS platform, the DHS cache or the LSMS library.	Family D is taught as specification. The chapter shows that every right-hand-side variable is available and the left-hand side is not, and refuses to substitute a proxy.
EG-03	The attendance flag means 'attended during the current school year', not 'currently attending'	The household-member recode defines its attendance variable as 0 no, 1 currently attending, 2 attended at some time. Both surveys this course uses record only 0 and 2, so the code that would carry 'currently attending' never appears.	Every participation estimate in this course is labelled with what the recode supports, and Laboratory 02 has learners verify the coding themselves before using it.
EG-04	The school register locates its schools badly	64 % of records carry a coordinate inside the country, and 53 % of pupil-bearing schools reach a county. The pupil-teacher ratio on the mappable schools is 48.2 against 49.8 on the full register.	Every map built from the register states the share mapped and the bias that dropping the rest introduces. Counts by level and ownership are used freely; per-school geography is not.
EG-05	A quarter of the register's schools record no pupils	4,126 of 16,884 primary schools have a pupil count of zero, which is why the register's enrolment total is far below the model run's.	The reconciliation chapter puts the two side by side and explains the difference rather than choosing between them.
EG-06	The estate's estimation engine does not cluster its standard errors	Its own documentation says point estimates are weighted and standard errors are model-based, with design-based clustering a follow-up. This course computes both and the design correction widens intervals materially.	Chapter 4 derives the correction, Laboratory 02 computes it, and every interval quoted in this course is design-based.
EG-07	The macro exchange leaves three contract fields empty	Projecting the committed run onto the exchange columns fills 7 of 10; learning score, completers and skilled workers are empty, and the human-capital index is the run's gross enrolment ratio.	The integration chapter shows the contract against the payload and requires a handover note that names every empty field.
EG-08	The CGE coupling cannot be executed on this host	The estate documents that the operational coupling runs an external process whose native stack is blocked by this machine's Application Control policy. The contract and the export are valid; the solve is not run.	The course teaches the contract and the handover, and never presents a coupled result.

ID	What it is	Evidence	How the course handles it
EG-09	The wage sample is one working-age person in seven	14 % of the working-age sample has a usable pay rate; 52 % of workers are on a household farm and have no wage at all.	Family E reports the return as a return within paid employment, with the selection share on the same slide.
EG-10	A selection correction without an exclusion restriction is not a correction	Run with the same regressors in both equations, the two-step correction raises the estimated return from 11.8 % to 34.3 % with a decisive-looking Mills coefficient. With household size and children under five excluded from the wage equation the correction disappears.	Both are shown side by side. This is the course's worked example of a result that is impressive because it is unidentified.
EG-11	Four sources describing the same country do not reconcile	Primary participation, the pupil-teacher ratio and primary enrolment each differ between the survey, the register and the committed run, by definition, by year or by coverage.	The reconciliation table is shipped as data and worked through in Laboratory 08. No source is declared correct.
EG-12	A shipped figure is labelled per student and is per household	The estate's data audit reports household education spending of about one million shillings per student per year. Re-deriving from the same file gives 1.03 m UGX per household with students and 375k UGX per student; the mean household has 2.69 students.	Taught in Chapter 6 as the course's denominator case. The arithmetic is right and the label is wrong, which is the failure mode hardest to catch.
EG-13	The accessibility surface is straight-line, not routed	Every settlement-to-school distance this course reads is measured as the crow flies. The estate implements a routable graph over OpenStreetMap roads and the spatial infrastructure holds travel surfaces, but the shipped table is straight-line.	Every accessibility figure says so, and Laboratory 06 asks what a river between the settlement and the school would do to it.
EG-14	The settlement population field sums well below the national total	The settlement register carries a population for 80,160 settlements summing to 5,570,285, against a national population several times larger.	Population-weighted coverage is labelled as a share of the settlement population that carries a value, never of the country.
EG-15	The estate's own cross-sector link finds essentially nothing	Its education-to-NEET association across fifty countries returns a slope of -0.002 and an R^2 of about zero, and the estate reports it as null rather than suppressing it.	Taught as the ecological-fallacy case and as the reason the individual-level labour estimates in Family E exist.
EG-16	The continent-scale household harmonisation holds no microdata yet	The catalogue covers 55 countries and maps education variables for 37, and its downloaded-microdata count is zero. A loader is ready and reports readiness rather than fabricating.	The data chapter distinguishes catalogued from ingested, and the course runs on the platforms that do hold microdata.

I.2 The five found by re-derivation

ID	What re-deriving it showed
EG-03	Reading the recode's own code list showed the attendance variable never takes the value that means 'currently attending' in either survey.
EG-04	Joining the school register to administrative polygons showed that only about half of pupil-bearing schools can be placed, and that the placeable half has a different pupil-teacher ratio.
EG-10	Running the selection correction twice showed that the version without an exclusion restriction triples the estimated return and reports a decisive t-statistic.
EG-11	Putting four sources for the same country side by side showed that none of the three substantive pairs reconciles.
EG-12	Recomputing a shipped audit figure from the same file showed it is a per-household quantity carrying a per-student label.

I.3 The automatic fails

Any one of these in an assessed deliverable fails that deliverable, whatever else it contains. Each corresponds to a limitation above or to an interpretation rule in chapter 29.

16. Quoting an estimate without its interval, its denominator and its disclosure status.
17. Quoting a per-household figure under a per-student label, or the reverse.
18. Describing an adjusted association as an effect, or presenting a microsimulation counterfactual as a treatment effect.
19. Reporting a capacity gap without the norm class size that generated it, or an optimisation result without its objective.
20. Publishing a map built on the school register without stating the share of schools it could place.
21. Presenting a model this course specifies but does not estimate as though it had been estimated.
22. Reporting any figure from a survey cluster coordinate, or distributing microdata of any kind.

Appendix J. The estate's documentation, and what it verifies

The estate ships 40 documentation files. Most of them end with a section headed *Verified*, in which the authors state what they ran and what came out — including, repeatedly, results that did not support the thing being built. This appendix collects those statements, because they are the estate's own account of what it can do and they are not repeated anywhere else in this manual.

WHY THIS APPENDIX EXISTS
A model estate that publishes only its successes cannot be audited. Several of the blocks below record a null result, an uncalibrated module or a blocked dependency. Those are the entries worth reading first: they tell you where the estate believes its own weight is thin, and they are where a new analyst should look before promising anything.

J. End-to-End Multi-Country Acceptance Run

File	‘docs/ACCEPTANCE.md‘
Command	33
Length	54 lines

real panel

Country	Status	Final GER	Primary PTR
Gambia (GMB)	✓ completed	0.99	32.8
Uganda (UGA)	✓ completed	1.06	42.7
Nowhereland (XXX)	⏩ skipped	—	—

Skip note (honest): ‘ArgumentError: no DHS PR dataset found for 'Nowhereland'‘.

Two distinct real countries — one small (Gambia, 49k–56k DHS persons) and one large (Uganda, ~51 M population, 8.8 M primary enrolment) — pass the full ingest → harmonize → demography → teachers → scenario → outputs chain deterministically, each producing 9 fingerprinted artifacts. The default ‘ACCEPTANCE_COUNTRIES‘ panel offers six candidates (GMB, KEN, GHA, SEN, RWA, MWI) for a broader run.

CAVEATS, IN THE ESTATE'S OWN WORDS
Each country runs the full DHS harmonization (seconds to a minute) — a large panel takes several minutes; run a subset for quick acceptance. Sanity bounds are deliberately wide (catch gross errors, not fine calibration). Countries lacking a DHS PR dataset or World Bank series skip by design — read the notes to see coverage, and cross-check with the Command-31 data-quality dashboard. The scenario is the baseline; pass a ‘Scenario‘ for policy acceptance runs.

J. School Accessibility & Catchment Areas

File	‘docs/ACCESSIBILITY.md‘
Command	14
Length	64 lines

real — The Gambia, GMB

1,586 OSM settlements against 624 OSM schools:

Metric	Value
Coverage < 2 km	0.332
Coverage < 5 km	0.658
Coverage < 10 km	0.948
Median nearest-school distance	3.49 km
Mean nearest-school distance	4.05 km
Settlements > 5 km from a school (underserved)	543 (34.2%)
Schools serving ≥1 settlement	319 (busiest catchment: 42 settlements)

Coverage is monotone across thresholds; catchment loads sum to the assigned settlements. Gate: **18/18 pass**.

CAVEATS, IN THE ESTATE'S OWN WORDS
Straight-line distance understates real travel where terrain/rivers/roads intervene — coverage figures are optimistic upper bounds until a road network is added. Settlement weights use OSM place points (no population attached in Educationdata); population-weighted coverage becomes available once settlement populations are joined (WorldPop grids, or a settlement gazetteer). School capacity is not yet modelled — catchment demand here is settlement count, not enrolment vs places (that is the teacher/infrastructure commands).

J. Macro / CDCGE Exchange Schema

File	‘docs/CGE_EXCHANGE.md‘
Command	23
Length	58 lines

real — The Gambia, GMB

Real canonical outputs assembled into a payload: a Command-17 primary enrolment forecast, Command-16 teacher counts (enrolment/PTR), Command-18 recurrent spending and measured population — validated, JSON round-tripped, and projected onto the legacy CDCGE columns. Gate: **23/23 pass**.

