

Child-Direct Disability Screening in Tanzanian. Primary Schools: Estimating Administrative Undercounts and Implications for Disability Prevalence in Low- and Middle-Income Countries.

This repository contains the code to reproduce all the results presented in the paper titled "Child-Direct Disability Screening in Tanzanian. Primary Schools: Estimating Administrative Undercounts and Implications for Disability Prevalence in Low- and Middle-Income Countries."

Contents

1. [Overview](#)
2. [Data Availability](#)
3. [Instructions for Replicators](#)
4. [List of Exhibits](#)
5. [Requirements](#)
6. [Key Variables Descriptions](#)

Overview

This replication package contains the original (raw) data files, provided in their unaltered forms, to ensure the verification of reproducibility. Additionally, it includes all necessary code files required to replicate the analysis from the original dataset to the results presented in the associated paper.

Data Availability

This paper uses both publicly available and data that is embargoed until publication, as detailed below.

Data Sources:

Publicly available data include:

1. IdentificationPilotMap: Consolidated_Primary_EnrolmentbyGrade_PTR_2023_PSLE2022.xlsx, sheet "TSM_Enrol2023", is from Republic of Tanzania President's Office Regional Administration and Local Government (PO-RALG) Basic Education Education Statistics 2022. Available from [PO-RALG website](#) and accessed in 2024.
2. Best: BEST_2022_Regional_Data_Final.xlsx, sheet "T2.6EnrolStages" is from PO-RALG – Pre-Primary, Primary, Secondary, Adult, and Non-Formal Education Statistics 2022. Typically available from [PO-RALG website](#), the website is currently not working for downloads. The team is checking with the government about when the data may be back online. Accessed in 2024. The data will be included in the reproducibility package.
3. UnicefMics: xxx_MICS6_SPSS_Datasets_fs.sav, where xxx is one of the 51 countries used in Analysis (Afghanistan, Algeria, Argentina, Bangladesh, Belarus, Benin, Central African Republic, Chad, Costa Rica, Cuba, Dominican Republic, DR Congo, Eswatini, Fiji, Georgia, Ghana, Guinea Bissau, Guyana, Honduras, Iraq, Jamaica, Kiribati, Kosovo (UNSCR 1244) (Roma, Ashkali and Egyptian Communities), Kosovo (UNSCR 1244), Kyrgyzstan, Lao PDR, Lesotho, Madagascar, Malawi, Mongolia, Montenegro, Montenegro (Roma Settlements), Nepal, Nigeria, Pakistan (Baluchistan), Pakistan Khyber Pakhtunkhwa, Pakistan Punjab, Pakistan Sindh, Republic of

North Macedonia, Republic of North Macedonia (Roma Settlements), Samoa, Sao Tome and Principe, Serbia, Serbia (Roma Settlements), Sierra Leone, State of Palestine, Suriname, Thailand, The Gambia, Togo, Tonga, Trinidad and Tobago, Tunisia, Turkmenistan, Turks and Caicos Islands, Tuvalu, Uzbekistan, Viet Nam, Yemen, Zimbabwe). Available for download from <https://mics.unicef.org/surveys> with an account. Please apply the following filters, Round: MICS6, Data type: MICS and Status: Completed. The MICS6 surveys were conducted between 2017 and 2023, with most of the data collected in 2018 and 2019.

4. UnicefMics: MICS_population.xlsx. Accessed in 2024.

- total_5to17 and total_6to13: total population in 2023 for the respective age groups, sourced from the United Nations, Department of Economic and Social Affairs, Population Division (2024). Available at <https://population.un.org/dataportal/>.
- gdp_p_c: GDP per capita in 2023, as reported in the World Bank's World Development Indicators (2025). GDP per capita figures for the State of Palestine (2023) and Cuba (2022) are obtained from Trading Economics (2023). Available at <https://databank.worldbank.org/source/world-development-indicators> and <https://tradingeconomics.com/country-list/gdp?continent=world>.

SB Tanzania

Forthcoming in the Microdata Library. Embargoed until journal publication (estimated release: 2026). The data include surveys conducted by an implementing partner as part of the Shule Bora (SB) baseline, to which researchers have been granted access. The Shule Bora baseline survey data were de-identified by the local Monitoring and Evaluation (M&E) team and confidentially shared with the DIME research team. These data cannot be made publicly available without prior authorization from the relevant program authorities; such authorization will be requested once the related paper is published.

