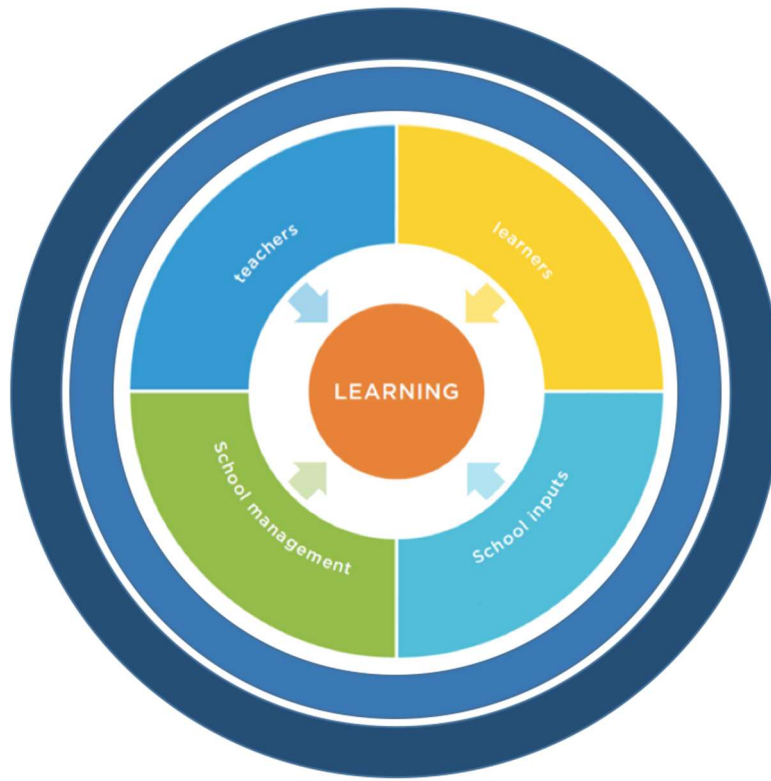


**Technical Note**

**Last Updated: April 19, 2021**

**Global Education Policy Dashboard**



## Table of Contents

Introduction .....	3
Section 1 – Development of the GEPD Survey Instruments .....	4
Developing the GEPD 1st-Grade Assessment.....	5
Developing the GEPD 4th-Grade Assessment .....	7
Developing the GEPD Teacher Assessment .....	8
Section 2 – Technical Specifications of GEPD Field Work.....	10
Sampling .....	10
Survey Weights .....	14
Power Considerations .....	15
Automated Data Collection.....	19
Anonymization Process .....	20
Data Quality Checks.....	21
Section 3 – Indicator Construction.....	21
Proficiency Cutoffs.....	22
1st-Grade Assessment .....	22
4th-Grade Assessment Cutoffs .....	22
Teacher Assessment Cutoffs .....	23
Question Scoring and Calculation of Indicators .....	23
Annex 1: Detailed Indicator Calculation.....	38

## Introduction

Many countries, despite having significantly increased access to education for their children and youth, now realize that they are facing a learning crisis. In low- and middle-income countries, where enrollment in primary school is nearly universal, even before the COVID-19 pandemic hit, 53% of children suffered from Learning Poverty—meaning that they could not read and understand a short age-appropriate story by age 10. This reality underlines that schooling is not the same as learning, even though education policy often assumes that it is. The learning crisis has only deepened with the extended school closures and the sharp recessions caused by the pandemic.

The *World Development Report 2018* argued that the learning crisis has multiple causes: poor service delivery in schools and communities, policies that are not aligned toward learning for all, and unhealthy politics and low bureaucratic capacity. To tackle the crisis and improve student learning for all, countries need to know where they stand on these three key dimensions—practices (or service delivery), policies, and politics. But providing such a systemwide overview requires better measurement. Many drivers of learning are not captured by existing administrative systems. And although new measurement tools capture some of these drivers well, no single instrument pulls together data on all these areas. This gap leaves policymakers in the dark about what is working and what isn't.

To fill this gap, the World Bank, with support from the Bill and Melinda Gates Foundation, the UK's Department for International Development, and the Government of Japan, has launched a Global Education Policy Dashboard, which measures the key drivers of learning outcomes in basic education around the world. In doing so, the GEPD highlights where systems are falling short in providing quality education for all children, identifies gaps between current practice and what evidence suggests would be most effective in promoting learning for all, and helps governments in setting priorities and tracking progress as they work to close those gaps.

This technical note outlines some technical aspects of the overall project as well as its field work, instruments, indicators, and scores. The purpose is to be transparent about the technical decisions that have been made in collecting the data and developing the indicators of the Global Education Policy Dashboard. The World Bank is committed to continuing to improve the Dashboard, so comments on these materials are very welcome. (See the link on the website, [www.EducationPolicyDashboard.org](http://www.EducationPolicyDashboard.org)) The note is divided into 3 sections that cover the: 1) development of the instruments, 2) field work, and 3) indicator computation. These three sections are followed by an annex that provides links to additional relevant information.

## Section 1 – Development of the GEPD Survey Instruments

The Dashboard project collects new data in each country using three new instruments: A School Survey, a Policy Survey, and a Survey of Public Officials. Data collection involves school visits, classroom observations, legislative reviews, teacher and student assessments, and interviews with teachers, principals, and public officials. In addition, the project draws on some existing data sources to complement the new data it collects. A major objective of the GEPD project was to develop focused, cost-effective instruments and data-collection procedures, so that the dashboard can be inexpensive enough to be implemented (and re-implemented) in many countries. The team achieved this by streamlining and simplifying existing instruments, and thereby reducing the time required for data collection and training of enumerators.

More information pertaining to each of the three instruments can be found below:

- **School Survey:** The School Survey collects data primarily on Practices (the quality of service delivery in schools), but also on some *de facto* Policy indicators. It consists of streamlined versions of existing instruments used by the World Bank and partners—including Service Delivery Indicators (SDI) Surveys on teachers and inputs/infrastructure, TEACH on pedagogical practice, Global Early Child Development Database (GECDD) and Measuring Early Learning Quality and Outcomes (MELQO) on school readiness of young children, and the Development World Management Survey (D-WMS) on management quality—together with new questions to fill gaps in those instruments. Though the number of modules in the School Survey is similar to the full version of the Service Delivery Indicators (SDI) Survey, the number of items and the complexity of the questions within each module have been reduced to streamline the survey, while additional items and assessments have been added from other instruments. The School Survey includes 8 short modules: School Information, Teacher Presence, Teacher Survey, Classroom Observation, Teacher Assessment, 1<sup>st</sup>-Grade Direct Assessment, School Management Survey, and 4th-Grade Student Assessment.
- **Policy Survey:** The Policy Survey collects information to feed into the Policy *de jure* indicators. This survey is filled out by key informants in each country, drawing on their knowledge to identify key elements of the policy framework (as in the SABER approach to policy-data collection that the Bank has used over the past 9 years). The survey includes questions on policies related to teachers, school management, inputs and infrastructure, and learners.
- **Survey of Public Officials:** The Survey of Public Officials collects information about the capacity and orientation of the bureaucracy, as well as political factors affecting education outcomes. This survey is a streamlined and education-focused version of the civil-servant surveys that the Bureaucracy Lab (a joint initiative of the Governance Global Practice and the Development Impact Evaluation unit of the World Bank) has implemented recently in

several countries. The survey includes questions about technical and leadership skills, work environment, stakeholder engagement, impartial decision-making, and attitudes and behaviors.

To guide the streamlining process and the development of the three instruments, the project included rigorous qualitative research to identify the most relevant questions on which to base the indicators. Throughout this process, numerous experts within and outside the World Bank were consulted. This process alone resulted in a significant reduction in terms of the length and complexity of the surveys. Additionally, for some of the modules of the School Survey, extensive psychometric analyses of existing data were conducted to inform the streamlining process. These modules were primarily those involving assessments – 1st-Grade Direct Assessment, 4th-Grade Assessment, and Teacher Assessment.

### **Developing the GEPD 1st-Grade Assessment**

The GEPD 1st-Grade Direct Assessment is based on instruments from the MELQO initiative. The MELQO initiative was produced by a partnership of UNESCO, the World Bank, the Center for Universal Education at the Brookings Institution, and UNICEF that began in 2014. The MELQO instrument was designed for 4- to 6-year-old children and measures early childhood development along several dimensions. MELQO has been fielded in a number of countries around the world.

The GEPD 1st-Grade Assessment was developed by streamlining the MELQO Direct Assessment Module. A number of alterations were made to shorten the length of the instrument as well as to adapt the assessment to slightly older children. Whereas MELQO targeted 4- to 6-year-olds, the GEPD targets 1st-grade students, who are typically 6 to 8 years old. For the psychometric analysis on which the streamlining was based, the team used MELQO data previously collected in Peru in 2017, Mongolia in 2016, Lao PDR in 2016, Tanzania in 2017, and Ethiopia in 2018.

Using that data, the team estimated item difficulty and discrimination parameters using a two-parameter logistic model in order to refine the set of items. In addition to the IRT analysis, a number of other considerations factored into which items to include in the GEPD 1st-Grade Assessment, including feedback from WBG ECE specialists on whether each item passed a face validity check. For instance, items that would be hard to explain to an education minister were dropped, as were items that would be too difficult to adapt to new countries, whether because of translation or cultural differences. Another consideration was whether enumerators would have difficulty implementing each item, either because the instructions were difficult to convey to children or because extensive or specialized training was needed for the enumerator. Finally, items that were targeted to younger children than expected in 1st-grade classrooms were excluded.