CAVEATS, IN THE ESTATE'S OWN WORDS
'human_capital_index' here is an enrolment-based **proxy** (documented) — the CGE may substitute its own human-capital measure. 'skilled_workers' needs a real labour-supply mapping (Command-21 graduate output + an LFS skill definition), not a fixed share. The operational coupling ('run_cdcge_coupling') runs the external CDCGE process, whose native stack is blocked under this machine's Application-Control policy — the **contract and export are valid**; execute the solve on a host where CDCGE's dependencies load. Any change to field names/units is a **breaking version bump** (major) by rule 10.

J. Education Costing & Unit Costs

File	'docs/COSTING.md'
Command	18
Length	66 lines

real — The Gambia, GMB, 2025

Quantity	Value
GDP per capita	$919
Measured primary unit cost ('SE.XPD.PRIM.PC.ZS')	$52/pupil
Modelled unit cost (salary 4× GDP p.c., non-salary 5%)	$158/pupil
Modelled / measured ratio	3.01
Salary share of recurrent	0.71
Education envelope ('SE.XPD.TOTL.GD.ZS')	$66.9 M
Modelled primary recurrent	$70.0 M
Affordability (recurrent / envelope)	1.05

The modelled cost lands within a factor of 3 of the measured benchmark — and the gap is **informative**: matching the measured $52/pupil implies an effective salary of ≈1.3× GDP per capita, not the illustrative 4× assumed. This is exactly how the tool is meant to be used — **calibrate the assumed salary multiple to the measured unit cost**, then project. Gate: **16/16 pass**.

CAVEATS, IN THE ESTATE'S OWN WORDS
Salary multiple / non-salary rate / classroom cost are assumptions — calibrate to the measured benchmark before forecasting (as above). Capital cost is classroom-only (no land, furniture, WASH, or rehabilitation of the poor-condition schools flagged in the EMIS 'condition_score'). One national average per level hides subnational and remote-school cost premia. Costs are nominal in the anchor year; no inflation/discounting (that is the CBA/MTEF command).

J. Cost-Effectiveness & Cost-Benefit Analysis

File	'docs/COST_BENEFIT.md'
Command	20
Length	59 lines

real — The Gambia, GMB

One extra year of schooling, measured anchors (GDP p.c. **$919**, unit cost **$52/pupil-year**), Mincer return **10%** (assumption), 40-year career after a 6-year delay, 5% discount:

	Value
Annual earnings gain	~$97
Benefit-cost ratio	24.8
Internal rate of return	0.375
Net present value	~$1,247

The very high BCR reflects a **low direct unit cost** ($52) against a lifetime earnings gain — consistent with the well-known finding that basic schooling in low-income settings has high measured private returns. Gate: **17/17 pass**.

CAVEATS, IN THE ESTATE'S OWN WORDS
Returns are **associational** (Mincer), not causal — the coefficient is an assumption to calibrate to microdata (a Mincer regression on LSMS earnings is the estimation upgrade, and would compose with the Command-09 attainment models). Costs here are direct public unit cost only (no opportunity cost of the pupil's time, no private costs); including foregone earnings lowers the BCR. Constant earnings profile (no experience curve, unemployment or mortality). CEA effects must be commensurable within a league table (don't mix enrolment-gain and learning-gain effects in one ranking).

J. Cross-Sector Linkages (Education ↔ Health / Labour)

File	‘docs/CROSS_SECTOR.md‘
Command	24
Length	66 lines

real — education → labour, World Bank, 50 countries

Cross-country ecological association of education (upper-secondary attainment, adult literacy fallback) against **NEET** (‘SL.UEM.NEET.ZS‘):

	Value
Countries (n)	50
Slope	−0.002
Correlation	−0.004
R^2	~0.000

The association is **essentially null at the ecological level** — reported honestly. Two reasons make this the *expected* result, not a failure: the x-variable mixes attainment and literacy proxies across countries (inconsistent construct), and national NEET is multi-causal. As with the learning models (Command 11), this reinforces that **individual-level microdata** (a woman's education vs her child's health in DHS; a person's schooling vs their employment in an LFS/LSMS) is required for a credible linkage — the schema and gradient estimator are ready to carry exactly those estimates. Gate: **18/18 pass**.

CAVEATS, IN THE ESTATE'S OWN WORDS
Ecological correlations do not imply individual-level relationships (ecological fallacy) — they are labelled ‘ecological‘ for that reason. The health linkage (mother's education → child anthropometry/mortality) uses the same schema but needs the DHS child/woman recodes (‘KR‘/‘IR‘) beyond the household roster used elsewhere. Mixing indicator constructs in one regressor (attainment/literacy) attenuates associations — align constructs before headline use. All linkages are associational; none may be relabelled causal without an identification strategy.

J. Diagnostics & Data-Quality Dashboard

File	‘docs/DATA_QUALITY.md‘
Command	31
Length	51 lines

real snapshot

	Value
Educationdata raw sources	16

Sources with 54-country coverage	geoBoundaries, OSM schools/settlements, UIS, World Bank, WorldPop age, learning, GIGA, MoE register, HDX (9)
DHS holdings	44 countries · 3,835 dataset files
LSMS holdings	1 country (Uganda)
Data-poor sources (≤1 country)	hydrosheds, naturalearth, worldclim, jmp__wash (stubs)
GIS gaps flagged	roads **absent**, population raster **absent**, hazards **absent**

The dashboard makes the honesty audit **checkable**: the system is data-rich for boundaries, schools, settlements and national education/population indicators across 54 countries, and **data-poor** for roads, population rasters, climate hazards and multi-country LSMS — exactly the gaps flagged throughout the model docs.

CAVEATS, IN THE ESTATE'S OWN WORDS

Country counts use ISO3-named folders; sources organized differently (‘AFRICA‘, single archives) report 0 countries — a coverage signal, not an error. Availability checks presence, not internal completeness (a present source may still have sparse years/fields — use ‘field__missingness‘ on the loaded frame). DHS country count is by dataset filename prefix; LSMS count is country subdirectories under the configured root.

J. Population & Demography — School-Age Cohorts

File	‘docs/DEMOGRAPHY.md‘
Command	13
Length	63 lines

real — The Gambia, GMB

DHS roster of **55,640 persons** scaled to World Bank total **2,822,093** (latest):

Quantity	Value
Total population (SP.POP.TOTL)	2,822,093
Population 0–14 (SP.POP.0014.TO)	1,121,027
School-age 3–24	1,521,173 (share **0.539**)
Primary-age 6–11	505,796
Age-heaping index (ages 1–24)	0.86 (mild — reported, not smoothed)

The 0–14 band reconciles exactly to the measured World Bank figure. Gate: **18/18 pass**.

CAVEATS, IN THE ESTATE'S OWN WORDS

The pyramid *shape* is the DHS survey year; the *level* is the World Bank target year — a level reconciliation, not a full cohort-component projection (that is the forecasting command). Age heaping (index $\neq 1$) reflects real roster reporting; enable smoothing only deliberately and flag it. Subnational cohorts require DHS region weights joined to boundaries (Command 12) — supported by the roster's ‘region‘ column when present.

J. Dropout & Progression Models

File	‘docs/DROPOUT__PROGRESSION.md‘
Command	08
Length	72 lines

on real data (DHS Gambia 2019

* **Binary dropout model:** n = 14,184; wealth-quintile coefficient **−0.263** (richer children less likely to have dropped out); mean predicted risk 0.46 (cross-sectional prevalence among ever-attended school-age children — high because many are not *currently* attending; reported as-is, not smoothed).
* **Grade-flow Markov** from Gambia administrative rates (promotion 0.90, repetition 0.05, dropout 0.05): an 8-state chain that **passes the consistency gate** (all rows

sum to 1).

CAVEATS, IN THE ESTATE'S OWN WORDS

* Binary dropout is prevalence, not incidence — the hazard is the right tool once panel/retrospective histories are available.
* Associational, not causal (Command 11 note). SEs weight- not cluster-adjusted.
* State assignment from DHS levels is a heuristic mapping; the transition machinery itself is exact and validated.

J. Econometric Models

File	‘docs/ECONOMETRIC_MODELS.md‘
Command	07 (attendance & enrolment)
Length	85 lines

on real data (DHS Gambia 2019, weighted

‘fit_attendance(harmonize_dhs(cfg; country=”Gambia”).persons)‘:

* n ≈ school-age members with complete covariates;
* **wealth quintile AME ≈ +0.057** — each step up the wealth ladder raises the probability of attending by ~5.7 pp (poorer children less likely in school);
* **urban coefficient ≈ +0.10** (urban children more likely attending);
* **McFadden R^2 ≈ 0.061**.

Real, survey-weighted estimates from the harmonized microdata — not placeholders.