The dataset comprises two main components: the Student Survey and the Early Grade Reading and Mathematics Assessment (EGRA/EGMA).

1. ShuleBoraCambEdBaseline: Student_Survey.dta

1. Name or title of the survey: Children Enrolled in Public Primary Schools in Shule Bora (SB) Tanzania Regions
2. Brief description: This dataset covers students enrolled in public primary schools across nine Tanzanian regions under the Shule Bora program (Dodoma, Katavi, Kigoma, Mara, Pwani, Rukwa, Simiyu, Singida, and Tanga). SB is a Government of Tanzania program supported by UK Aid to enhance pre-primary and primary education quality. The baseline survey, implemented by the M&E team and DataVision International Ltd., collected data in 180 schools between 20 February and 6 March 2023. Sampling included both lower (Standards 1–3) and upper (Standards 4–7) grades, resulting in a total of 2,160 students. The survey used the Child Functioning Module (CFM) developed by the Washington Group and UNICEF to assess child functioning and disability.
3. Dates accessed or survey conducted: 20 February 2023 – 6 March 2023
4. Citation: Government of Tanzania, Shule Bora Program, and DataVision International Ltd. (2023). *Shule Bora Baseline Student Survey (SB Tanzania)*. Forthcoming in World Bank Microdata Library.

2. ShuleBoraCambEdBaseline: EGRA_EGMA_Assessment.dta

1. Name or title of the survey: Early Grade Reading Assessment (EGRA) / Early Grade Mathematics Assessment (EGMA)
2. Brief description: This dataset includes the results of EGRA and EGMA tests conducted as part of the Shule Bora baseline in 2023. Following the same sampling protocol as the student survey, 12 Grade 3 students per school were randomly selected, with six reserves, totaling 1,082 children. The assessments and accompanying

questionnaires were developed by the Shule Bora M&E team, approved by the Government of Tanzania, and implemented by trained enumerators from DataVision International Ltd.

3. Dates accessed or survey conducted: 20 February 2023 – 6 March 2023

4. Citation: Government of Tanzania, Shule Bora Program, and DataVision International Ltd. (2023). *Shule Bora Baseline EGRA/EGMA Assessment (SB Tanzania)*. Forthcoming in World Bank Microdata Library.

Additionally, the implementing partner confidentially shared the following datasets:

1. IdentificationPilotMap: CSEEtCtRegions18-20.dta (used as ShuleBora region reference ID for matching between datasets)

1. Name or title of the survey: Certificate of Secondary Education Examination (CSEE) test results 2018 – 2020.
2. Brief description: National achievement examination administered to students who have completed four years of secondary education. In this project, the dataset CSEEtCtRegions18–20.dta is used exclusively as a regional reference file for the Shule Bora program to match school regions across datasets. Only region ID and school region variables are used; no achievement or subject-level outcome variables are included.
3. Dates accessed or survey conducted: Administered annually during the first week of November; dataset covers examination years 2018–2020.
4. Citation: United Republic of Tanzania, National Examinations Council of Tanzania (NECTA). Certificate of Secondary Education Examination (CSEE), 2018–2020. Dataset: CSEEtCtRegions18–20.dta.

2. IdentificationPilotMap: SB_baseline_list_of_schools.xlsx, sheet "Primary_Gov"

1. Name or title of the survey: Shule Bora Baseline List of Schools, Government Primary Schools
2. Brief description: List of government primary schools by location and type of special needs alignment used as the baseline school roster for the Shule Bora program. The dataset originates from the Annual School Census (ASC) system managed by PO-RALG (now PMO-RALG) and is used to identify government primary schools and their characteristics for baseline sampling and matching purposes.
3. Dates accessed or survey conducted: Data reflect school information as of 31 March 2022 and were entered into the ASC system at the school level by heads of schools.
4. Citation: United Republic of Tanzania, President's Office – Regional Administration and Local Government (PMO-RALG). Annual School Census (ASC), 2022. Shule Bora Baseline List of Schools (Primary_Gov), unpublished administrative data.