The resulting GEPD 1st-Grade Assessment contains four domains—literacy, numeracy, executive functioning, and socio-emotional skills—that are broken into the following sub-domains:

Literacy:

- Expressive Vocabulary

- Letter Identification
- Word Recognition
- Sentence Reading
- Listening Comprehension Story
- Name Writing
- Print Awareness

Numeracy:

- Verbal Counting
- Producing a Set
- Number Identification
- Number Comparison
- Simple Addition

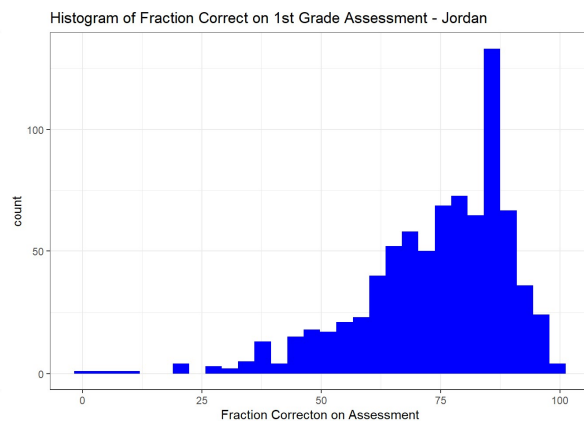
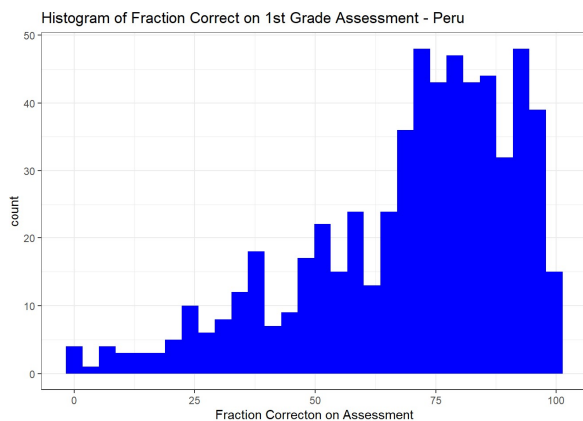
Socio-emotional:

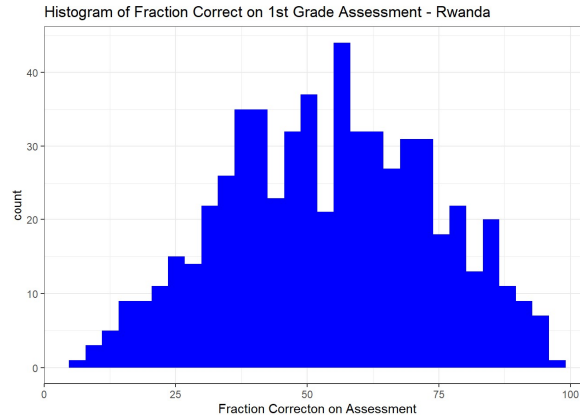
- Perspective-Taking/Empathy
- Conflict Resolution

Executive functioning:

- Working Memory/Backward Digit Spans
- Follow Instructions/Head, Toes, Knees, Shoulders Task

Even with the adaptations described above, there were concerns about ceiling effects with these 6- to 8-year-old children, but data evaluation from the first three countries where data was collected (Peru, Jordan, and Rwanda) shows no indication of such effects. See the histograms below, which show the percentage of the items correct on the GEPD 1st-Grade Direct Assessment. Very few students either get all of the questions wrong or all correct. In Peru, there is some rightward skew to the distribution, but only 2 students out of 603 who took the assessment got perfect scores.





## Developing the GEPD 4th-Grade Assessment

The GEPD 4th-Grade Assessment is based on the Student Assessment module from the Service Delivery Indicators (SDI) Survey Instruments. The SDI survey has been fielded in many countries in Africa, including Kenya in 2012, Morocco in 2016, Madagascar in 2016, Mozambique in 2014 and 2018, Nigeria in 2013, Niger in 2015, Senegal in 2010, Togo in 2013, Tanzania in 2014 and 2016, and Uganda in 2013. The original instrument included 15 mathematics questions and 13 literacy questions, and the assessment was conducted one-on-one, with the enumerator assessing each student individually.

The GEPD 4th-Grade Assessment incorporates some adaptations to the SDI model and questions. First, the assessments were designed to be conducted in a group setting rather than one-on-one, to simplify data collection for the enumerators and reduce costs. Second, 11 new literacy items were added to align the instrument better with the Global Proficiency Framework (GPF) developed by a multiagency partnership for the purpose of monitoring progress toward Sustainable Development Goal 4.1. These items were selected from among the publicly released PIRLS items for 4<sup>th</sup> grade.<sup>1</sup> More work is being done in collaboration with the UNESCO Institute of Statistics to further refine this assessment and better align it with the GPF.

In total, the 4th-Grade Assessment includes the following items:

### 24 Literacy Items:

- Letter Identification (3 items)
- Word Recognition (7 items)
- Reading Comprehension Story (3 items)
- Reading Comprehension Story II (11 items)

### 15 Numeracy Items:

---

<sup>1</sup> In Peru, we did not include these additional PIRLS items, so only 13 literacy items were included for this country.

- Number Sense (4 items)
- Arithmetic (11 items)
- Word Problem (1 item)
- Sequences (1 item)

## Developing the GEPD Teacher Assessment

The GEPD Teacher Assessment is based on the SDI Teacher Assessment, which has two main objectives. First, it aims to examine whether teachers have mastery of the subjects they are teaching. This is interpreted as the minimum knowledge required for the teacher to be effective. Second, it also examines the extent to which teachers demonstrate mastery of subject content skills that are above the level they are teaching at.

To simplify administration, the teacher assessment was designed as a marking exercise, in which teachers are asked to mark and correct a hypothetical student's exam. There are two versions of the teacher assessment, one for language and one for mathematics. The language assessment is administered to teachers teaching language (or language and other subjects), and the mathematics assessment is administered to teachers teaching mathematics (or mathematics and other subjects). The original SDI assessment was validated against 13 Sub-Saharan primary curricula.<sup>2</sup>

The original SDI language assessment consists of two sections. The first section asks teachers to assess pupil language literacy by correcting a primary school pupil language assessment. The teachers are asked to mark whether the “student answer” is correct and, if it is incorrect, to write the correct answer. The second section asks teachers to correct a letter written by a child in 4th grade. The teachers have to correct the letter for grammar, punctuation (between sentences and within sentences), spelling, syntax, and salutation by circling the mistakes and writing the correction on the line. The mathematics assessment asks teachers to assess pupil numeracy literacy by correcting a primary school pupil mathematics test. Again, the teachers have to indicate whether the “student answer” is correct and, if not, to write the correct answer. The original SDI instrument included 22 language questions, 23 math questions, and a pedagogical assessment.

The GEPD Teacher Assessment includes some adaptations to the SDI assessment. First, because the GEPD measures teachers' pedagogical skills directly through classroom observation, the pedagogical assessment is excluded from the GEPD Teacher Assessment. Second, the correct-the-letter exercise that was part of the literacy assessment was dropped, as it was very difficult to maintain comparability in different languages. Third, the team analyzed the statistical performance of the items using IRT methods to determine if any questions needed to be replaced due to low