CAVEATS, IN THE ESTATE'S OWN WORDS

* SEs are weight-adjusted, **not** cluster-adjusted (needs DHS PSU/strata on ‘persons‘).
* These are **descriptive/associational** models — no causal claim without an explicit identification strategy (Command 11 note).
* ‘save_model‘ serializes parameters (portable CSV/JSON); the live GLM object can also be ‘Serialization.serialize‘d if a binary reload is needed.

J. Enrolment / Cohort Forecasting

File	‘docs/ENROLMENT_FORECAST.md‘
Command	17
Length	59 lines

real — The Gambia, GMB

Measured base primary enrolment **442,901**, measured entry-age (6) cohort **91,705**, primary-age (6–11) population **505,796**; scenario rates π=0.88, ρ=0.08 (δ=0.04) over 5 years:

Year	0	1	2	3	4	5
GER	0.88	0.89	0.92	0.94	0.97	1.01

5-year cumulative completers ≈ 299,829; dropouts ≈ 93,049 (at the assumed rates). GER rises toward ~1.0 as the (large, measured) entry cohort feeds the system. Gate: **41/41 pass**.

CAVEATS, IN THE ESTATE'S OWN WORDS
Transition rates here are **assumptions**, not measured flows — credible projections need grade-specific promotion/repetition/dropout from an EMIS flow file (or estimated via the Command-08 progression model on longitudinal data). Base enrolment is spread evenly across grades for the demo; a real grade-structure vector should replace that. Intake is held constant; linking intake to the projected single-year population path (cohort-component) is the demographic-projection extension. No mortality/migration in the school-age cohort (short horizons).

J. Education Equity & Deprivation

File	‘docs/EQUITY_AND_DEPRIVATION.md‘
Command	10
Length	62 lines

Verified

Hand-calculated (the gate): ‘gini([1,2,3,4]) = 0.25‘, ‘gini([0,0,0,4]) = 0.75‘; ‘concentration_index‘ = −0.75 (outcome among poorest) / +0.75 (richest); ‘EDI‘ for a 2-area, 2-component example = 0.28 / 0.24 exactly, and identical on re-run.

Real DHS Gambia 2019 (weighted):
* primary-age attendance rises from **66.1% (poorest quintile) → 86.2% (richest)** — a 20 pp rich–poor gap; attendance **concentration index +0.053** (pro-rich);
* **attainment Gini = 0.74** — very unequal years of schooling (a large share with none).

Real region composite (Gambia country data): an EDI over teacher-shortage, overcrowding and classroom-gap components (min-max normalized) is computed per region and is reproducible on re-run — illustrating the index mechanics on measured sub-national data (component choice/weights are configurable and should be set with national data before policy use).

CAVEATS, IN THE ESTATE'S OWN WORDS
Concentration/Gini SEs are not reported (point estimates); disability gaps require a disability variable (available in some DHS/LSMS modules — flagged, added when mapped). EDI component selection and weights are policy choices — set and sensitivity-test them with national stakeholders; the defaults here are illustrative.

J. GIS / Spatial Foundation

File	‘docs/GIS_MANUAL.md‘
Command	12
Length	70 lines

Verified

Synthetic: Web-Mercator round-trip, geodesic 1°≈111 km, PIP in/out, spatial join, grid ‘nearest‘/‘within_radius‘.

Real (The Gambia, GMB): 10 ADM1 regions and 624 OSM schools loaded from Educationdata; the spatial join assigns **all 624 schools** to a region by point-in-polygon; every boundary vertex reprojects to finite Web-Mercator metres.

Gate (CRS conversion + PIP + spatial join): **17/17 pass**.

J. Household Survey Harmonization

File	‘docs/HOUSEHOLD_HARMONIZATION.md‘
Command	04

Length	100 lines

on real data

‘harmonize_dhs(load_config(); country=”Gambia”)‘ →

* survey **GM2019DHS** (‘GMPR81DT.ZIP‘, latest phase), **55,640 persons**, **6,549 households**;
* canonical output **validates** against the ‘persons‘ schema (0 errors);
* unavailable flagged: literacy, parental education, expenditure, consumption, poverty, occupation, wages;
* **survey-weighted primary-age (6–11) attendance = 0.794** — a real, weighted indicator computed directly from the harmonized microdata (not a placeholder).

J. Education Indicator Engine

File	‘docs/INDICATOR_ENGINE.md‘
Command	06
Length	80 lines

on real data (DHS Gambia 2019, weighted

* **Primary attendance (6–11)** = **0.794** — SE 0.0051, 95% CI 0.784–0.804, n = 10,426, n_eff = 6,317;
* **Mean years of schooling (25+)** = **7.88** (n = 20,457);
* 8 DHS-region estimates via ‘indicators__by__group‘.

These come straight from the harmonized microdata with survey weights — real, not placeholder, values with design statistics attached.

J. Learning Outcome & Education Production Models

File	‘docs/LEARNING_MODELS.md‘
Command	11
Length	71 lines

Verified

Synthetic student assessment (25 schools × 40 students, known DGP):
* OLS/FE recover the within-school student slope (≈ 3.0);
* multilevel variance components — between-school 4.12, within 16.08, **ICC ≈ 0.20**;
* ‘school_performance‘ returns only ‘school, school_effect, n‘ (no student columns);
* the causal gate holds (associational only; ‘iv‘/‘did‘/‘rct‘ refused).

Real cross-country HLO (54 countries, ecological): the production OLS *runs* on the real learning tables but is **uninformative ($R^2 \approx 0.008$)** — the national mean-score proxies (with inconsistent scales/definitions across the country builds) do not support an input–outcome model. This is reported honestly and reinforces the core limitation:

> **Real student-level assessment microdata** (e.g. PASEC, SACMEQ, EGRA, national exams)
> is required for credible learning-production estimates. The models here are built,
> tested and ready; they should be run on such microdata (student nested in school,
> with teacher/infrastructure covariates) rather than on aggregate proxies.

CAVEATS, IN THE ESTATE'S OWN WORDS
Associational, not causal (by construction). SEs are model-based, not cluster/design-adjusted. Quantile regression was not added (no reliable dependency adopted under the policy) — deferred with the mixed-models upgrade.

J. Prioritization — Multi-Criteria Decision Analysis

File	‘docs/MCDA.md‘
Command	26
Length	53 lines

real — 51 African countries, World Bank indicators

Composite performance score on primary GER (benefit, 0.4), adult literacy (benefit, 0.4) and pupil-teacher ratio (cost, 0.2):

	Ranking
Weighted-sum, top 5	MUS, SWZ, NAM, SYC, TUN
Weighted-sum, bottom 5	SSD, NER, CAF, SOM, TCD
TOPSIS, top 5	NAM, SWZ, MUS, RWA, STP
Top-rank stability (±0.1 weight)	**1.0**

The rankings are **face-valid** (high-performing middle-income economies at the top, conflict/low-income states at the bottom), the two methods agree on the leading cluster and differ only at the margin (weighted-sum leads MUS, TOPSIS leads NAM), and the top rank is fully stable to weight perturbation. Gate: **19/19 pass**.

CAVEATS, IN THE ESTATE'S OWN WORDS

Min-max normalization is sensitive to outliers (one extreme unit compresses the rest) — consider rank or z-score normalization for heavy-tailed criteria. Weights encode value judgements: always report the sensitivity, and prefer criteria that are directly decision-relevant (coverage gap, equity, cost-effectiveness from Commands 14/10/20) over convenient proxies. TOPSIS and weighted-sum can disagree on close units — present both. Prioritization ranks *given* the criteria; it does not choose the criteria or their weights.

J. Household Policy Microsimulation

File	‘docs/MICROSIMULATION.md‘
Command	22
Length	60 lines

real — The Gambia, DHS, n = 22,626

Policy shock: **upward wealth mobility (+1 quintile)**, ever-enrolment logit on sex/urban-rural/wealth (weighted). National ever-enrolment **80.8% → 84.9%** (+4.1pp; wealth coefficient +0.30). Incidence by baseline quintile:

Quintile	Baseline	Counterfactual	Change
Q1 (poorest)	0.687	0.747	**+6.0pp**
Q2	0.764	0.814	+5.0pp
Q3	0.823	0.862	+3.9pp
Q4	0.868	0.898	+3.0pp
Q5 (richest)	0.899	0.923	+2.4pp

The impact is **strongly pro-poor** — the largest enrolment gains accrue to the poorest quintile — which is exactly the incidence signal a targeting decision needs. Gate: **12/12 pass**.

CAVEATS, IN THE ESTATE'S OWN WORDS

Counterfactuals are model-based and **associational**, not causal — magnitudes inherit the logit's specification and the covariate set (here sex/urban-rural/wealth only; richer covariates change the numbers). The data must contain every model covariate (derived terms like ‘age^2‘ must be rebuilt before simulating). A ”+1 quintile” shock is illustrative economic mobility, not a costed programme; linking a specific policy (fee abolition, CCT) to its covariate change and cost (Commands 18–20) is the scenario command. Expenditure incidence via the Command-09 two-part model composes the same way.