3. IdentificationPilotMap: SB_Inclusive_schools_in_9_regions_v3.xlsx, sheet "Schools_with_CWD_by_region"

1. Name or title of the survey: Ofisi ya Rais Tamisemi – Orodha ya Shule Zenye Wanafunzi Wenye Mahitaji Maalum kwa Aina ya Ulemavu / Office of the President – Regional Administration and Local Government (PO-RALG) – List of Schools with Students with Special Needs by Type of Disability
2. Brief description: Administrative dataset containing government disability statistics on the number of children with disabilities (CWDs) enrolled in government primary schools across nine regions, disaggregated by type of disability. The data are not publicly available and are used in the Shule Bora program to identify and classify inclusive schools for sampling, weighting, and analytical purposes.
3. Dates accessed or survey conducted: Collected by PO-RALG directly from schools and compiled at the regional level; specific collection dates are not publicly documented.
4. Citation: United Republic of Tanzania, President's Office – Regional Administration and Local Government (PO-RALG). Administrative data on Children with Disabilities in Primary Schools, nine regions. Shule Bora Inclusive Schools in 9 Regions dataset, unpublished administrative data.

4. VisionCare: School screening overall statistics (2025) _WB sharing data.xlsx

Data from Vision Care Tanzania include the file School screening overall statistics (2025) _WB sharing data.xlsx, which contains information on first and second vision screenings, expected and actual follow-up screenings, and provision of spectacles. These data were shared directly by the implementing partner (Vision Care Tanzania) and are available upon request, subject to approval from both Vision Care Tanzania and the Korea International Cooperation Agency (KOICA). Access year: 2025-11-21. URL: <https://vcs2020.org/> (Shared directly by institutional contact by email).

5. VisualAcuity: Figure1Data.xlsx

The package also includes data from the study Baseline MNREAD Measures for Normally Sighted Subjects From Childhood to Old Age, specifically the file Figure1Data.xlsx, which provides information on reading speed (words per minute) and print size (logMAR). The underlying figure is publicly available from the article repository (<https://pmc.ncbi.nlm.nih.gov/articles/PMC4961000/>) and is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (accessed on 2026-03-24). The data reported in Figure1Data.xlsx were digitized by pasting Figure 1 from the paper into PlotDigitizer.

6. IdentificationPilotMap: Pupils with Disabilities in Primary Schools, 2024_2.xlsx

The file Pupils with Disabilities in Primary Schools, 2024_2.xlsx was obtained from the Government of Tanzania through a direct institutional contact. It is associated with the Ministry of Education data platform available at <https://esmis.moe.go.tz/> (shared directly by institutional contact via email), and was accessed on 2024-09-27. The file contains school-level administrative information with variables including Region, Council, Ward, School, Reg. No., Ownership, and counts of children with disabilities disaggregated by sex and disability type, including Albino, Deaf, Hard of Hearing, Deaf-Blind, Blind, Low Vision, Physical Impairment, Autism, Intellectual Impairment, and Multi-Impairment for boys and girls separately. The dataset is under a restricted-use license by request only and requires approval from the Government of Tanzania for access or redistribution.

7. IdentificationPilotMap: Consolidated Primary Education Data, 2024.xlsx

The dataset Consolidated Primary Education Data, 2024.xlsx was obtained from the Government of Tanzania through a direct institutional contact. It is associated with the Ministry of Education data platform available at <https://esmis.moe.go.tz/> (shared directly by institutional contact via email), and was accessed on 2024-09-30. The file contains school-level administrative information with variables including Region, Council, Ward, School, Reg. No., NECTA Centre No., Ownership, Latitude, Longitude, and student enrollment counts by grade and sex for Standard I through Standard VII (boys and girls separately), as well as Total Boys and Total Girls. The dataset is under a restricted-use license by request only and requires approval from the Government of Tanzania for access or redistribution.

Statement about Rights:

- I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Instructions for Replicators

Below are the steps to follow to run the package successfully:

- Repicator must update "DataWork\Inclusion\Dofiles\Main_dofile.do" do file with your directory paths for the OneDrive folder with the raw data and the Github where you clone this repo, for example:

```
global user 1
```

```

* Root folder globals
* -----

if $user == 1 {
    global projectfolder "C:\Users\wb570123\WBG\Maria Valentina Daza Pardo - SEN
Reproducibility" // filepath to data and output files
    global Dofiles "C:\Users\wb570123\Github\shule-bora\DataWork\Inclusion\Dofiles"
// filepath for github repo
}