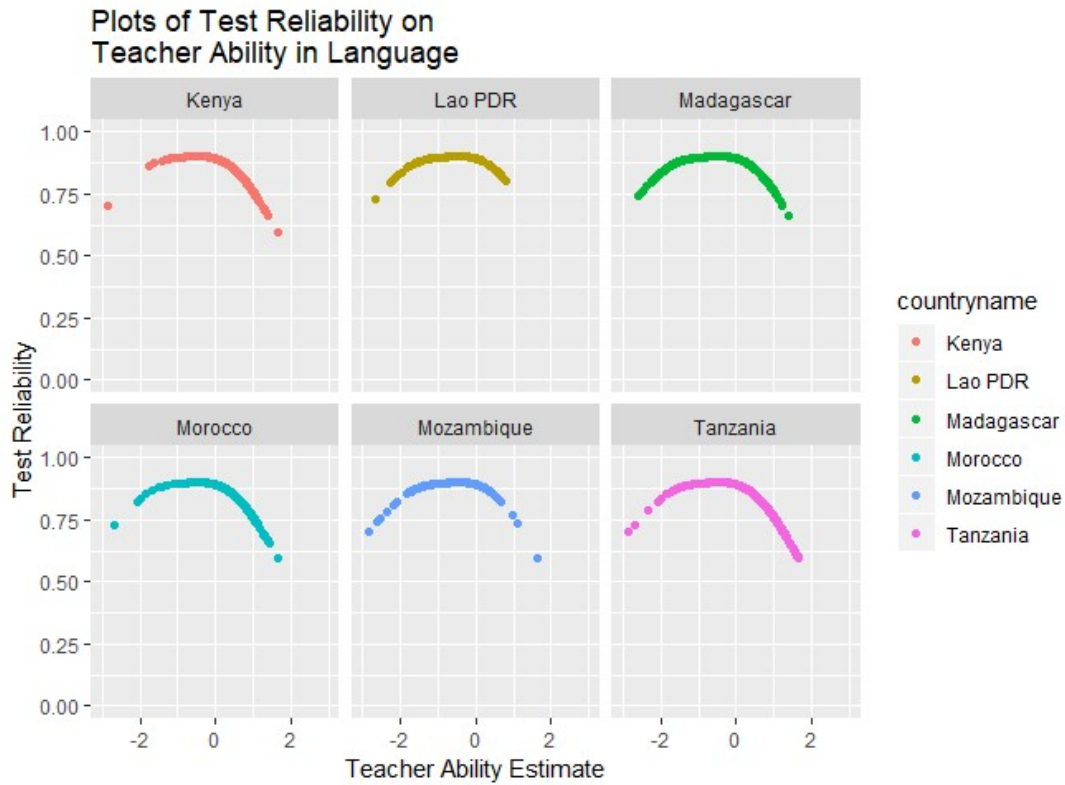
---

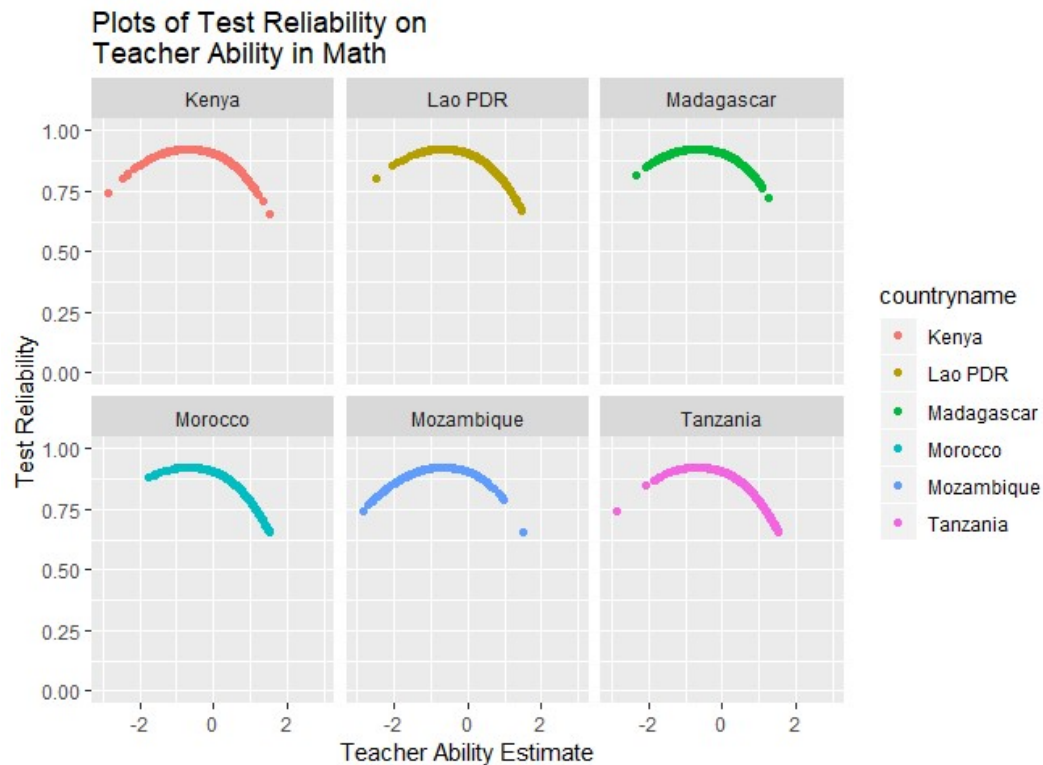
<sup>2</sup> The countries included for the review were: Botswana, Ethiopia, Gambia, Kenya, Madagascar, Mauritius, Namibia, Nigeria, Rwanda, Seychelles, South Africa, Tanzania, and Uganda. See David Johnson, Andrew Cunningham and Rachel Dowling (2012), “Teaching Standards and Curriculum Review”, prepared as background document for the SDI Survey. While some SDI countries were not part of the review, before doing the SDI survey the SDI team verified that the school curriculum was compatible with the SDI teacher and student test. For more information see SDI Technical report for each specific country.



performance. Using a two-parameter logistic IRT model, data was analyzed from Kenya, Mozambique, Lao PDR, Madagascar, Morocco, and Tanzania to estimate item discrimination and difficulty parameters.

The figures below show the test reliability of the original SDI survey in these countries. The reliability for the average teacher in math and language in each country ranged from 0.87 to 0.91 in math and 0.85 to 0.90 in language; all of these exceed the 0.8 that is commonly used as a rule of thumb to indicate a well-performing set of items.





Based on these analyses of SDI data, the GEPD instruments incorporate some small adaptations to ensure that the assessment can discriminate sufficiently between teachers at the top of the achievement distribution. As can be seen in the graphs, the SDI test reliability is much lower for teachers above 1 standard deviation the mean in the country. With this in mind, the GEPD team replaced two of the easier mathematics items in the SDI test with two more difficult items.

## Section 2 – Technical Specifications of GEPD Field Work

The aim of the data collection is to produce high-quality data that allow construction of nationally representative indicators, with sufficient precision to allow detection of differences across time and by urban/rural location and gender. The GEPD does this by using electronic in-person data collection and automating the data collection process and quality checks. The following sections describe the sampling strategy and the technical specifications of the data collection protocols and processes.

### Sampling

The aim of the GEPD surveys is to produce nationally representative estimates with enough precision to allow detection of changes over time at a minimum power of 80% and at a 0.05

significance level. The GEPD also sets aims to detect differences by urban/rural location and by gender on relevant indicators. In some cases, we can provide statistics by region, but this will depend on the geography of the country and the number of regions making up that country.

For the GEPD School Survey, a two-stage random sample design is used. In the first stage, Bank staff select a random sample of around 200 schools.<sup>3</sup> The sample is stratified based on urban/rural classification and the region in which schools are located. When stratifying by region, the GEPD team works with partners within the country to make sure all relevant geographical divisions are included. In the second stage, a random sample of teachers and students is selected to answer questions from the survey modules; this sampling is done by enumerators in the field at each school. Ten teachers per school are sampled for attendance checks, and five teachers are interviewed and given a teacher assessment. Three randomly selected 1<sup>st</sup> grade students are assessed, as is a randomly selected classroom of 4<sup>th</sup> grade students. We could only interview three 1<sup>st</sup> grade students, compared to an entire 4<sup>th</sup> grade classroom, because the 1<sup>st</sup> grade assessment is done face to face by our enumerators and takes a considerable amount of time per student (~15-20 minutes). The 4<sup>th</sup> grade assessment can be given to an entire class simultaneously.

For the GEPD Survey of Public Officials, 200 public education officials in each country are randomly selected for interviews. These public officials are typically professional staff at the Ministry of Education central office, as well as from the regional or district offices. Roughly 60 officials are surveyed at the federal (or central-government) level, while 140 officials are surveyed at the regional/district level.<sup>4</sup> To select officials at the regional and district level, the team employs a cluster sampling strategy, where 10 regional offices are chosen at random from among the regions in which schools were sampled. Among these 10 regions, 10 districts are selected (one in each region) from among the districts in which schools are sampled. The result of this sampling approach is that for 10 clusters, the GEPD captures the links from the school to the district office, to the regional office, and then to the central office. In each regional/district office, 7 officials are sampled: the head of the office, the HR director, and 5 officials working in finance and planning chosen at random.<sup>5</sup> At the federal level, the GEPD team works with the Ministry of Education to identify the offices that should be included in the sample according to the functions they serve. In each office, the director of the office is interviewed along with a randomly selected number of public officials. The team aims to maintain roughly the same sampling strategy for the Survey of Public Officials for all countries, with some adaptation to country characteristics. For instance, in countries with smaller district offices, a larger number of district offices will be sampled, with fewer public officials interviewed at each of them. In countries with smaller offices at the federal level, fewer public officials will be interviewed at the federal level and more will be interviewed

---

<sup>3</sup> The exact sample size for any specific country depends on local conditions and funding.

<sup>4</sup> Based on analysis of the data from the first four countries, the number interviewed at the federal level may be reduced in future applications of the survey.

<sup>5</sup> We chose to do the random sampling in the field, where enumerators enter a staff list and the tablet randomly chooses among the list. An alternative would have been to rely on the HRMIS staff lists and choose in advance, but we had concerns that the staff list would become outdated due to staff turnover and so chose the former approach.

at the decentralized level. Ultimately, these adaptations are the result of extensive dialogue with the country counterparts.

To develop this methodology, several sampling experts at the World Bank were consulted on the best strategies for school sampling. As mentioned, the goal of the GEPD School Survey is to be able to detect differences across time and by urban/rural location. Based on the consultation with experts, our team uses a stratified sampling approach. In cases where we have access to rich data collected prior to our survey, our team may use the optimal stratified sampling technique in Barcaroli (2014), which uses the genetic algorithm to produce an optimal number of strata and an optimal number of units within strata.<sup>6</sup> In our initial set of pilot countries, only Peru contained the prior data needed to use the optimal stratification approach. This will be discussed further in the discussion on further sampling considerations. Optimal stratification can reduce sampling error compared to alternative approaches such as cluster sampling or simple random sampling.<sup>7</sup> After stratification, we will reweight as appropriate to produce a nationally representative estimate for each indicator. This will be discussed more below.

Before producing the sample using the stratification approach, the sampling frame is defined. Typically, the sampling frame includes all schools with at least three 1st-grade and at least three 4th-grade students, according to the latest school census. This minimum size was set to ensure that enough teachers and students would be interviewed on the limited budget available for each country.<sup>8</sup> To develop this sampling frame, the GEPD team works with the World Bank country teams and the country counterparts to compile an up-to-date and detailed database containing information on schools. Depending on the country, the sampling frame may include private schools, typically because a large share of students attends private schools. This sampling decision is jointly made by World Bank and country counterparts. As an example, in Peru only public schools were sampled, because private schools represent a small share of the schools in the

---

<sup>6</sup> Barcaroli G (2014). “SamplingStrata: An R Package for the Optimization of Stratified Sampling.” *Journal of Statistical Software*, 61(4), 1–24. <http://www.jstatsoft.org/v61/i04/>.