J. MTEF — Medium-Term Expenditure Framework

File	‘docs/MTEF.md‘
Command	19
Length	57 lines

real — The Gambia, GMB

Measured base GDP with the current education share of **2.81% of GDP**, projected at 5% GDP growth:

	Value
Base-year envelope	$72.8 M
Year-5 envelope	$93.0 M
NPV of envelope (6 yr, r=5%)	$437 M
NPV of a cost path 10% above base envelope	$427 M
NPV financing gap	−$10 M (envelope keeps pace as GDP grows)

A flat cost 10% above the *base* envelope is initially underfunded, but 5% envelope growth closes it over the horizon (negative NPV gap). Gate: **22/22 pass**.

CAVEATS, IN THE ESTATE'S OWN WORDS

GDP growth and share targets are assumptions — vary them for scenario envelopes; the tool does not forecast GDP. Financing gap here compares one cost path to one envelope; linking it to the Command-17/18 cost projection under alternative access/quality scenarios is the scenario command. Sub-sector allocation is a single-period LP (no inter-year carryover or capital/recurrent split). Nominal flows discounted at a real rate should be made consistent (deflate or use a nominal discount rate) before headline use.

J. Direct Link — EducationModelling ↔ MTEFModelling (FINEX)

File	‘docs/MTEF_LINK.md‘
Length	79 lines

real — The Gambia, baseline scenario

	Value
MTEF scenarios available	17
Recurrent envelope (base yr)	4.17×10^{12}
Capital envelope (base yr)	6.50×10^{11}
MTEF-funded enrolment places (base yr)	105,161
Education school-age demand (DHS × WB)	505,796
Funded coverage	**20.8%** — the financing gap the two models jointly reveal

CAVEATS, IN THE ESTATE'S OWN WORDS

The link consumes whatever the MTEF has published — run the MTEF education bridge first (‘MTEFModelling/edis_mtef_bridge‘) to refresh scenarios. Funded-vs-demand coverage uses the MTEF's own unit-cost translation; feeding education's measured unit costs (Command 18) back via ‘export_education_to_mtef‘ lets the MTEF replace its generic costs with measured ones. A schema/units change on either side is a breaking version bump (rule 10).

J. Facility-Location Optimization (School Siting)

File	‘docs/OPTIMIZATION.md‘
Command	15
Length	66 lines

real — The Gambia, GMB subset

100 settlement demand points, 40 existing schools pinned open, 20 candidate new sites, 5 km coverage radius, budget = 45 (add up to 5):

	Value
Solver status	‘OPTIMAL‘
Status-quo coverage (40 existing)	0.360
Optimized coverage (+ new sites)	0.400
New sites opened	4
Additional demand covered	+4 points

The optimizer weakly improved coverage (0.36 → 0.40) within budget. Gate: **20/20 pass**.

CAVEATS, IN THE ESTATE'S OWN WORDS
Geodesic (not routed) distances — coverage gains are upper bounds until a road network is added. School **capacity** and construction **cost heterogeneity** are not yet in the objective (equal-cost sites, uncapacitated coverage); capacitated/cost-weighted siting composes with the costing command. Large instances (national, all settlements × all candidates) should use a coverage radius to sparsify ‘A‘, a ‘time_limit‘, or candidate pre-filtering — the p-median ‘x_{ij}‘ matrix in particular grows with demand×candidates.

J. End-to-End Pipeline Orchestration

File	‘docs/PIPELINE.md‘
Command	27
Length	60 lines

real — The Gambia, GMB

“‘
run_id : edu_GMB_20260101T000000_664b64f1
stages : harmonize → demography → teacher_demand → scenario → outputs
files : cohorts.csv, config_snapshot.toml, costs.csv, enrolment.csv, finance.csv, run.log, run_metadata.json, summary.json, teachers.csv
result : persons 55,640 · base enrolment 442,901 · primary PTR 32.8 · final GER 0.99 · cumulative completers 283,879
“‘

One command turns raw DHS + World Bank inputs into a fingerprinted, logged, reloadable set of planning outputs. Gate: **18/18 pass**.

CAVEATS, IN THE ESTATE'S OWN WORDS
The default run uses the primary sub-sector and a baseline scenario; pass a ‘Scenario‘ (Command 25) for policy runs, and extend the driver with secondary/TVET/tertiary stages as needed. Run directories accumulate under ‘outputs/runs/‘ — prune old runs in batch use. Inherits every stage's caveats (assumed transition rates/cost drivers; national base spread across grades). Reproducibility covers the deterministic modelling path; upstream data edits change the config fingerprint (by design) and therefore the ‘run_id‘.

J. TVET & Tertiary Sub-Sector Models

File	‘docs/POST_BASIC.md‘
Command	21
Length	51 lines

real — The Gambia, GMB

Measured tertiary enrolment **5,001**. With an *illustrative* 15% secondary-to-tertiary share and 4-year duration, the derived tertiary stock is **36,978** — a ratio of **7.4**.

This overshoot is **informative**: Gambia's tertiary GER is very low, so the realistic transition share is ~2%, not 15%. The tool exposes exactly this calibration — **fit the onward-transition shares to the measured tertiary stock** before projecting TVET/tertiary scenarios. Gate: **19/19 pass**.

CAVEATS, IN THE ESTATE'S OWN WORDS
Transition shares, durations and survival are assumptions to calibrate against measured tertiary/TVET stocks (as above). No field-of-study detail in local data, so ‘skills_balance‘ and ‘absorption‘ need a programme×field mapping and labour-demand estimates (from an LFS or employer survey) to be operational. Private/religious tertiary providers may be outside the World Bank series. TVET has no dedicated local count — its outputs here are entirely assumption-driven and must be flagged as such in any report.

J. Reporting Layer

File	‘docs/REPORTING.md‘
Command	30
Length	56 lines

Verified

A full planning report for The Gambia is produced in one call, stamped with its run id and scenario fingerprint, and saved alongside the pipeline's CSV/JSON artifacts — a decision-ready document whose every number is traceable to a measured source or a flagged assumption.

CAVEATS, IN THE ESTATE'S OWN WORDS
Markdown output; render to PDF/HTML with any converter for distribution (self-contained, no external assets). One scenario per report — call again with a different ‘Scenario‘ for comparisons, or extend the module to embed a baseline-vs-scenario delta table. The report summarizes the primary sub-sector path; secondary/TVET/tertiary sections follow the same pattern once those stages are added to the pipeline.

J. Scenario Engine

File	‘docs/SCENARIO_ENGINE.md‘
Command	25
Length	64 lines

real — The Gambia, GMB

Baseline vs an ”expand access” scenario (50% dropout reduction, 6% GDP growth) on measured base-year inputs (primary enrolment, entry cohort, primary-age population, GDP):

Metric	Baseline	Reform	Δ
GER, year 0	0.88	0.88	—
GER, year 5	0.99	1.06	+0.07
Cumulative completers	283,879	307,200	**+23,321**
Extra recurrent cost	—	—	+$16.5 M
Extra NPV financing gap	—	—	+$17.0 M

The reform lifts completers and GER at a quantified extra cost and financing gap — the full policy trade-off in one run, with distinct fingerprints (‘a4030be7…‘ vs ‘8c6bcd92…‘). Gate: **16/16 pass**.

CAVEATS, IN THE ESTATE'S OWN WORDS
Scenarios inherit every upstream caveat (assumed transition rates/cost drivers; national base spread evenly across grades; single-PTR teacher model). Levers are uniform across grades/regions here — subnational scenarios need region-indexed base inputs. The engine composes the deterministic planning models; stochastic/uncertainty bands (fan charts) are a later extension. CBA (Command 20) can be layered on a scenario's cost/benefit streams for NPV/BCR of the reform.

J. EMIS / School Data Harmonization

File	‘docs/SCHOOL_HARMONIZATION.md‘
Command	05
Length	80 lines

on real data

Country	source	input	harmonized	dupes removed	invalid coords	with region	with level	schema
Gambia	EDIP export	599	599	0	0	599	175	✓
Kenya	full OSM (Educationdata)	15,243	15,194	49	0	15,194 (ADM1 PIP)	9,355	✓

Kenya demonstrates using the **full** Educationdata register (15k schools) rather than the 600-sample, with every school assigned to its ADM1 county by point-in-polygon.

J. API / Services Layer

File	‘docs/SERVICES.md‘
Command	28
Length	61 lines

real — The Gambia

‘svc_scenario(”GMB”, ”Gambia”, {dropout_reduction: 0.5})‘ returns a serialized envelope with the scenario summary (final GER, completers, NPV cost), the enrolment and finance paths, and the run fingerprint. ‘svc_country_summary‘, ‘svc_accessibility‘ and ‘svc_list_countries‘ all return ‘ok:true‘ on real data; ‘iso3=”ZZZ”‘/‘””‘ return ‘ok:false‘ with a message. Gate: **50/50 pass**.

CAVEATS, IN THE ESTATE'S OWN WORDS
‘svc_scenario‘ runs the full DHS→forecast→cost path per call (seconds) — cache or pre-compute for interactive use. Services expose a curated subset of the model API; add services rather than letting the web layer reach into modules directly (keeps the seam clean). Error messages are developer-facing — the UI should map ‘error.type‘ to user-friendly text. Envelopes are transport-agnostic (HTTP/CLI/queue).