```

- Ensure all required software and dependencies are installed as listed in the [Requirements](#) section.
- Run the "Github\shule-bora\DataWork\Inclusion\Dofiles\Main_dofile.do" do file.
- Tables will be exported to Output and figures to Output\Figures subfolders of the OneDrive folder. Below is the list of tables and figures with their respective output locations.
 - Note that, for each table, raw results are exported into an Excel sheet "TableXData", and the formulas in "Output\Tables.xlsx" automatically populate the raw values into a formatted table in sheet "TableX". All tables are in "Output\Tables.xlsx", except for tables A7 and A8, which follow a slightly different format and can be found in "Output\TableA7.xlsx" and "Output\TableA8.xlsx", respectively.
- To produce the final tables in the Overleaf manuscript we convert the formatted Excel tables ("Output\Tables.xlsx", "Output\TableA7.xlsx" and "Output\TableA8.xlsx") into .tex files using Excel2LaTeX, a free Excel add-in that converts Excel tables into LaTeX code. Please follow instructions here on how to use the add-in: <https://ctan.org/pkg/excel2latex?lang=en>

List of Exhibits

Exhibit name	Output filename	Script	Note
Table 6.1	"Output\Tables.xlsx"	Stata008AnalyzeFinalSampleTable1.do	Formatted table in sheet "Table1" and Stata export in sheet "Table1Data"
Table 6.4	"Output\Tables.xlsx"	Stata008AnalyzeFinalSampleTable2.do	Formatted table in sheet "Table2" and Stata export in sheet "Table2Data"
Table 6.5	"Output\Tables.xlsx"	Stata008AnalyzeFinalSampleTable3.do	Formatted table in sheet "Table3" and Stata export in sheet "Table3Data"
Table 6.2	"Output\Tables.xlsx"	Stata008AnalyzeFinalSampleTable4.do	Formatted table in sheet "Table4" and Stata export in sheet "Table4Data"

Exhibit name	Output filename	Script	Note
Table 6.3	"Output\Tables.xlsx"	Stata008AnalyzeFinalSampleTable5.do	Formatted table in sheet "Table5" and Stata export in sheet "Table5Data"
Table 6.6	"Output\Tables.xlsx"	Stata008AnalyzeFinalSampleTable6.do	Formatted table in sheet "Table6" and Stata export in sheet "Table6Data"
Table 6.7	"Output\Tables.xlsx"	Stata008AnalyzeFinalSampleTable7.do	Formatted table in sheet "Table7" and Stata export in sheet "Table7Data"
Table 6.8	"Output\Tables.xlsx"	Stata008AnalyzeFinalSampleTable8.do	Formatted table in sheet "Table8" and Stata export in sheet "Table8Data"
Table 6.9	"Output\Tables.xlsx"	Stata008AnalyzeFinalSampleTable9.do	Formatted table in sheet "Table9" and Stata export in sheet "Table9Data"
Table 6.10	"Output\Tables.xlsx"	Stata010TabulateAscVsCfmLocal.do	Formatted table in sheet "Table10" and Stata export in sheet "Table10Data"
Table 6.11	"Output\Tables.xlsx"	Stata011TabulateVisionCare.do	Formatted table in sheet "Table11" and Stata export in sheet "Table11Data"
Table 6.12	"Output\Tables.xlsx"	Stata012TabulateSeeingAndHearingGlobal.do	Formatted table in sheet "Table12" and Stata export in sheet "Table12Data"

Exhibit name	Output filename	Script	Note
Table 6.13	"Output\Tables.xlsx"	Stata013ExamineVisualAcuity.do	Formatted table in sheet "Table13" and Stata export in sheet "Table13Data"
Table A1	"Output\Tables.xlsx"	Stata008AnalyzeFinalSampleTableA1.do	Formatted table in sheet "TableA1" and Stata export in sheet "TableA1Data"
Table A2	"Output\Tables.xlsx"	Stata008AnalyzeFinalSampleTableA2.do	Formatted table in sheet "TableA2" and Stata export in sheet "TableA2Data"
Table A3	"Output\Tables.xlsx"	Stata008AnalyzeFinalSampleTableA3.do	Formatted table in sheet "TableA3" and Stata export in sheet "TableA3Data"
Table A4	"Output\Tables.xlsx"	Stata008AnalyzeFinalSampleTableA4.do	Formatted table in sheet "TableA4" and Stata export in sheet "TableA4Data"
Table A5	"Output\Tables.xlsx"	Stata008AnalyzeFinalSampleTableA5.do	Formatted table in sheet "TableA5" and Stata export in sheet "TableA5Data"
Table A6	"Output\Tables.xlsx"	Stata008AnalyzeFinalSampleTableA6.do	Formatted table in sheet "TableA6" and Stata export in sheet "TableA6Data"
Table A7	"Output\TableA7.xlsx"	Stata008AnalyzeFinalSampleTableA7A8.do	Formatted table in sheet "Sheet1"