<sup>7</sup> The optimal stratification approach is particularly useful if relevant data on schools or students is available from previous surveys. In such cases, the optimal stratification algorithm will assign extra units to any stratum that has particularly high variance in the target variables. For instance, in Peru, data was used from a previous nationally representative MELQO survey on early childhood development, which closely resembles the GEPD 1st-Grade Direct Assessment. Using this data, the mean and standard deviation of the student-level test scores by department were calculated. Based on these results, extra sampling units were assigned to strata with relatively high variances in these outcomes. To the extent that regions with a high variance in the MELQO test also have a high variance in other outcomes (such as the 4th-Grade test scores, teacher knowledge, principal practices), the expectation is that this sampling approach improves the precision in these measures as well.

A quick introduction to the technical aspects of the optimal stratification approach can be found here: <https://cran.r-project.org/web/packages/SamplingStrata/vignettes/SamplingStrata.html>

<sup>8</sup> In Jordan, this restriction to include only schools with at least 3 1<sup>st</sup>- and 4<sup>th</sup>-grade students resulted in 30 schools (and 106 4<sup>th</sup>-grade students) being dropped from the sampling frame, out of 3313 schools. In Peru, this restriction eliminated 4.7% of 4<sup>th</sup>-grade students in these smaller schools. In Rwanda, we lacked data to apply this filter, so no schools were dropped because of these limits.

country. Alternatively, in Jordan, the percentage of private schools was larger, and so a decision was made to include both public and private schools.

For the stratified sampling, the stratification variables used are the rural/urban status of each school as well as the 1<sup>st</sup>- or 2<sup>nd</sup>-level administrative division (while the denomination changes, in most cases, the 1<sup>st</sup> administrative division is the province/department-level and the 2<sup>nd</sup> division corresponds to districts). For example, in Peru, the department and urban/rural status of the school made up the stratification variables. In cases where private schools are also included in the sampling frame, as in Jordan, the public/private status of the school is also included as a stratification variable.

In each country, the sampling strategy is slightly customized to reflect ongoing efforts and meet country needs. Table 1 offers some examples of how samples have been customized.

**Table 1: GEPD Sampling Customization**

Country	Sample
Peru	MELQO data was merged with the Peru school sampling frame to allow optimal stratification. The stratification was done on the basis of urban/rural location and department. There are 25 departments in Peru. In 2017, Peru conducted an examination of around 4,500 children aged 5-8, with a median age of 6. The MELQO assessment is quite similar to the GEPD 1st-Grade Direct Assessment. Using data from this 2017 survey, the team calculated means and standard deviations by province and fed this information into the optimal stratification algorithm. Provinces with low standard deviations among students in terms of their MELQO development scores are allocated fewer schools compared to an allocation that is simply based on population, and provinces with high standard deviations are allocated more schools. 205 schools were chosen for the GEPD School Survey after optimally stratifying.
Jordan	For the GEPD School Survey, both public and private schools that are supervised by the Ministry of Education were included in the sampling frame. Schools supervised by Ministry of Defense, Ministry of Endowments, Ministry of Higher Education, or Ministry of Social Development were excluded from the sampling frame. This left a sampling frame containing 3,330 schools, with 1297 private schools and 2003 public schools. Schools kept needed to have at least three 1 <sup>st</sup> -grade students, three 4 <sup>th</sup> -grade students, and three teachers. Southern schools were oversampled (to reach a total of 50 Southern schools) to allow regional comparisons. Additionally, second-shift (evening) schools were oversampled, for a total of 40 second-shift schools, to allow reporting on this unique type of school. Second-shift schools make up around 16% of our total sample, while they make up 7.6% of the schools in our sampling frame. When producing a national estimate, we reweight the schools to be reflective of their totals in the population. 250 schools were sampled for the School Survey after the discussed adaptations.

Rwanda	In the case of Rwanda, the GEPD team tested the possibility of each field team visiting 2 schools per day. To allow visits to two schools per day, the team clustered at the sector level and chose two schools per cluster. With a sample of 200 schools, this means that 100 PSUs had to be allocated. This clustering was combined with stratification by district and by the urban/rural status of the schools. The number of PSUs allocated to each stratum is proportionate to the number of schools in each stratum (i.e., the district X urban/rural status combination).
--------	---

## Survey Weights

Survey weights are constructed using the original sampling frame for each country. School weights are the inverse probability that a 4th-Grade student in the school is randomly selected for the school survey. The exact formula depends on the selection procedure for choosing schools. In nearly all cases, there is some form of stratification, in which case the probabilities referenced above are computed within each stratum. The weight for school  $i$  within stratum  $j$  is, where  $m$  is a measure of school size and  $n$  is the number of schools selected per stratum:

$$SW_i^j = \frac{\sum_{i=1}^{N_j} m_i}{n * m_i}$$

In a simple case where schools are selected at random proportional to size, this is calculated by summing the total student enrollment in the school and dividing by the total enrollment of 4th-Grade students in the sampling frame. In cases where there are many non-responses or enumerators are denied entry to the school, an adjustment term to account for non-response is added. This could happen for instance, if an originally sampled school was selected but refused to allow entry and then the two replacement schools for that school also refused.

For student-level or teacher-level data, the team also adjusts for the random selection of classrooms, teachers, or students that takes place within schools. To derive these weights at the individual level, the school weight is multiplied by the number of units sampled (five teachers for the teacher interview, 10 for the teacher absence module, one for the class of 4<sup>th</sup>-graders, and three for the three randomly selected 1st-grade students) divided by the total number of units available to be sampled in the school.

After the stratum weights are formed, to produce a nationally representative mean for an indicator, we produce a mean within each stratum and then average these stratum means with weights based on the size of the stratum:

$$\bar{X} = \frac{1}{N} \sum_{j=1}^N N_j * \bar{X}_j$$

where  $\bar{X}_j$  is the weighted mean for stratum  $j$ ,  $N_j$  is the size of the stratum and  $N$  is the total size at the national level. Standard errors are calculated as follows:

$$SE(\bar{X}) = \sqrt{\sum_{j=1}^N \left(\frac{N_j}{N}\right)^2 \left(\frac{N_j - n_j}{N_j}\right) \frac{\hat{s}_j^2}{n_j}}$$

where  $\hat{s}_j^2$  is the sample standard deviation in stratum  $j$  and  $n_j$  is the number of sampled units in stratum  $j$ .

## Power Considerations

All surveys must balance the cost of conducting the survey with the resulting precision of the estimates. One goal of the GEPD survey was to provide a snapshot of a comprehensive set of indicators at low cost and to be able to track progress over time. The GEPD team set a target of around \$150,000 for the total cost of data collection, which includes the training, enumeration, travel, and cleaning of the data necessary to complete the survey. With this budget, the GEPD team was able to visit roughly 200 schools in Peru and Rwanda, and 250 schools in Jordan. The data collection also includes an interview of roughly 200 public officials in the country. The number of schools in a GEPD survey is roughly comparable, if not larger, than the number of schools typically visited for TIMSS and PIRLS surveys.<sup>9</sup>

While survey cost is an important consideration, the minimum detectable effect size is also important for understanding the tradeoff between budget and precision. Any effect size below the minimum detectable effect size will not be statistically distinguishable from an effect of zero. In this section, we examine how well the GEPD survey will be able to detect changes in an indicator over time.

Existing data from surveys that use a similar methodology to the GEPD were incorporated in order to assess the likely detectable effect sizes for some of the GEPD key indicators. The team examined the likely detectable effect sizes for the following indicators: teacher absence teacher content knowledge, 4th-grade student knowledge, early childhood development, and teacher pedagogical skills.

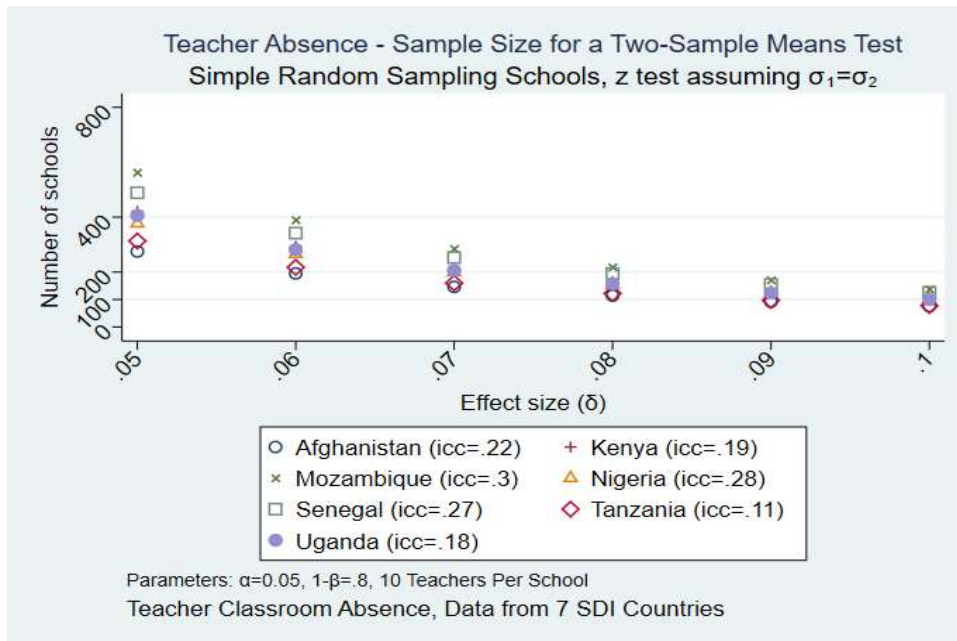
For teacher absence, the relevant data was previously collected as part of the Service Delivery Indicators (SDI) survey (<https://www.sdindicators.org/>). Data came from the countries of Afghanistan (SABER SD), Tanzania, Kenya, Mozambique, Nigeria, Uganda, and Senegal. In each country, data on teacher absence was collected using a very similar questionnaire. One difference is that for the SDI survey, absence data was collected during a second unannounced visit to the school. In the GEPD's case, the school was notified that a team of enumerators would come to the school within a two-week period, but schools were not provided with an exact date for the visit. Setting aside this difference, the team was able to calculate the intra-class correlation (ICC) in

---

<sup>9</sup> The typical school survey for TIMSS and PIRLS contains roughly 150 schools and a student sample of 4,000 students. [https://timssandpirls.bc.edu/methods/pdf/tp\\_sampling\\_design.pdf](https://timssandpirls.bc.edu/methods/pdf/tp_sampling_design.pdf)

teacher absence within the same school for each country and then use these ICCs to calculate expected detectable effect sizes by number of schools sampled. With a benchmark sample of around 200 schools, the expected detectable difference in teacher absence between two years is 6-8 percentage points.