J. Teacher Demand & Allocation

File	‘docs/TEACHER_DEMAND.md‘
Command	16
Length	61 lines

real — The Gambia, GMB, World Bank measured

At a target PTR of 40:1:

Level	Year	Enrolment	Teachers	Current PTR	Required@40	Gap
Primary	2025	442,901	13,495	32.8	11,073	−2,422 (surplus)
Secondary	2021	369,783	12,452	29.7	9,245	−3,207 (surplus)

Both levels sit below a 40:1 target nationally — the binding problem is **distribution**, not aggregate numbers, which is exactly what the allocator addresses. Gate: **15/15 pass**.

CAVEATS, IN THE ESTATE'S OWN WORDS

National PTR masks subnational variation; the allocator needs **measured regional enrolment/teacher counts** to target real deployment (not in local data at region level — supply from an EMIS establishment file). Demand here is a single target PTR per level; subject-specialist and multigrade requirements, teacher attrition and pre-service supply (the pipeline to *fill* gaps) are later commands. Allocation treats teachers as homogeneous and freely movable — deployment frictions/costs compose with the costing command.

J. Network Travel-Time Accessibility

File	‘docs/TRAVEL_TIME.md‘
Length	60 lines

real — Kampala region

Analysis box around Kampala with real OSM schools & settlements:

	Value
Road nodes / edges	306,294 / 318,571
Schools in box	2,424
Settlements routed	1,417
Median travel to nearest school	0.7 min
Coverage < 15 / 30 / 60 min	0.992 / 0.992 / 0.992

Near-universal short travel time is expected in the dense capital (2,424 schools); a rural box shows longer times. Live in the app: **Models · simulate → Network Travel-Time** (Uganda; ~30 s first run, then cached).

CAVEATS, IN THE ESTATE'S OWN WORDS

Uganda only (the OSM road extract is Uganda; other countries need their Geofabrik extract at ‘data_sources.osm_roads_shp‘). The graph is **undirected** (one-way tags not yet applied) and uses free-flow class speeds (no congestion, condition or surface adjustment — the ‘edis-africa-transport-data‘ road_conditions/surfaces configs could refine them). Snapping is to the nearest network node (Euclidean); very remote points off the network route from their nearest node. National routing should be precomputed and cached, not run per request.

J. Genie UI Integration

File	‘docs/WEB_INTEGRATION.md‘
Command	29
Length	55 lines

Verified

Handler tests (34 assertions) pass on real data; ‘app.jl‘ parses after the edit; the real ‘/api/v2/scenario‘ endpoint returns a full scenario envelope (final GER, fingerprint) for The Gambia. Gate: **34/34 pass**.

CAVEATS, IN THE ESTATE'S OWN WORDS

The new endpoints are JSON APIs; a front-end page consuming them (charts/forms) is a presentation add-on left to the UI. ‘/api/v2/scenario‘ runs the full DHS→forecast→cost path per request (seconds) — add caching for interactive use. The endpoints are unauthenticated like the existing read APIs; apply the app's auth middleware if these should be protected. Running the server requires Genie's stack to load on the host (as the existing app already does).

J. EDIS Education — Architecture

File	‘docs/ARCHITECTURE.md‘
Command	01 — Canonical architecture (additive)
Length	131 lines

J. Attainment & Household Education-Expenditure Models

File	‘docs/ATTAINMENT_EXPENDITURE.md‘
Command	09
Length	77 lines

J. Baseline Test Status

File	‘docs/BASELINE_TEST_STATUS.md‘
Command	00 — baseline capture (no code changed)
Length	90 lines

J. Data Dictionary — Canonical Schemas

File	‘docs/DATA_DICTIONARY.md‘
Command	03 (regenerated)
Length	191 lines

J. Data Usage Audit — SDInfrastracture, D:\EDIP\Data, LSMS

File	‘docs/DATA_USAGE_AUDIT.md‘
Length	85 lines

J. EDIS Education Planning, Modelling & Decision-Support System — Documentation Index

File	‘docs/INDEX.md‘
Length	98 lines

J. Mathematical Specification

File	‘docs/MATHEMATICAL_SPECIFICATION.md‘
Length	56 lines

J. Performance & Benchmarking

File	‘docs/PERFORMANCE.md‘
Command	35
Length	53 lines

CAVEATS, IN THE ESTATE'S OWN WORDS
‘@elapsed‘ has ~microsecond resolution, so trivial operations legitimately read ‘0.0‘ s
(the harness asserts non-negative + finite, and requires a measurable op to be ‘> 0‘).
Timings are wall-clock and machine-dependent; use them for *relative* regression detection,
not absolute SLAs. Precompilation under ‘--pkgimages=no‘ dominates cold-start time and is not
counted here (it is a policy constraint, documented in the [index](INDEX.md)).

J. Consolidated Test Suite / Regression Runner

File	‘docs/REGRESSION_SUITE.md‘
Command	32
Length	54 lines

CAVEATS, IN THE ESTATE'S OWN WORDS
The full run executes real DHS/LSMS harmonization, JuMP/HiGHS solves and Genie-handler
paths, so it takes several minutes and requires the data roots configured in
‘config/defaults.toml‘. Use the argument filter for fast, targeted re-runs during
development. Data-dependent suites degrade to ‘@test true‘ skips (with an ‘@info‘) when a
source is absent, so the runner stays green on a partial data install rather than reporting
false failures — check the skip logs to confirm real coverage.

J. Repository Audit — EDIS Integrated Education System Model

File	‘docs/REPOSITORY_AUDIT.md‘
Command	00 — Audit the Existing Repository (no code changed)
Length	202 lines

Appendix K. The configuration

10 configuration files, verbatim. A result whose configuration is not recorded cannot be reproduced, and this course treats an irreproducible result as no result. These are the files that decide what the model specifications estimate, which survey variables map to which canonical names, and what the macro exchange will accept.

K. ‘config/cdcge.yml‘

The macro exchange configuration.

config/cdcge.yml

```
# CDCGE bridge configuration
# ---------------------------------------------------------------------------
# Couples this Education twin to the authoritative CDCGE hub (a GTAP-11 CGE
# solved with PATHSolver). The bridge is NON-INVASIVE: it writes an education
# results CSV in CDCGE's documented ‘read_edu_results‘ schema and calls CDCGE's
# public human-development linkage (‘edu_skill_growth → g_ls‘, SDG4 ‘HDParams‘)
# via a driver run under CDCGE's own project. CDCGE equations are never rewritten.
model_path: ”D:/EDIP/models/CDCGE”
driver: ”bridge/run_cdcge_education.jl” # our driver, run with --project=<CDCGE>
gtap_dir: ”D:/EDIP/Data/MRIO/GTAP”
focus: ”uga” # Uganda (GTAP region code)
base_year: 2017
nperiods: 6
# Files that must exist inside model_path for a real run.
required_files:
- ”CDCGE.jl”
- ”src/CDCGEData.jl”
- ”src/HumanDevLink.jl”
# CDCGE's read_edu_results schema — the interchange contract (long CSV).
edu_results_columns:
- ”year”
- ”population”
- ”enrolled”
- ”teachers”
- ”classrooms”
- ”learning_score”
- ”completers”
- ”education_spending”
- ”skilled_workers”
```

K. ‘config/defaults.toml‘

Global run defaults: horizon, base year, targets and the parameters every stage inherits.

config/defaults.toml

```
# EDIS Education — default configuration (template)
#
# Command 01 template; the Config subsystem that reads/validates this file is
# implemented in Command 02. These values mirror the constants currently hard-
# coded in src/Web/Service.jl and are moved here so no country-specific or
# standards value lives in source. Country files in config/countries/ override.
#
# All paths are resolved relative to the project root at load time — never hard-code
# absolute paths here.
schema_version = ”1.0.0”
[run]
base_year = 2026
default_country = ”UGA” # ISO3; must exist under data/countries/
solver = ”HiGHS” # optimisation solver
[standards]
target_class_size = 40.0 # pupils per classroom
target_pupil_teacher = 40.0 # target PTR
access_threshold_min = 45.0 # primary access standard (network travel minutes)
shift_factor = 1.0 # teaching sessions per classroom per day
[finance]
discount_rate = 0.05 # for CBA/present value (Command 24)
currency = ”local” # never assume USD; label explicitly
usd_to_local = 72.0 # example FX; overridden per country/year
# Per-unit education-input costs (local currency). Overridden per country.
[unit_costs]
teacher_salary = 6000.0
classroom_construction = 25000.0
teacher_training = 3000.0
materials_per_student = 45.0
textbook_per_student = 30.0
ict_per_student = 25.0
feeding_per_student = 60.0
furniture_per_student = 15.0
[optimization]
unmet_demand_penalty = 3000.0 # social value of enrolling one child (siting)
[paths]
# Relative to project root; resolved by Core/Paths.jl (Command 02).
data_raw = ”data/raw”
data_processed = ”data/processed”
data_spatial = ”data/spatial”
countries = ”data/countries”
outputs = ”outputs”
[data_sources]
# External data roots (machine-specific). Absolute paths are permitted HERE (in
# configuration), never in source. Empty string = that source is not available on
# this machine, and dependent steps must report it (never fabricate).
dhs_root = ”D:/EDIP/Data/Healthdata/DHSdata”
afrobarometer_dir = ”D:/EDIP/Data/HouseholdSurveys/Afrobarometer/country”
arabbarometer_dir = ”D:/EDIP/Data/HouseholdSurveys/ArabBarometer”
educationdata_root = ”D:/EDIP/Data/Educationdata/africa-education-gis”
# LSMS-ISA microdata (World Bank). Present for UGA, GHA, MWI, NGA, TZA.
lsms_root = ”D:/EDIP/Data/HouseholdSurveys/HouseholdsSurveys”
# Canonical World Bank LSMS microdata library (16 countries, multiple rounds each).
lsms_study_root = ”D:/EDIP/Data/HouseholdSurveys/Living Standards Measurement study”
```