Exhibit name	Output filename	Script	Note
			and Stata export in sheet "Sheet1Data"
Table A8	"Output\TableA8.xlsx"	Stata008AnalyzeFinalSampleTableA7A8.do	Formatted table in sheet "Sheet1" and Stata export in sheet "Sheet1Data"
InTextNumbers	"Output\Tables.xlsx"	Stata009InTextNumbers.do	Found in sheet "InTextNumbers"; produces in-text statistics not directly tied to tables.
Figure 1	"Output\Figures\Figure1.pdf"	Stata008AnalyzeFinalSampleFigure1.do	Combined figure panels 1a and 1b.
Figure 2	"Output\Figures\Figure2.pdf"	Stata008AnalyzeFinalSampleFigure2.do	Combined figure panels 2a and 2b.
Figure 3	"Output\Figures\Figure3.pdf"	Stata008AnalyzeFinalSampleFigure3.do	Combined Domain, Indicator, Score, and Severity panels.
Figure 4	"Output\Figures\Figure4.pdf"	Stata008AnalyzeFinalSampleFigure4.do	Combined Domain, Indicator, Score, and Severity panels.

Requirements

Computational Requirements

To successfully replicate the results, the following computational environment must be set up. This includes the operating system, software dependencies, and environment configurations. Ensuring alignment with these specifications is crucial for reproducibility.

Operating System: Windows 10/11, macOS

Environment Setup: Instructions for setting up virtual environments, containers (Docker), or conda environments if applicable

Software Requirements

The codes are run using Stata version 14.1. Program files (.ado) that are required to run the codes are included as part of the code package.

Memory and Runtime and Storage Requirements

Memory: The computational environment should have at least 3 GB of RAM to ensure smooth execution without memory overflow.

Runtime: Approximate runtime for the complete analysis is 40 minutes on the specified hardware configuration (Intel i7 processor, 16GB RAM).

Storage:

- The repository itself occupies about 450 MB of space.
- Additional disk space of 2 GB is required for unzipping datasets, storing temporary files, and intermediate results.
- Data downloads may require stable internet and sufficient space to avoid interruptions.

Key Variables Descriptions

Variable Name	Variable Description	Variable Construction
Seeing_5to17	Binary (0/1), child aged 5–17 has functional difficulty in seeing.	CF2=3 or 4, or CF3=3 or 4
Hearing_5to17	Binary (0/1), child has functional difficulty in hearing.	CF5=3 or 4, or CF6=3 or 4
Walking_5to17	Binary (0/1), child has functional difficulty in walking.	(CF8=3 or 4) or (CF9=3 or 4) or (CF12=3 or 4) or (CF13=3 or 4)
Self-care_5to17	Binary (0/1), child has functional difficulty in self-care.	CF14=3 or 4
Communication_5to17	Binary (0/1), child has functional difficulty in communication.	CF15=3 or 4, or CF16=3 or 4
Learning_5to17	Binary (0/1), child has functional difficulty in learning.	CF17=3 or 4
Remembering_5to17	Binary (0/1), child has functional difficulty in remembering.	CF18=3 or 4
Concentrating_5to17	Binary (0/1), child has functional difficulty in concentrating.	CF19=3 or 4
AcceptingChange_5to17	Binary (0/1), child has functional difficulty in accepting change.	CF20=3 or 4
Behaviour_5to17	Binary (0/1), child has functional difficulty in controlling behaviour.	CF21=3 or 4
MakingFriends_5to17	Binary (0/1), child has functional difficulty in making friends.	CF22=3 or 4