The 2010 and 2014 rounds of the SDI survey in Tanzania give us some sense of whether a 6-8 percentage point effect size is reasonable for detecting changes in absence in practice.<sup>10</sup> Between the 2010 and 2014 rounds of the SDI survey, school absence dropped 9 percentage points from 23% to 14%. Teacher classroom absence dropped 7 percentage points from 53% to 46%, which would be in the range of what we would be able to say is a statistically significant difference.

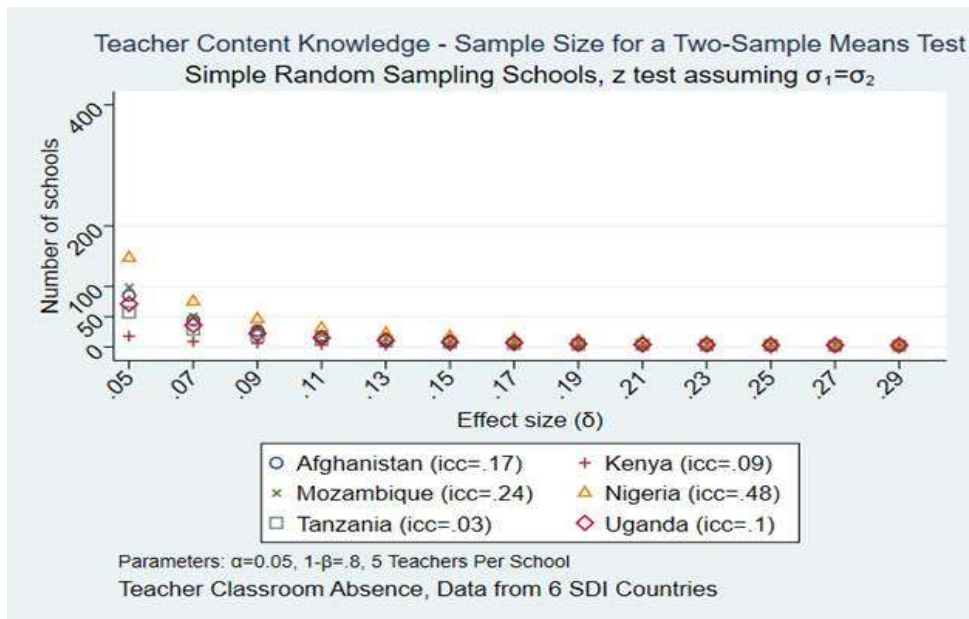


For teacher content knowledge, the same data sources and methodology were used. Assuming a range of intra-class correlations similar to that found in previous SDI surveys, with a sample of 200 schools, the expected detectable change in Teacher Content Knowledge between two years is of 0.05 standard deviations. While there may be disagreements about what effect sizes are small or large, and answering this question requires important context, according to the classifications in Kraft (2020), a 0.05 standard deviation impact would translate into a borderline small/medium size effect.<sup>11</sup> Any change larger than this 0.05 standard deviation threshold would be detectable by our study.

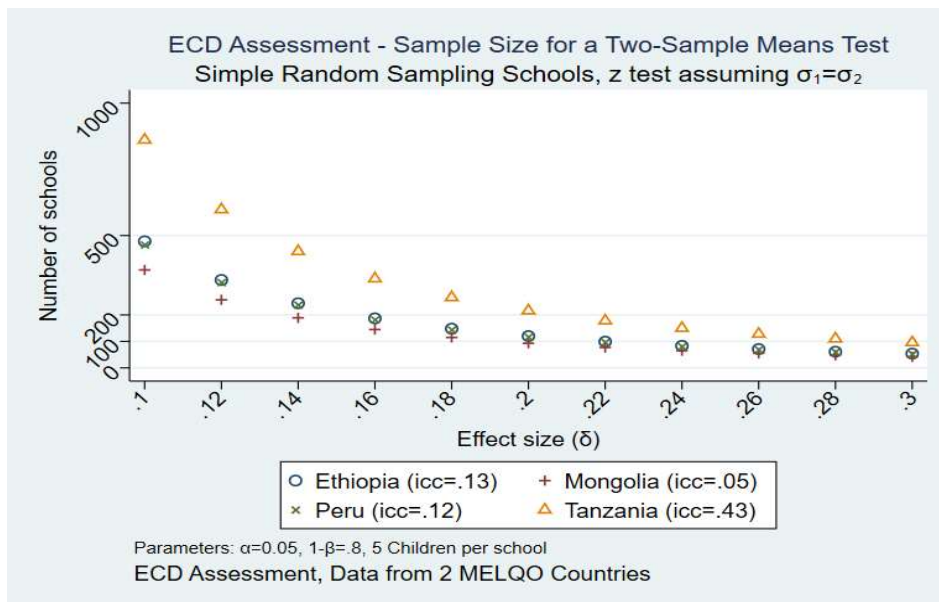
<sup>10</sup> "World Bank Group. 2015. Education Service Delivery in Tanzania. World Bank, Washington, DC. © World Bank. <https://openknowledge.worldbank.org/handle/10986/24797> License: CC BY 3.0 IGO."

<sup>11</sup> Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4), 241-253.





For the GEPD 1st-Grade Direct Assessment, the team used direct assessment data from the MELQO survey in four countries: Ethiopia, Peru, Mongolia, and Tanzania. The ICCs for these countries ranged from 0.05 in Mongolia to 0.43 in Tanzania. Assuming a sample size of 200, the expected detectable change in early grade knowledge is between 0.14 and 0.2 standard deviations. According to Kraft (2020), a 0.14 to 0.2 standard deviation effect would be classified as a medium effect size.



Finally, for teacher pedagogical skills, the relevant data came from the Teach survey in Mozambique and the Punjab province of Pakistan. Based on data from these countries, with a

sample of 200 schools the expected detectable change in teacher pedagogy is between 0.16 and 0.18 units on the 1-5 scale of teacher pedagogical performance.

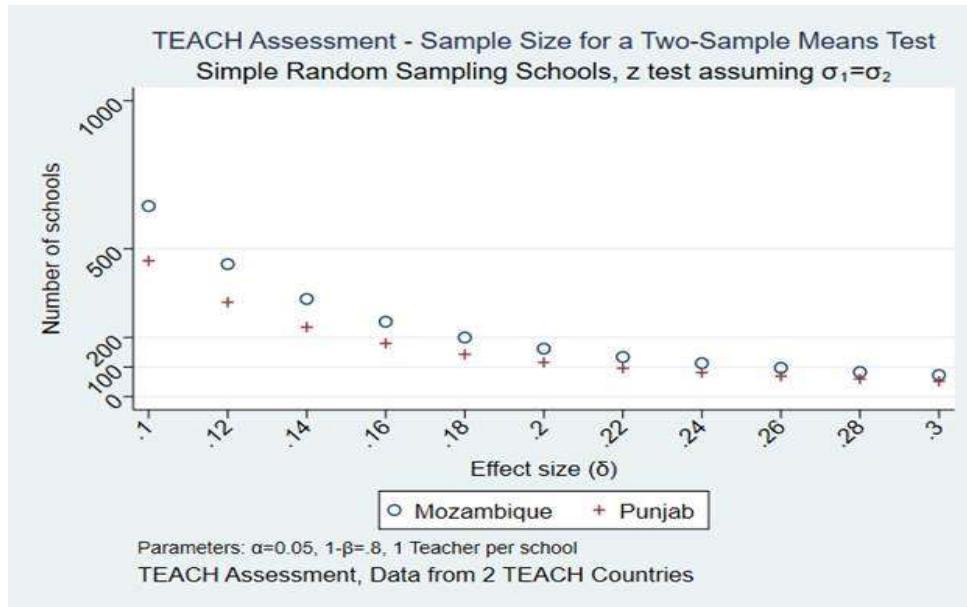


Table 2. Minimum Detectable Effect sizes for Select Indicators.

Indicator	Minimum Detectable Effect
Teacher Absence	6-8 percentage points
Teacher Content Knowledge	0.05 standard deviations
1 <sup>st</sup> Grade Assessment	0.14 - 0.2 standard deviations
Teacher Pedagogical Skills (1-5 scale)	0.16 - 0.18 units

The precision assumptions implied by the power analysis can be compared to data actually collected from GEPD School Surveys in Peru, Jordan, and Rwanda in 2019 and 2020. As discussed in Table 1, the school samples included 205 schools in Peru, 250 schools in Jordan, and 200 schools in Rwanda. While the power analyses were primarily concerned with the detectable effect size over a two-year period, data from one year can be used to assess whether projections were on target.