K. ‘config/mappings/household/dhs_pr.toml‘

The household-member recode to canonical variables. This is the file chapter 3's attendance-code finding comes out of.

config/mappings/household/dhs_pr.toml

```
# DHS Household Member (PR) recode → canonical persons/households
#
# Source→canonical mapping for the DHS ”Household Member Recode” (PR). Standard DHS
# recode variable names (hv*) are stable across surveys, so one mapping serves every
# country/phase. Original source variables are preserved by the harmonizer; this file
# only declares the canonical translation. Nothing is imputed — codes with no mapping
# become ‘missing‘ and are flagged.
survey = ”DHS-PR”
recode = ”PR”
description = ”DHS Household Member Recode → canonical persons + households”
weight_divisor = 1000000 # canonical weight = hv005 / 1e6 (DHS convention)
# columns to read from the .dta (raw codes)
usecols = [”hv001”, ”hv002”, ”hvidx”, ”hv005”, ”hv024”, ”hv025”,
”hv104”, ”hv105”, ”hv106”, ”hv108”, ”hv121”, ”hv122”, ”hv123”, ”hv270”]
[keys]
cluster = ”hv001”
household = ”hv002”
line = ”hvidx”
# canonical persons field -> { source var, optional recode name / transform }
[persons]
region = { source = ”hv024”, transform = ”string” } # DHS region code as geo id
age = { source = ”hv105” }
sex = { source = ”hv104”, recode = ”sex” }
urban_rural = { source = ”hv025”, recode = ”urbrur” }
enrolled = { source = ”hv121”, recode = ”attending” }
grade = { source = ”hv123” }
attainment_years = { source = ”hv108” }
wealth_quintile = { source = ”hv270” } # 1 (poorest) … 5 (richest)
weight = { source = ”hv005”, transform = ”weight” }
# value-code recodes (raw code as string key -> canonical value)
[recodes.sex]
”1” = ”M”
”2” = ”F”
[recodes.urbrur]
”1” = ”urban”
”2” = ”rural”
[recodes.attending] # hv121: 0 not attending, 1 attending current yr, 2 attended
”0” = false
”1” = true
”2” = true
# Canonical variables NOT present in the PR recode. Declared here so the harmonizer
# flags them as unavailable (rule 5: never imputed). The note says where each lives.
[unavailable]
consumption = ”not in PR; DHS wealth quintile hv270 is a relative proxy only”
expenditure = ”not collected in DHS core recodes”
poverty = ”DHS has no monetary poverty line; use hv270 wealth quintile as proxy”
wages = ”IR (women, v-series) / MR (men, mv-series) recode”
occupation = ”IR/MR recode (v716/mv716)”
literacy = ”IR/MR recode (v155/mv155)”
parental_education = ”requires linking child (KR) records to parent PR lines”
```

K. ‘config/mappings/household/lsms_uga_unps.toml‘

The household panel to canonical variables, including every expenditure item chapter 6 sums.

config/mappings/household/lsms_uga_unps.toml

```
# LSMS-ISA Uganda National Panel Survey 2019/20 → canonical households/persons.
# Files live inside the World Bank distribution zip. Section paths are matched by
# suffix (robust to the top folder name). Categorical fields are read as their value
# LABELS (e.g. ”Male”/”Urban”); numeric fields (expenditure, age) as numbers.
survey = ”LSMS-UGA-UNPS-2019/20”
country = ”UGA”
zip = ”UGA_2019_UNPS_v03_M_STATA14.zip”
[sections] # relative path suffix inside the zip
education = ”HH/gsec4.dta”
roster = ”HH/gsec2.dta”
identification = ”HH/gsec1.dta”
consumption = ”pov2019_20.dta”
[keys]
household = ”hhid”
person = ”pid”
# person-level education (gsec4)
[education]
ever_attended = ”s4q05”
grade_completed = ”s4q07” # highest grade/class completed (attainment)
current_grade = ”s4q10” # currently attending grade (enrolment)
literacy = ”s4q04”
school_type = ”s4q12”
scholarship = ”s4q16”
# per-person education expenditure items (annual, past 12 months); summed per household
[education.expenditure]
fees = ”h4q15h”
registration = ”h4q15i”
exam = ”h4q15j”
boarding = ”h4q15e”
uniforms = ”h4q15b”
books = ”h4q15c”
transport = ”h4q15d”
daycare = ”h4q15k”
other = ”s4h15f”
total = ”h4q15g” # reported total per person
# roster demographics (gsec2)
[roster]
age = ”h2q8”
sex = ”h2q3”
# identification / location / weights (gsec1)
[identification]
region = ”region”
urban = ”urban”
weight = ”wgt”
# consumption / welfare aggregate (pov2019_20)
[consumption]
monthly_expenditure = ”nrrexp30” # nominal monthly household expenditure
welfare = ”welfare”
adult_equiv = ”equiv”
poor = ”poor_2020”
annualize_factor = 12 # education spend is annual; consumption is monthly
[numeric] # fields read as numbers (unwrapped)
fields = [”age”, ”grade_completed”, ”current_grade”,
”fees”,”registration”,”exam”,”boarding”,”uniforms”,”books”,”transport”,”daycare”,”other”,”total”,
”monthly_expenditure”,”welfare”,”adult_equiv”,”weight”]
[categorical] # fields read as value labels (strings)
fields = [”sex”,”urban”,”region”,”literacy”,”school_type”,”ever_attended”,”scholarship”,”poor”]
```

K. ‘config/models/attendance_logit.toml‘

The attendance model specification. Compare it with the enrolment one: the difference is the universe.

```
config/models/attendance_logit.toml
# School attendance model specification (Command 07).
# Outcome: currently attending school (canonical 'enrolled') among school-age children.
# Estimated on harmonized DHS 'persons' with survey weights.
name = "attendance_logit"
description = "Weighted logit of current school attendance on household/child covariates"
outcome = "enrolled"
link = "logit" # logit | probit | lpm
weights = "weight"
age_lo = 6
age_hi = 17
add_age_terms = true # include age and age^2
# covariates present in the canonical persons frame (categoricals auto-dummied)
covariates = ["sex", "urban_rural", "wealth_quintile"]
```

K. 'config/models/edi_weights.toml'

Deprivation index weights, which must sum to one. The index is reproducible from them and from nothing else.

```
config/models/edi_weights.toml
# Education Deprivation Index (EDI) — component weights & orientation (Command 10).
# Components are supplied as [0,1] deprivation rates (1 = most deprived), so
# method = "identity". EDI = Σ_k w_k · z_k, weights renormalized to sum to 1.
# The index is REPRODUCIBLE from these documented weights.
method = "identity"
[weights] # must sum to 1
non_enrolment = 0.20
dropout = 0.15
low_completion = 0.15
poor_access = 0.15
overcrowding = 0.10
teacher_shortage = 0.10
low_learning = 0.15
[directions] # all oriented so higher = more deprived
non_enrolment = "higher_worse"
dropout = "higher_worse"
low_completion = "higher_worse"
poor_access = "higher_worse"
overcrowding = "higher_worse"
teacher_shortage = "higher_worse"
low_learning = "higher_worse"
```

K. 'config/models/enrolment_logit.toml'

The enrolment model specification: outcome, link, weights, age band and covariates.

```
config/models/enrolment_logit.toml
# Ever-enrolment model specification (Command 07).
# Outcome: ever attended school (attainment_years > 0 OR currently enrolled).
name = "enrolment_logit"
description = "Weighted logit of ever-enrolment on household/child covariates"
outcome = "ever_enrolled" # derived by EnrolmentModel
link = "logit"
weights = "weight"
age_lo = 7
age_hi = 24
add_age_terms = true
covariates = ["sex", "urban_rural", "wealth_quintile"]
```

K. 'config/models/learning_production.toml'

The learning production function's parameters. Read this beside chapter 8: these coefficients were chosen, not fitted.

```
config/models/learning_production.toml
# Learning production-function model specification (Command 11).
# Estimates assessment ‘score‘ on student/household/school/teacher/infrastructure inputs.
name = ”learning_production”
description = ”Education production function (associational unless identification set)”
outcome = ”score”
covariates = [”attendance”, ”teacher_quality”, ”materials”, ”digital_access”, ”class_size”]
group = ”school_id” # for multilevel/FE
weights = ”” # optional student weight column
# Causal identification: ”none” = associational (default). Any real strategy (rct/iv/
# did/rd/panel_fe) must be implemented before results may be labelled causal — the
# engine refuses to mislabel otherwise.
identification = ”none”
```