Variable Name	Variable Description	Variable Construction
Anxiety_5to17	Binary (0/1), child has functional difficulty in anxiety.	CF23=1
Depression_5to17	Binary (0/1), child has functional difficulty in depression.	CF24=1
FunctionalDifficulty_5to17	Binary (0/1), indicates functional difficulty if at least one domain shows difficulty.	Any domain variable above = 1
FunctionalDifficulty_5to17_Miss0	Binary (0/1), same as FunctionalDifficulty_5to17, with missing = 0.	FunctionalDifficulty_5to17, missing treated as 0
FunctionalDifficulty_without	Binary (0/1), same as FunctionalDifficulty_Miss0 excluding anxiety & depression.	FunctionalDifficulty_5to17_Miss0 excluding Anxiety_5to17 and Depression_5to17
S01AnxietyScore	Index (0–3), child experiences anxiety daily (3), weekly (2), monthly (1), or rarely/never (0).	Based on frequency of anxiety
S02DepressionScore	Index (0–3), child experiences depression daily (3), weekly (2), monthly (1), or rarely/never (0).	Based on frequency of depression
S03BehaviourScore	Index (0–3), difficulty controlling behaviour: severe (3), moderate (2), mild (1), none (0).	CF21 response
S04AcceptingChangeScore	Index (0–3), difficulty accepting change: severe (3), moderate (2), mild (1), none (0).	CF20 response
S05WalkingScore	Index (0–3), difficulty walking: severe (3), moderate (2), mild (1), none (0).	CF8–CF13 responses
S06LearningScore	Index (0–3), difficulty learning: severe (3), moderate (2), mild (1), none (0).	CF17 response
S07RememberingScore	Index (0–3), difficulty remembering: severe (3), moderate (2), mild (1), none (0).	CF18 response
S08MakingFriendsScore	Index (0–3), difficulty making friends: severe (3), moderate (2), mild (1), none (0).	CF22 response
S09ConcentratingScore	Index (0–3), difficulty concentrating: severe (3), moderate (2), mild (1), none (0).	CF19 response

Variable Name	Variable Description	Variable Construction
S10SelfcareScore	Index (0–3), difficulty with self-care: severe (3), moderate (2), mild (1), none (0).	CF14 response
S11CommunicationScore	Index (0–3), difficulty in communication: severe (3), moderate (2), mild (1), none (0).	CF15–CF16 responses
S12SeeingScore	Index (0–3), difficulty seeing: severe (3), moderate (2), mild (1), none (0).	CF2–CF3 responses
S13HearingScore	Index (0–3), difficulty hearing: severe (3), moderate (2), mild (1), none (0).	CF5–CF6 responses
TotalCfmScore	Sum of functional difficulty scores across 13 domains.	Row total of S01–S13 scores
MaxCfmScore	Maximum score among 13 domains.	Row maximum of S01–S13 scores
NoDisability	Binary (0/1), child has no disability.	MaxCfmScore == 0
MildDisability	Binary (0/1), child has mild disability.	MaxCfmScore == 1
ModerateDisability	Binary (0/1), child has moderate disability.	MaxCfmScore == 2
SevereDisability	Binary (0/1), child has severe disability.	MaxCfmScore == 3
"Variable"_Miss0	Corresponding variable, missing treated as 0.	Replace missing values with 0
"Variable"_without	Corresponding variable, excluding anxiety and depression.	Exclude Anxiety_5to17 and Depression_5to17
MicsEnrolled	Binary(0/1), indicating a proxy for a child's school enrollment status.	Constructed if the child is reported to have attended either (i) an educational level classified as primary, or
	(ii) a grade level between 1 and 7 (inclusive) during the current school year.	
Physical	Binary (0/1), physical difficulty (SB)	=1 if moderate/severe difficulty in walking or self-care (CFM)
Lowvision	Binary (0/1), low vision (SB)	=1 if moderate/severe difficulty in seeing (CFM)