Assuming that the sampling variance of the indicators is constant across time, one can project the detectable effect size under this assumption. Let the following be the standard error of the difference in means of one of the indicators, where  $\bar{X}_j$  is the sample mean for year j.

$$SE(\bar{X}_1 - \bar{X}_0) = \sqrt{(SE(\bar{X}_1) + SE(\bar{X}_2))}$$

If we assume the precision of the two estimators is the same, (i.e.  $SE(\bar{X}_1) = SE(\bar{X}_2)$ ) the Standard Error of the difference of the means of the two samples is:

$$SE(\bar{X}_1 - \bar{X}_0) = \sqrt{(2 * SE(\bar{X}_1))}$$

In Peru, for teacher presence, the GEPD assessed the presence of 1461 teachers in the sampled schools. The mean presence rate for this sample was 90.68, with a standard error of 0.98. Therefore, the standard error for the two-means difference is just  $\sqrt{2}$  times this standard error, or 1.39. Finally, to determine the detectable effect, one can reject the null of an identical mean for teacher presence if the presence measure in the second Peru survey differs by more than 5.48 percentage points from the original mean of 91.17. In Jordan, one can reject the null of an identical mean for teacher presence if the presence measure in the second Jordan survey differs by more than 4.6 percentage points from the original mean of 81.21. And in Rwanda, one can reject the null of an identical mean with a difference of more than 5.1 percentage points from the original mean of 80.08.

This suggests that in the case of Peru, Jordan, and Rwanda, forecasts of detectable effect sizes were slightly too pessimistic, as the expectation was for a detectable size of 6-8 percentage points. This difference could be attributed to lower-than-typical intra-class correlations in these countries for absence or to improved precision due to the optimal stratification technique.

For teacher content knowledge, conducting the same exercise, one can reject the null of an identical mean for teacher content knowledge if the measure in the second survey differs by more than the following amounts from the first-survey mean: for Peru, 10.47 percentage points; for Jordan, 5.58 percentage points; and for Rwanda, 6.09 percentage points.

For the GEPD 1st-Grade Direct Assessment, one can reject the null of an identical mean for early childhood education (ECE) if the following changes in means are detected in the means from the second survey: for Peru, 15.98 percentage points; for Jordan, 10.3 percentage points; and for Rwanda, 5.32 percentage points.

### Automated Data Collection

An important element of the Global Education Policy Dashboard project is the pursuit of automation and efficiency in data processing to save time and cost. The GEPD has almost fully automated the storage and processing of data by integrating several WBG products that enable this functionality. As the data collection platform, the GEPD uses Survey Solutions, which is a tablet-based, free, open source survey tool designed by the World Bank. The raw data is hosted on a

secure server through Amazon AWS, which is certified by numerous international bodies, including the EU Data Protection Directive. The data flow is as follows:

1. Data is collected by an enumerator
2. The enumerator or their supervisor uploads the collected data to the cloud secure server on AWS
3. The data is encrypted on the server
4. The GEPD team downloads the encrypted data using the Survey Solutions API
5. The data is stored in a secure folder within the World Bank network
6. The GEPD team cleans the data and creates encrypted unique IDs for students, teachers, principals, schools, and public officials using cryptographic hashing and running a fully automated R-code
7. Data quality checks are run using R-code to minimize missing values, identify enumerator errors, and ensure accurate coding
8. The data is then anonymized using R-code that removes all personal identifiable information (PII) information
9. The final indicators are calculated and aggregated to the national level (and relevant breakdowns) using R-code
10. This aggregated data is uploaded to the World Bank's Edstats open data platform
11. The GEPD website pulls the data from EdStats using the EdStats API system

### **Anonymization Process**

Immediately following download of data from Survey Solutions, a process is followed to save as an anonymized version of the data. The following variables are dropped:

1. Enumerator name
2. School name, address, official school codes (EMIS codes)
3. Principal name, phone numbers
4. Geo-code info
5. Unique responses, such as when respondent is asked to specify other as a choice

The following are produced:

1. Crypto-hashed school ID, province ID, district ID
2. Categorical respondent age
3. Categorical year began teaching
4. Categorical number of students in school/class
5. Categorical top-coded variables

## 6. Categorical educational degrees

### Data Quality Checks

While survey data is collected, the team runs daily quality checks to ensure the data is high-quality. These include both built-in checks available through the Survey Solutions Headquarters app and a custom-designed data quality check tool.

Survey Solutions has several built-in quality checks. These include a mapping tool based on the geo-coordinates entered by the enumerators during data collection and an interview management system that allows supervisors or the headquarters to see whether all questions that should have been completed by the enumerator are completed. Supervisors then have the option to send feedback to the enumerator and reject the interview until the data is successfully collected.

Additionally, the team has designed a set of custom tools for the GEPD Survey of Public Officials and School Survey. All code for the data quality checks can be found on the GEPD GitHub site (<https://github.com/worldbank/GEPD>). These tools allow the field team to calculate the final primary indicators in real time and check for any anomalies, including missing values and data entered at strange times (e.g. after regular business hours), as well as to look at data collected by specific enumerators.

## Section 3 – Indicator Construction

The GEPD produces a total of 39 indicators, built almost entirely from the data collected using the three survey instruments. Indicators are presented in a scale from 1 to 5 or as a percentage. On the GEPD, indicators will be illustrated using the traffic light system: green if the indicator is *On Target*, yellow if *Caution*, and red for *Needs Improvement*.

For our indicators ranging from 0-100, the display cutoffs are:

- On Target - Values at least 90%
- Caution - Values between 85 and 90%
- Needs improvement - Values under 85%

For our indicators ranging 1-5, the display cutoffs are:

- On Target - Values of at least 4
- Caution - Values between 3 and 4
- Needs improvement - Values 3 and under

The following section discusses the indicator-specific proficiency cutoffs and how the broader set of indicators are constructed.

## Proficiency Cutoffs

For indicators that involve assessment—the 1st-Grade Direct Assessment, 4th-Grade Direct Assessment, and Teacher Content Knowledge Assessment—minimum proficiency thresholds had to be created. Below there is a brief description of how these proficiency cutoffs were determined.

### *1st-Grade Assessment*

The threshold for proficiency on the 1st-Grade Direct Assessment is a score of 80%, meaning that that students must answer 80% of the questions correctly to be counted as having an appropriate level of preparedness for learning. Because the MELQO/GECDD initiative has not set any proficiency threshold and there is no Global Proficiency Framework in early learning to establish such a cutoff, the GEPD team had to estimate its own. The 80% cutoff was validated by calculating the percent correct in the top 25% highest-scoring schools in Peru on the 4th-Grade dashboard assessment and finding that students in these schools answered roughly this percentage of items correct. Because students in these top-performing schools typically achieve proficiency on the 4<sup>th</sup>-grade assessment, this provides some reassurance that 80% is a reasonable proficiency cutoff on the 1<sup>st</sup>-grade assessment. However, it is admittedly a provisional cutoff, and it will likely be revised as further data (especially longitudinal data on outcomes for children with different MELQO scores) becomes available.

Table 2. Percent correct for students in the highest-performing schools (top 25%)

<b>Indicator</b>	<b>Mean</b>	<b>Std Dev</b>	<b>Median</b>
1 <sup>st</sup> Grade Overall Score	84.00	13.90	87.72
1 <sup>st</sup> Grade Math Score	94.10	10.44	99.42
1 <sup>st</sup> Grade Literacy Score	84.55	13.62	88.53
1 <sup>st</sup> Grade Executive Functioning Score	74.96	16.25	77.19
1 <sup>st</sup> Grade Socio-Emotional Score	82.40	24.63	93.33

### *4th-Grade Assessment Cutoffs*

To determine whether or not a child was proficient based on the GEPD assessment, the team consulted with a set of experts in psychometrics and pedagogy in low- and middle-income countries. The experts were asked to complete a standard setting exercise in which each expert was asked to rate whether a “minimally proficient” student in the subject should be able to answer the item correctly, using the descriptions in the global proficiency framework being designed by UIS. The result of that exercise was that one would expect a minimally proficient child to answer 20/24 questions correctly on the literacy section and 14/17 points on the mathematics assessment. Students meeting these minimum scores are rated as proficient, others as not proficient.

### Teacher Assessment Cutoffs

The teacher assessment is based on the SDI assessment, which was validated against 13 Sub-Saharan primary curricula.<sup>12</sup> For the GEPD Teacher Assessment, the threshold for minimum proficiency was set at 80% on both math and language. This means that a teacher must correctly grade 80% of the questions in either math or language (depending on the assessment they complete) for them to be considered minimally proficient in the subject they teach. This was the benchmark used in the Service Delivery Indicators (SDI) initiative, and the Global Education Policy Dashboard initiative has kept this threshold.

### Question Scoring and Calculation of Indicators

As noted above, the GEPD includes a total of 39 indicators. These can be categorized into 4 general categories: Outcomes, Practices, Policies, and Politics. While the GEPD Reference Guide contains information regarding the measurement approach for each of them as well as the source for the information used, this section provides more information regarding how questions are scored, where appropriate, as well as how points are distributed across the different sub-aspects.

Table 4 provides information on how the five Outcome Indicators are calculated. As noted below, only one of them is computed using data collected through the dashboard instruments; the other four come from existing data sources.

Table 4: Scoring of GEPD Outcome Indicators

Indicator	Scoring Approach
Proficiency on GEPD Assessment	Each question on the 4th-Grade Student Assessment is scored as 1 point. The indicator reports the fraction of students scoring at least 20 out of 24 points on the 4th-grade language assessment and at least 14 out of 17 points on the math assessment.
Proficiency by Grade 2/3	Fraction of students minimally proficient according to UIS Database for SDG 4.1.1.a.

---

<sup>12</sup> The countries included for the review were: Botswana, Ethiopia, Gambia, Kenya, Madagascar, Mauritius, Namibia, Nigeria, Rwanda, Seychelles, South Africa, Tanzania, and Uganda. See David Johnson, Andrew Cunningham and Rachel Dowling (2012), "Teaching Standards and Curriculum Review", prepared as a background document for the SDI Survey. While some SDI countries were not part of the review, before doing the SDI survey the SDI team verified that the school curriculum was compatible with the SDI teacher and student test. For more information, see the SDI technical report for each specific country at [sdindicators.org](http://sdindicators.org).