K. ‘config/scenario.yml‘

Scenario engine settings.

```
config/scenario.yml
name: accelerated_learning
start_year: 2026
end_year: 2035
population_growth: 0.025
enrolment_push: 0.02
teacher_hiring_growth: 0.06
classroom_growth: 0.05
learning_reform_effect: 0.08
dropout_reduction: 0.02
annual_budget_growth: 0.08
```

K. ‘config/scenarios/baseline.toml‘

The baseline scenario — the one the committed run used.

```
config/scenarios/baseline.toml
# EDIS Education — baseline scenario (template)
#
# Every scenario inherits from this baseline and records only its changed
# parameters (Command 31). The scenario engine assigns a scenario_id and keeps
# outputs isolated. This baseline changes nothing — it is the reference run.
id = ”baseline”
inherits = ”” # baseline inherits from nothing
description = ”Reference run: current standards and trends, no policy change.”
[horizon]
start_year = 2026
num_years = 10
# System-dynamics levers (per year unless a vector is supplied). These match the
# projection engine's parameters; the baseline uses trend defaults.
[levers]
population_growth = 0.025
enrolment_push = 0.02
teacher_hiring_growth = 0.05
classroom_growth = 0.04
learning_reform_effect = 0.05
dropout_reduction = 0.02
annual_budget_growth = 0.06
[changes]
# Policy scenarios list only what they change relative to baseline, e.g.
# [changes.standards]
# target_pupil_teacher = 35.0
```

Appendix L. The country input bundles

The estate ships an input bundle for each of 54 countries: the tables a run reads before it can project anything. This appendix is the inventory, because the first question about running the model for a new country is what already there.

L.1 What a bundle contains

File	Rows across all countries	What it carries
‘settlements.csv‘	44,923	Settlements with population, poverty rate and travel time to the nearest school
‘schools.csv‘	28,857	Schools with classrooms, desks, water and electricity indices, coordinates and hazard exposure
‘enrolment.csv‘	1,536	Enrolment by region and level, with promotion, repetition and dropout rates
‘teachers.csv‘	768	Teachers by region and subject, with salary and attrition
‘candidate_sites.csv‘	433	Candidate school sites with capacity and construction cost
‘finance.csv‘	54	Government budget, donor funds and household spending
‘learning.csv‘	54	Mean score, attendance, teacher quality and digital access — one aggregate row, which is why chapter 8 estimates nothing
‘population.csv‘	54	Population by region and year, with the school-age share
‘labour_demand.csv‘	0	Labour demand by sector and skill
‘road_edges.csv‘	0	Road network edges for routing
‘travel_time.csv‘	0	Origin-to-school travel times

TWO OF THOSE FILES ARE EMPTY ALMOST EVERYWHERE
‘road_edges.csv‘ and ‘travel_time.csv‘ carry no rows for most countries, which is why the accessibility figures in this course are straight-line rather than routed (EG-13). ‘learning.csv‘ carries one aggregate row per country-region-year and no pupil records, which is why family D estimates nothing (EG-02). The inventory is how you find that out in a minute rather than a week.

L.2 Coverage by country

Country	Schools	Settlements	Enrolment rows	Teachers	Candidate sites
AGO	600	900	34	17	8
BDI	463	900	38	19	12
BEN	600	900	26	13	5
BFA	600	900	26	13	12
BWA	600	707	20	10	12
CAF	600	900	34	17	1
CIV	600	900	28	14	7
CMR	600	900	22	11	1
COD	600	900	48	24	3
COG	600	900	20	10	12
COM	195	262	6	3	9
CPV	429	688	42	21	9
DJI	404	203	14	7	12
DZA	600	900	94	47	12
EGY	600	900	52	26	11
ERI	459	721	14	7	12
ETH	600	900	20	10	7
GAB	600	900	20	10	12
GHA	600	900	32	16	8
GIN	600	900	18	9	4
GMB	599	1,281	16	8	12
GNB	97	900	16	8	12
GNQ	52	900	8	4	10
KEN	600	900	94	47	8
LBR	600	900	32	16	1
LBY	600	900	44	22	12

Country	Schools	Settlements	Enrolment rows	Teachers	Candidate sites
LSO	600	900	20	10	1
MAR	600	900	24	12	8
MDG	600	900	44	22	3
MLI	600	900	18	9	5
MOZ	600	900	20	10	12
MRT	593	900	26	13	5
MUS	600	500	20	10	12
MWI	600	900	8	4	6
NAM	600	900	28	14	12
NER	600	900	12	6	12
NGA	600	900	72	36	4
RWA	600	900	10	5	3
SDN	600	900	38	19	3
SEN	600	900	30	15	9
SLE	600	900	8	4	12
SOM	396	900	34	17	2
SSD	600	900	22	11	0
STP	79	424	4	2	6
SWZ	392	492	8	4	6
SYC	99	45	38	19	3
TCD	600	900	46	23	5
TGO	600	900	12	6	8
TUN	600	900	48	24	12
TZA	600	900	60	30	12
UGA	600	900	10	5	12
ZAF	600	900	18	9	12
ZMB	600	900	20	10	12
ZWE	600	900	20	10	12

L.3 What the committed runs used

Run	Country	Stages	Files written
edu_GMB_ 20260101T000000_664b64f1	Gambia	harmonize, demography, teacher_demand, scenario	9
edu_GMB_ 20260201T000000_664b64f1	Gambia	harmonize, demography, teacher_demand, scenario	10
edu_GMB_ 20260301T000000_664b64f1	Gambia	harmonize, demography, teacher_demand, scenario	9
edu_UGA_ 20260101T000000_664b64f1	Uganda	harmonize, demography, teacher_demand, scenario	9
edu_UGA_ 20260301T000000_664b64f1	Uganda	harmonize, demography, teacher_demand, scenario	9
edu_XXX_ 20260301T000000_664b64f1		—	3

Appendix M. The specification's reference tables

The specification document carries 13 reference tables that are not model cards: the layer definitions, the minimum data requirements, the canonical schema, the diagnostics each family owes, the standard scenarios, the registry fields and the macro mapping template. They are its operational content and they are reproduced here in full, because a reader working from this manual alone would otherwise have to go back to the source document for them.

TWO OF THESE TABLES DESERVE A SECOND READING

Minimum diagnostics by model family is the checklist a reviewer should hold you to, and most of it is not statistical.
The CGE mapping template is the missing link chapter 22 describes: it specifies exactly what the skilled-workers field needs, and family E's education-to-occupation table is the empirical content that would fill it.

M.1 The system layers

Chapter 1 works through these six.

Layer	Function
Data layer	Raw LSMS, HBS, LFS, MICS, DHS, census, EMIS, assessments, GIS and macro-fiscal data.
Harmonization layer	Canonical person, household, school, geography and labour-market identifiers and derived variables.
Econometric layer	Participation, expenditure, learning, returns, transitions and inequality models.
Spatial layer	Accessibility, catchments, capacity gaps, school siting and teacher allocation.
Projection layer	Cohort flows, demographic demand, teacher/classroom requirements and financing.
Microsimulation layer	Household income, poverty, inequality and benefit-incidence simulations.
CGE integration layer	Skill supplies, factor earnings, household factor ownership, productivity and government education accounts.
Decision layer	Scenario comparison, budget optimization, investment ranking and policy recommendations.

M.2 Data source families and their minimum information

What each kind of source must carry before a model can use it. Chapter 3 checks the platforms this course reaches against it.

Source family	Minimum information	Primary modelling role
LSMS / LSMS-ISA	Roster, education, expenditure, employment, agriculture, shocks, geography	Participation, costs, poverty, child labour, returns, microsimulation
Household Budget Survey	Consumption, income, education expenditure, demographics	Affordability, demand systems, poverty, benefit incidence
Labour Force Survey	Education, occupation, industry, employment, wages, hours	Returns, transitions, skills mismatch, labour factor construction
MICS	Attendance, completion, literacy, child characteristics	Access, completion, ECD, inequality
DHS	Attainment, literacy, fertility, health, household wealth, GPS	Education-health-fertility links, spatial inequalities
Population Census	Age-sex-location structure, schooling, occupation	Small-area denominators and long-run education demand
EMIS / School Census	Schools, enrolment, teachers, classrooms, facilities	Supply, capacity, staffing, unit costs
Learning Assessments	Student scores, teacher/school inputs	Education production functions and learning inequality
GIS / OSM / WorldPop	School coordinates, roads, settlements, population grids	Travel time, accessibility, catchments and siting
Fiscal / Budget	Education spending by level/economic classification	Unit costs, projections, fiscal scenarios and optimization

M.3 The canonical microdata schema

The variable names every loader harmonises to. A mapping file translates a survey's own names into these, and Appendix K carries two such files.