Variable Name	Variable Description	Variable Construction
Blind	Binary (0/1), blindness (SB)	=1 if severe difficulty in seeing (CFM)
Lowhearing	Binary (0/1), low hearing (SB)	=1 if moderate/severe difficulty in hearing (CFM)
Deaf	Binary (0/1), deafness (SB)	=1 if severe difficulty in hearing (CFM)
Intellectual	Binary (0/1), intellectual difficulty (SB)	=1 if difficulty in ≥ 2 of: learning, remembering, concentrating
Autism	Binary (0/1), socio-behavioral difficulty (SB)	=1 if difficulty in ≥ 2 of: behavior, social interaction, communication, adapting
Total	Binary (0/1), any disability (SB)	=1 if any domain indicator =1
Rate_Physical_2023	Physical disability rate (ASC)	Share of children with physical disability (ASC 2023)
Rate_Lowvision_2023	Low vision rate (ASC)	Share of children with low vision (ASC 2023)
Rate_Blind_2023	Blindness rate (ASC)	Share of children classified as blind (ASC 2023)
Rate_Lowhearing_2023	Low hearing rate (ASC)	Share of children with low hearing (ASC 2023)
Rate_Deaf_2023	Deafness rate (ASC)	Share of children classified as deaf (ASC 2023)
Rate_Intellectual_2023	Intellectual disability rate (ASC)	Share of children with intellectual disability (ASC 2023)
Rate_Autism_2023	Autism rate (ASC)	Share of children with autism (ASC 2023)
Rate_Total_2023	Total disability rate (ASC)	Share of children with any disability (ASC 2023)
Screen1Total	Students screened in the first round using teacher-administered visual acuity charts.	Direct from VisionCare2025.dta; totals row added via collapse (sum)
Screen2Expected	Students expected to receive second-round screening by visiting ophthalmologists based on first-round results.	Direct from VisionCare2025.dta; totals row added via collapse (sum)
Screen2Actual	Students who were available and received second-round screening	Direct from VisionCare2025.dta; totals row added via collapse (sum)

Variable Name	Variable Description	Variable Construction
ReceivedSpectaclesStudents	Students with visual acuity $\leq 6/12$ who were prescribed and received corrective lenses.	Direct from VisionCare2025.dta; totals row added via collapse (sum)
SeeingDiff	Binary (0/1): child aged 5–17 reports "a lot of difficulty" or "cannot do at all" on the CFM seeing domain (moderate-or-severe threshold).	gen SeeingDiff = Seeing_5to17_IND_Miss0 (pre-constructed from CFM items CF2 and CF3 per WG/UNICEF guidance).
UseOfCorr	Binary (0/1): child currently uses corrective lenses (glasses or contacts).	gen UseOfCorr = CF1 == 1
HearingDiff	Binary (0/1): child reports moderate or severe hearing difficulty on CFM.	gen HearingDiff = Hearing_5to17_IND_Miss0 (from CFM items CF5 and CF6).
UseOfHear	Binary (0/1): child uses a hearing aid.	gen UseOfHear = CF4 == 1
Snellen	Snellen acuity fraction (6/6, 6/7.6, ..., 6/120): ratio of testing distance (m) to the distance at which normal vision resolves the same optotype. Lower denominators = better acuity.	Built from SnlnDnm_m: "6/" + string(SnlnDnm_m) with integer-vs.-decimal formatting via cond().
LogMar	LogMAR acuity $\equiv \log_{10}(\text{SnlnDnm_m}/6)$. Standard clinical scale where 0 = normal (6/6), $0.3 \approx 6/12$, $1.0 \approx 6/60$.	gen LogMar = -x.
MPS, phi2, phi3	Gompertz reading-speed parameters calibrated from Cheung et al. (2008) / Calabrèse et al. (2016): maximum print size (MPS), rate (phi2), and critical-print-size intercept (phi3).	Estimated via nl (y = {MPS=225}/exp(1)exp(1 - exp(-{phi2=2.3}(x-{phi3=0}))))); saved as scalars.
ArAt3m / ArAt6m / ArAt9m	Acuity reserve at viewing distance d (eq. 2.6 in paper): ratio of 3.4 cm blackboard-letter height to the optotype size subtending a 5-arcminute visual angle at d.	$0.034 / (2^* (\text{SnlnDnm_m} * d / 6) \tan(5/60/2_pi/180))$ for d $\in \{3, 6, 9\}$ m.
LogMarAt3m / LogMarAt6m / LogMarAt9m	LogMAR-equivalent acuity reserve at each distance, expressed relative to the $10^{0.2}$ threshold.	$\log(\text{ArAt}\{d\}m / 10^{0.2}) / \log(10)$.

Variable Name	Variable Description	Variable Construction
RsPctAt3m / RsPctAt6m / RsPctAt9m	Predicted reading speed as % of maximum (eq. 2.7): Gompertz function of acuity reserve. Benchmarks: 25% = threshold reading, 50% = survival reading, 90% = fluent/sustained classroom reading.	$1/\exp(1) * \exp(1 - \exp(-\phi^2 * (\text{LogMarAt}\{d\}m - \phi_3)))$, formatted with 3 significant digits and % suffix.