Proficiency by End of Primary	Fraction of students minimally proficient according to UIS Database for SDG 4.1.1.b.
Participation	Adjusted primary net enrollment rate, as reported in the UIS Database
Learning Poverty	Learning poverty indicator, as reported in the Learning Poverty Database, capturing schooling and learning at the end of primary.

Table 5 provides information on the scoring of the 11 Practice Indicators. These are all based on data collected through the GEPD School Survey. More information on the measurement approach can be found in the GEPD Reference Guide.

Table 5: Scoring of GEPD Practice Indicators

Indicator	Scoring Approach
Teacher Presence 0% - 100%	Percent of sampled teachers who are coded as being in school at the time of the visit.
Teacher Content Knowledge 0% - 100%	Percent of teachers scoring at least 80% correct on the teacher assessment. In this assessment, each question is worth 1 point.
Teacher Pedagogical Practices 0% - 100%	Percent of teachers scoring an average of at least 3.5 (on a 1-5 scale) across the 9 components of the Teach tool. These include: Supportive Learning Environment, Positive Behavioral Expectations, Lesson Facilitation, Checks for Understanding, Feedback, Critical Thinking, Autonomy, Perseverance, Social & Collaborative Skills.
Readiness for Learning	Percent of sampled 1st-grade students scoring at least 80% on the GEPD Direct Assessment. In this assessment, total points (100) are allocated equally across the 4 domains—numeracy, literacy, socioemotional skills, and executive function. Within each domain, all questions are given equal weight.



Indicator	Scoring Approach
Student Attendance	Percent of registered students in observed 4th-grade class who are present during the school visit, measured out of the total students on the class list.
Basic Infrastructure	<p>A score of 0-5 based on the availability of 5 infrastructure aspects. Each aspect counts as 1, so 0 if there are none available and up to 5 depending on the availability of each. The aspects captured are the following:</p> <ul style="list-style-type: none"> <li>- Improved water source</li> <li>- Functional toilets</li> <li>- Access to electricity</li> <li>- Access to internet</li> <li>- Accessibility for children with disabilities</li> </ul>
Basic Inputs	<p>A score of 0-5 based on the availability of 5 inputs. Each input counts as 1, so 0 if there are none available and up to 5 depending on the availability of each. The inputs considered are the following:</p> <ul style="list-style-type: none"> <li>- Functional blackboard</li> <li>- Pens/pencils and exercise books</li> <li>- Textbooks</li> <li>- Classroom furniture</li> <li>- ICT</li> </ul>
Operational Functions	<p>A score of 1-5 based on two hypothetical scenarios. The two scenarios are given equal weight (2 points each). The points will be given according to the clarity with which the responsibility to solve the problem is allocated (1 point) and the perception that the problem will be addressed in a timely manner (1 point). Partial credit is possible for co-sharing of responsibility or the problem only being partially resolved.</p>

Indicator	Scoring Approach
Instructional Leadership	<p>A score of 1-5 based on the presence of 4 practices as reported by teachers. The four practices, which are given equal weight, are the following:</p> <ul style="list-style-type: none"> <li>- Had a classroom observation in past year</li> <li>- Had a discussion based on that observation that lasted longer than 30 min</li> <li>- Received actionable feedback from that observation</li> <li>- Teacher had a lesson plan and discussed it with another person</li> </ul>
School Knowledge	<p>A score of 1-5 based on 6 questions directed to the principal about their school. These answers are then compared to the data collected in other modules. The score is computed according to accuracy:</p> <ul style="list-style-type: none"> <li>- 5 points if all questions accurate</li> <li>- 4 points if 5/6 questions accurate</li> <li>- 3 points if 4/6 questions accurate</li> <li>- 2 points if 3/6 questions accurate</li> <li>- 1 point if fewer than three accurate</li> </ul>
Management Skills	<p>A score of 1-5 based on the quality of two skills: goal-setting and problem-solving. Both skills are allocated an equal number of possible points (2). The quality of each is assessed as follows:</p> <ul style="list-style-type: none"> <li>- Goal-setting receives 2 points if there are clear school goals, known to the school community and linked to student learning, that are being monitored.</li> <li>- For problem-solving, the 2 points are awarded based on the quality and thoroughness with which the principal addresses the root cause of a hypothetical problem. Principals are given a</li> </ul>

Indicator	Scoring Approach
	<p>scenario and asked how they would likely react to the situation. E.g. Imagine that you conduct a school-wide student assessment, and while the performance is good overall, you notice that there is one class in particular that is lagging behind the others. What would be the first thing you would do? What actions would you take? How would you monitor progress?</p>

Table 6 provides information on the scoring of the 18 Policy GEPD Indicators. For each policy indicator, there are two numbers reported—one based on *de facto* information, and one based on the *de jure* information collected through the Policy Survey. Table 6 focuses on the *de facto* version of the indicators, but the same specifications were applied to the *de jure* indicators. The information reported under *de facto* policy levers largely comes from the School Survey. One exception is the Learner policy indicators, which come primarily from existing data sources. Our preferred sources are the DHS or MICS surveys, but not all countries have such surveys. In that case, we look for other surveys conducted in the country that can help us to form the indicator. If no such information exists, then we cannot report on the indicator. Please refer to the GEPD Reference Guide for more information on these indicators as well as the data sources.

Table 6: Scoring of GEPD Policy Levers

Indicator	Scoring Approach
Teaching – Attraction	<p>A score of 1-5 based on 5 factors. Each factor receives an equal weight in terms of possible points (0.8). The factors are the following:</p> <ul style="list-style-type: none"> <li>- Job satisfaction</li> <li>- Community satisfaction</li> <li>- Perceived meritocracy</li> <li>- Financial incentives</li> <li>- Absence of salary delays</li> </ul>
Teaching – Selection & Deployment	<p>A score of 1-5 based on 2 factors. Each factor receives an equal weight in terms of points (2). The factors are the following:</p>

Indicator	Scoring Approach
	<ul style="list-style-type: none"> <li>- Meritocratic selection – with points assigned in accordance to the aspects considered in selecting candidates</li> <li>- Meritocratic transfers – with points assigned in accordance to the aspects considered in granting transfers</li> </ul>
Teaching – Support	<p>A score of 1-5 based on 4 factors. Each factor receives an equal weight in terms of points. The factors are the following (with points listed in parentheses):</p> <ul style="list-style-type: none"> <li>- Induction (1) <ul style="list-style-type: none"> <li>o Reported induction (0.5)</li> <li>o Reported useful (0.5)</li> </ul> </li> <li>- Practicum (1) <ul style="list-style-type: none"> <li>o Reported practicum (0.5)</li> <li>o Reported lasting longer than 3 months and having class component (0.5)</li> </ul> </li> <li>- In-Service training (1) <ul style="list-style-type: none"> <li>o Reported in-service training (0.5)</li> <li>o Reported longer than 2 days (0.25)</li> <li>o Reported at least 25% practical (0.125)</li> <li>o Reported at least 50% practical (0.25)</li> </ul> </li> <li>- Opportunities to collaborate with other teachers (1)</li> </ul>
Teaching – Evaluation	<p>A score of 1-5 based on 4 factors. Each factor receives an equal weight in terms of points. The factors are the following:</p> <ul style="list-style-type: none"> <li>- Reported evaluation in the past year (1)</li> </ul>

Indicator	Scoring Approach
	<ul style="list-style-type: none"> <li>- Reported multiple evaluation criteria (1)</li> <li>- Reported consequences for negative evaluation (1)</li> <li>- Reported consequences for positive evaluation (1)</li> </ul>
Teaching – Intrinsic Motivation	<p>A score of 1-5 based on 5 factors. Each factor receives an equal weight in terms of points. The factors are the following:</p> <ul style="list-style-type: none"> <li>- Beliefs about absenteeism (0.8)</li> <li>- Beliefs about students’ ability to learn (0.8)</li> <li>- Beliefs about growth mindset (0.8)</li> <li>- Main reason to become a teacher is intrinsic (0.8)</li> <li>- Probationary period exists (0.8)</li> </ul>
Teaching – Monitoring & Accountability	<p>A score of 1-5 based on 4 factors. Each factor receives an equal weight in terms of points. The factors are the following:</p> <ul style="list-style-type: none"> <li>- Attendance is monitored (1)</li> <li>- Attendance is rewarded (1)</li> <li>- Absenteeism is penalized (1)</li> <li>- Absenteeism is not caused by bureaucratic procedures (1)</li> </ul>
Inputs & Infrastructure – Standards	<p>A score of 1-5 based on 1 factor for all inputs and infrastructure. Principals are asked “Do you know if there are standards in place requiring all schools to have...?” This question is asked for all 5 inputs and all 5 infrastructure aspects. Each question receives an equal weight in terms of points.</p>

Indicator	Scoring Approach
Inputs & Infrastructure – Monitoring	<p>A score of 1-5 based on 3 factors. Each factor receives an equal weight in terms of points. The factors are the following:</p> <ul style="list-style-type: none"> <li>- Someone is monitoring</li> <li>- System for monitoring available (e.g. inventory)</li> <li>- Community involved in monitoring</li> </ul>
Learners – Nutrition Programs	<p>A score of 1-5 based on 4 factors. Each factor receives an equal weight in terms of points (1). The factors are the following:</p> <ul style="list-style-type: none"> <li>- Percentage of households with salt testing positive for any iodide among households</li> <li>- Percentage of children age 6–23 months who had at least the minimum dietary diversity and the minimum meal frequency during the previous day</li> <li>- Percentage of schools with a school feeding program</li> <li>- Percentage of children born in the five (three) years preceding the survey who were ever breastfed</li> </ul> <p>Each percentage is treated as the share of the possible points, and thus to create the total score, they are summed and added to 1.</p>
Learners – Health Programs	<p>A score of 1-5 based on 4 factors. Each factor receives an equal weight in terms of points (1). The factors are the following:</p> <ul style="list-style-type: none"> <li>- Percentage of children who at age 12-23 months had received all basic vaccinations recommended in the national immunization schedule</li> </ul>