Domain	Canonical variables
Identifiers	country, survey_year, household_id, person_id, cluster_id, school_id, district_id, adm1_id, adm2_id
Demographics	age, sex, relationship_to_head, marital_status, disability, migration_status
Education	ever_attended, currently_enrolled, level, grade, highest_grade, years_schooling, attendance_days, repeat, dropout, literacy, numeracy, score
School choice/access	school_type, ownership, distance_km, travel_time_min, transport_mode, boarding

Domain	Canonical variables
Education costs	fees, uniforms, books, transport, meals, boarding, tutoring, other__cost, total__education__cost
Household welfare	income, consumption, per__capita__consumption, adult__equivalent__consumption, poverty__status, wealth__index
Parents/household	mother__education, father__education, household__size, children__count, dependency__ratio, shocks
Labour	employment__status, occupation, industry, hours, wage, earnings, formal, self__employed
Survey design	weight, strata, psu
Geospatial	latitude, longitude, urban__rural, settlement__id, travel__time__school

M.4 Minimum diagnostics by model family

What each family owes before its output may be used. Chapter 4 covers the design-based ones; the rest are the model card's Diagnostics field.

Model family	Minimum diagnostics
Binary/multinomial	Weighted fit, calibration, ROC/AUC where predictive, marginal effects, rare-event checks, survey-design SEs
Linear/earnings	Residuals, heteroskedasticity-robust SEs, influential observations, functional form, wage definition
Hazard/survival	Censoring, proportional-hazards or time-varying effects, survival calibration, competing-risk consistency
Multilevel	ICC, random-effect variance, cluster counts, shrinkage sensitivity
Demand systems	Adding-up/homogeneity/symmetry, price/expenditure elasticities, positivity and budget-share checks
Causal/quasi-experimental	Identification statement, balance/pre-trends, placebo tests, sensitivity to bandwidth/specification
Spatial	Coordinate validity, routing completeness, travel-time plausibility, boundary and resolution sensitivity
Optimization	Feasibility, objective value, optimality gap, binding constraints, sensitivity to budgets/costs
Projection	Stock-flow identities, backcasting, demographic reconciliation, scenario uncertainty
CGE/microsimulation	SAM balance, benchmark replication, closure, factor-market balance, household-income reconciliation, decomposition

M.5 Standard policy scenarios

The shocks the scenario engine is built to run, and the outputs each produces.

Scenario	Typical shock	Principal outputs
Universal primary access	New schools + lower travel time + fee support	Enrolment, OOSC, cost, poverty
Secondary expansion	Classrooms, teachers and bursaries	Completion, skill supply, wage effects
TVET scale-up	Increased TVET seats and occupational alignment	Skill gaps, sector employment, productivity
Girls education package	Bursaries, safety/access, transition support	Gender completion gap, fertility link, labour income
School feeding	Per-student feeding subsidy	Attendance, costs, poverty targeting
Teacher redistribution	Reallocate existing teachers	PTR dispersion, learning proxy, recurrent cost
Infrastructure package	School construction, WASH, electricity, internet	Access, quality index, enrolment
Budget reallocation	Shift spending from lower-return uses toward education sub-sectors	GDP, fiscal balance, poverty, learning/skills
Climate shock resilience	Household income/agricultural shock plus social protection	Dropout, attendance, long-run human capital
Digital education	Connectivity/device investment with learning productivity assumption	Learning, access, unit cost, skills

M.6 Evaluation approach by use case

Which validation is appropriate for prediction, for inference and for projection. They are not interchangeable.

Use case	Preferred evaluation
Dropout risk	AUC, precision-recall, calibration, subgroup fairness
OOSC prediction	Calibration by district, spatial cross-validation
Learning prediction	RMSE/MAE, out-of-school validation
School congestion	Forecast error and capacity calibration
Skill shortage	Backcast accuracy and sector consistency
Policy targeting	Expected gain per beneficiary/cost, stability across folds

M.7 Implementation modules

The specification's own suggested module structure. Compare it with the module index in Appendix E, which is what was actually built.

Module	Suggested implementation
Data ingestion/harmonization	Julia DataFrames.jl/CSV.jl/XLSX.jl; R haven/arrow/data.table
Survey statistics	R survey/srvyr; Julia weighted estimators with explicit design handling
GLM/econometrics	Julia GLM.jl/FixedEffectModels.jl; R glm/fixest/survival/nnet
Spatial	Julia ArchGDAL.jl; R sf/terra/osrm/igraph as appropriate
Optimization	Julia JuMP + HiGHS/Gurobi/CPLEX where licensed
Projection	Julia DataFrames/StaticArrays or R data.table matrices
Microsimulation	Julia/R custom simulation engine with immutable scenario inputs
CGE interface	Julia JuMP/PATHSolver or existing EDIS CGE stack
Results	Parquet/CSV + metadata JSON + model registry

M.8 Model registry fields

What must be recorded about every fitted model. A result whose registry entry is missing cannot be reproduced.

Field	Description
model_id	Stable ID such as EDU_M01_ENROLMENT
country	ISO3 country code
survey	Source and wave
sample	Age/eligibility filters
outcome	Canonical dependent variable
specification	Formula or optimization objective
weight/design	Survey weight, strata, PSU treatment
estimator	Estimator/solver
diagnostics	Pass/fail and key metrics
scenario	Baseline or named policy scenario
run_timestamp	Execution timestamp
code_version	Git commit/tag or package version
output_path	Machine-readable result location

M.9 The phased development plan

The order the specification proposes for building the catalogue, and the priority models in each phase.

Phase	Priority models	Purpose
Phase 1: Household core	1, 4, 6, 9, 11, 12, 17, 20, 25, 43, 46	Create immediate survey-based analytics.
Phase 2: Spatial supply	53-60	Link household demand to schools, access, capacity and teachers.

Phase	Priority models	Purpose
Phase 3: Learning/quality	33-42	Move from schooling quantity to learning outcomes.
Phase 4: Projection/fiscal	62-68	Build medium- and long-term education planning model.
Phase 5: Policy simulation	69-72	Integrate transfers, fees, human capital, CGE and poverty.

M.10 Canonical derived variables

Quantities every implementation must derive the same way — including the age-for-grade construction that chapter 5's over-age finding rests on.

Variable	Definition
school_age	1 if person is within nationally defined compulsory/ school-age range
age_for_grade	age - official_age_for_current_grade
overage	1 if age_for_grade exceeds policy threshold
oosc	1 if school-age and not currently enrolled
education_cost_total	sum of fees, books, uniforms, transport, meals, boarding, tutoring and other costs
education_burden	education_cost_total / household_total_consumption
years_schooling	harmonized completed years derived from level and grade
education_level_harmonized	ISCED-compatible or EDIS level category
hourly_wage	cash and in-kind labour earnings divided by hours, after cleaning
formal_worker	country-specific formal employment indicator harmonized to EDIS rules
neet	youth not in employment, education or training
travel_time_school	estimated network travel minutes to nearest relevant school
school_capacity	seats or classrooms × target students per classroom
capacity_gap	demand minus capacity
learning_z	standardized test score within assessment/grade where appropriate

M.11 Standard scenario input parameters

The levers a scenario may move, with example values. Read this beside chapter 15: almost every one of them is a policy choice rather than a measurement.

Parameter	Example value	Notes
scenario_name	Secondary Expansion	Unique descriptive identifier
start_year	2027	First policy year
end_year	2035	Simulation horizon
budget_change_pct	20	Relative to baseline education budget
primary_share	0.40	Share of incremental budget
secondary_share	0.30	Share of incremental budget
tvet_share	0.20	Share of incremental budget
tertiary_share	0.10	Share of incremental budget
new_schools	250	Optional capital target
new_classrooms	2000	Optional capital target
new_teachers	5000	Optional staffing target
fee_subsidy_pct	50	Optional household-cost shock
target_group	poorest_40	Eligibility if targeted
productivity_link	scenario-specific	Must be documented; do not assume immediate productivity effects

M.12 The CGE mapping template

Education micro categories to CGE factors and household groups. This is the mapping chapter 22's empty skilled-workers field needs.

Education micro category	Suggested CGE factor	Household mapping
No schooling / incomplete primary	LAB_LOW	Income shares by survey household group
Primary completed	LAB_PRIMARY	Income shares by survey household group
Lower/upper secondary	LAB_SECONDARY	Income shares by survey household group
TVET/technical	LAB_TECH	Income shares by survey household group
Tertiary/professional	LAB_HIGH	Income shares by survey household group

M.13 Glossary

The specification's own definitions.

Term	Meaning
CGE	Computable General Equilibrium
EMIS	Education Management Information System
HBS	Household Budget Survey
LSMS	Living Standards Measurement Study
LSMS-ISA	LSMS Integrated Surveys on Agriculture
LFS	Labour Force Survey
MICS	Multiple Indicator Cluster Survey
DHS	Demographic and Health Survey
OOSC	Out-of-school children
PTR	Pupil-teacher ratio
TVET	Technical and Vocational Education and Training
SAM	Social Accounting Matrix
LAYS	Learning-adjusted years of schooling
NEET	Not in employment, education or training
MILP	Mixed Integer Linear Programming