Indicator	Scoring Approach
	<ul style="list-style-type: none"> <li>- Percentage of children under 5 covered by health insurance</li> <li>- Percentage of children age 6-59 months who received deworming medication.</li> <li>- Percentage of women age 15-49 years with a live birth in the last 2 years whose most recent live birth was delivered in a health facility</li> </ul> <p>Each percentage is treated as the share of the possible points, and thus to create the total score, their sum is added to 1.</p>
Learners – Early childhood Education	<p>A score of 1-5 based on 2 factors. Each factor receives an equal weight in terms of points (2). The factors are the following:</p> <ul style="list-style-type: none"> <li>- Percentage of children age 36-59 months who are attending ECE</li> <li>- Percentage of ECE Classrooms with Effective Practices</li> </ul> <p>Each percentage is treated as the share of the possible points, and thus to create the total score, the sum of the share times the possible points is added to 1.</p>
Learners – Caregiver Capacity – Skills	<p>A score of 1-5 based on 2 factors. Each factor receives an equal weight in terms of points (2). The factors are the following:</p> <ul style="list-style-type: none"> <li>- Percentage of children under age 5 who have three or more children’s books</li> <li>- Percentage of children age 24-59 months engaged in four or more activities to provide early stimulation and responsive care in the last 3 days with any adult in the household</li> </ul>

Indicator	Scoring Approach
	<p>Each percentage is treated as the share of the possible points, and thus to create the total score, the sum of the share times the possible points is added to 1.</p>
<p>Learners – Caregiver Capacity – Financial</p>	<p>A score of 1-5 based on 1 factor. The factor is the following:</p> <ul style="list-style-type: none"> <li>- Percentage of population participating in social protection and labor programs (includes direct and indirect beneficiaries)</li> </ul> <p>The indicator score is the multiplication of the share and the total possible points (4). That is added to 1 since the indicator is normalized from 1 to 5.</p>
<p>School Management – Clarity of Functions</p>	<p>A score of 1-5 based on 1 factor for 7 functions. To score this question, principals are asked “Do you know if the policies governing schools assign responsibility for the implementation of each of the following? Indicate for each, Yes/No as well as the level at which they are allocated: national, sub-national, local, or school.” The total score has been split among the 7 functions in the following way, so that Inputs &amp; Infrastructure elements, teacher elements, student elements, and school management elements have the same number of total points. The functions are the following:</p> <ul style="list-style-type: none"> <li>- Maintenance and expansion of school infrastructure (Inputs &amp; Infrastructure – 0.5 pts)</li> <li>- Procurement of materials (Inputs &amp; Infrastructure – 0.5 pts)</li> <li>- Teacher hiring and assignment (Teachers – 0.5 pts)</li> <li>- Teacher supervision, training, and coaching of teachers (Teachers – 0.5 pts)</li> <li>- Student learning assessments (Students – 1 pt)</li> <li>- Principal hiring and assignment (School Management – 0.5 pts)</li> <li>- Principal supervision and training (School Management – 0.5 pts)</li> </ul>



Indicator	Scoring Approach
	<p>Points are assigned for each function if the principal said that responsibility was allocated, regardless of the level.</p>
<p>School Management – Attraction</p>	<p>A score of 1-5 based on 2 factors. Each factor receives an equal weight in terms of points. The factors are the following:</p> <ul style="list-style-type: none"> <li>- Satisfaction in the community</li> <li>- Salary as a share of GDP per capita</li> </ul> <p>GDP per capita data is pulled from the World Development Indicators database of the World Bank. Principal salary points are put on a 1-5 scale as follows:</p> <ul style="list-style-type: none"> <li>- Salary less than 50% of GDP per capita (1)</li> <li>- Salary 50-75% of GDP per capita (2)</li> <li>- Salary 75-100% of GDP per capita (3)</li> <li>- Salary 100-150% of GDP per capita (4)</li> <li>- Salary more than 150% of GDP per capita (5)</li> </ul>
<p>School Management – Selection &amp; Deployment</p>	<p>A score of 1-5 based exclusively on how principals are selected. The scoring is as follows:</p> <ul style="list-style-type: none"> <li>- Most important factor is political affiliations or ethnic group (1)</li> <li>- Political affiliations or ethnic group is a consideration, but other factors considered as well (2)</li> <li>- Most important factor is years of experience, good relationship with owner/education department, and does not factor in quality teaching, demonstrated management qualities, or knowledge of local community (3)</li> </ul>

Indicator	Scoring Approach
	<ul style="list-style-type: none"> <li>- Quality teaching, demonstrated management qualities, or knowledge of local community is a consideration in hiring, but not the most important factor (4)</li> <li>- Quality teaching, demonstrated management qualities, or knowledge of local community is the most important factor in hiring (5)</li> </ul>
School Management – Support	<p>A score of 1-5 based on 4 factors. The factors and the way in which the points have been allocated are the following:</p> <ul style="list-style-type: none"> <li>- Principal has received training on how to manage the school (1)</li> <li>- Principal Training involved training for new principals, in-service training, and mentoring/ coaching (1)</li> <li>- Principal found such training useful and has applied the gained skills (1)</li> <li>- In-service training is offered at least once a year (1)</li> </ul>
School Management – Evaluation	<p>A score of 1-5 based on 4 factors. Each factor receives an equal weight in terms of points. The factors are the following:</p> <ul style="list-style-type: none"> <li>- Reported evaluation in the past year (1)</li> <li>- Reported multiple evaluation criteria (1)</li> <li>- Reported consequences for negative evaluation (1)</li> <li>- Reported consequences for positive evaluation (1)</li> </ul>

Table 7 provides information on the scoring for the last set of 5 indicators, the indicators for Politics & Bureaucratic Capacity. These indicators report information collected through the Survey of Public Officials, with the exception of Financing, which comes from existing data sources. Please refer to the GEPD Reference Guide for more information on these indicators as well as the data sources.

Table 7: Scoring of GEPD Politics & Bureaucratic Capacity Indicators

Indicator	Scoring Approach
Characteristics of Bureaucracy	<p>A score of 1-5 based on 4 factors. Each factor has been given an equal weight. Each factor is based on a set of 3-4 questions scored 1-5. For each factor, the average score across the questions is determined. To construct the total score, the average is taken of the 4 factor scores. The factors include:</p> <ul style="list-style-type: none"> <li>- Knowledge &amp; skills</li> <li>- Work environment</li> <li>- Merit</li> <li>- Motivation</li> </ul>
Impartial Decision-Making	<p>A score of 1-5 based on 4 factors. Each factor has been given an equal weight. Each factor is based on a set of 3 questions scored 1-5. For each factor, the average score across the 3 questions is determined. To construct the total score, the average is taken of the 4 factor scores. The factors include:</p> <ul style="list-style-type: none"> <li>- Politized personnel management</li> <li>- Politized policy-making</li> <li>- Politized policy implementation</li> <li>- Employee unions as facilitators</li> </ul>
Mandates & Accountability	<p>A score of 1-5 based on 3 factors. Each factor has been given an equal weight. Each factor is based on a set of 3 questions scored 1-5. For each factor, the average score across the 3 questions is determined. To construct the</p>

Indicator	Scoring Approach
	<p>total score, the average is taken of the 3 factor scores. The factors include:</p> <ul style="list-style-type: none"> <li>- Coherence</li> <li>- Transparency</li> <li>- Accountability of public officials</li> </ul>
National Learning Goals	<p>A score of 1-5 based on 4 factors. Each factor has been given an equal weight. Each factor is based on a set of 3 questions scored 1-5. For each factor, the average score across the 3 questions is determined. To construct the total score, the average is taken of the 4 factor scores. The factors include:</p> <ul style="list-style-type: none"> <li>- Targeting</li> <li>- Monitoring</li> <li>- Incentives</li> <li>- Community Engagement</li> </ul>
Financing	<p>A score ranging from 1 to 5 that considers the quality of financing using two lenses – adequacy and efficiency. This score is calculated using 3 sub-indicators: 1) per-child spending (adequacy), 2) public management financing performance (efficiency), and 3) outcomes per spending (efficiency). If reliable data are available on incidence of education spending by socioeconomic status and other sources of disadvantage, a fourth sub-indicator on equity will be included. The total score is a weighted sum of the adequacy and efficiency components.</p> <ul style="list-style-type: none"> <li>- 50% weight for adequacy</li> <li>- 50% weight for efficiency</li> </ul>

Indicator	Scoring Approach
	The efficiency score is an average of the public management financing performance (efficiency) and outcomes per amount spent (efficiency).

Underlying the 39 primary indicators are various sub-indicators. For instance, within the Basic Inputs indicator, sub-indicators on the availability of a functional blackboard and other classroom materials also exist. For a detailed description of the scoring for each of these sub-indicators, please see the project Github page at:

[https://github.com/worldbank/GEPD/blob/master/Indicators/indicators\\_details.csv](https://github.com/worldbank/GEPD/blob/master/Indicators/indicators_details.csv)

## **Annex 1: Detailed Indicator Calculation**

Details on how each of the indicators is calculated can be found in the GEPD team's Github repository. Code is available in R and Stata. The code is omitted from this technical note, both for brevity and to allow for updating, but all code can be accessed here:

<https://github.com/worldbank/GEPD>