# Girls' Education South Sudan
## Learning Assessment: Midline

Charlie
Goldsmith
**Associates**

## Acknowledgements

The following individuals contributed to the field research and analysis undertaken for the GESS Learning Assessment Midterm Report:

Natalie Forcier, Partner
Dr. Rens Twijnstra, Lead Methodologist
Dr. Clare Hollowell, Technical Advisor
Tim Berke, Research Manager
Lokiri Moses Lowuro, Research Officer
Jeff Longoria, Senior Researcher & Quality Control
Brenton Peterson, Chief Research Quality Officer
Hayley Umayam, Engagement Manager & Analyst

# Table of Contents

## List of Tables

## Acronyms

| | |
|---|---|
| **CPA** | Comprehensive Peace Agreement |
| **DiD** | Difference-in-differences |
| **EfC** | Education for Change |
| **GESS** | Girls' Education South Sudan |
| **GUN** | Greater Upper Nile |
| **SACMEQ** | Southern and Eastern Africa Consortium for Monitoring Educational Quality |
| **SSSAMS** | South Sudan Schools Attendance Monitoring System |

# Executive Summary

The 2016 Girls Education in South Sudan (GESS) midterm learning assessments were carried out in all 10 (former) states of South Sudan, unlike the Baseline of 2014 which was carried out in only the 7 states that were accessible at the time of study (Central Equatoria, Eastern Equatoria, Lakes, Northern Bahr El Ghazal, Warrap, Western Bahr El Ghazal, and Western Equatoria).[1] Literacy and numeracy tests were administered in September and October 2016 at Primary 5, Primary 8 and Secondary 2 grades in a sample of 59 schools, including 38 primary schools and 21 secondary schools. This school sample included assessments from 10,042 pupils of which 3,993 were girls, 5,297 were boys, and 752 had no sex listed.

In 2014, results indicated a significant gender gap at all grade levels in both literacy and numeracy. While girls' scoring on the learning assessments remains lower than that of their male counterparts at many grades, girls' scores increased between 2014 and 2016, which forms the basis for an overall positive programme effect. As in 2014, pupils perform 'better' (higher mean scores) on the literacy compared to the numeracy assessments. While there are challenges and limitations in evaluating the impact of a national programme, the results of the baseline, midterm and, eventually, end term learning assessments contribute to the Knowledge, Evaluation and Research strand of GESS.

It must be noted that the results presented in this report may differ slightly from those in preliminary results reported for 2016 data and in some cases, from those in the 2014 baseline report. This will be accounted for in the Methodology section.

**Results and Analysis**

- Overall, average scores decreased from 45.74% in 2014 to 44.70% in 2016. This result is likely influenced by selection effects of the baseline, in which three states were excluded from the sample due to insecurity. As the insecurity in the Greater Upper Nile (GUN) states of South Sudan affected the education sector, it is reasonable to assume that average scores in 2014 would also have been affected had the GUN states been included in the overall sample.
- A difference-in-differences (DiD) estimate shows that there is a sizable positive effect in the aggregate (coef: 0.029, p-value: <0.000). If each grade level and type is considered separately, 4 out of 6 tests show a positive and significant (at least marginally) effect. In the aggregate, **the GESS "treatment" has had a positive effect in girls test scores relative to boys.**
- Results appear to be largely driven by improvements - relative to boys- in girls' numeracy scores. **Girls' average scores increased** from an average of 44.2% to 45.7% **while boys' average scores decreased** from 46.7% in 2014 to 45.3% in 2016.
- **Numeracy**: Overall scores increased for numeracy tests among both boys and girls. In particular, girls' exhibited gains at the Primary 5 and Primary 8 numeracy level (with average scores of 48.6% and 43.4%, respectively).
- **Literacy**: Overall scores decreased for literacy tests from 55.51% in 2014 to 49.06% in 2016. Girls' scores remained behind boys' and, in the case of Primary 8, the gap widened.
- In terms of a DiD estimate on average scores, there are **strong effects for numeracy scores (at all grade levels), while weak or no effects were found among literacy scores**. Primary 5 is a (positive) exception with regard to literacy, but it fits the overall findings that the effects for literacy are weaker in every grade level when compared to numeracy. Average scores reflected

---

[1] The sample design, fieldwork, and analysis of the 2014 baseline were conducted prior to the Executive Order in October 2015 establishing 28 states in South Sudan. State-level findings in this 2016 report srefer to the previous administrative delineation of ten states for consistency,.

these DiD results. This is consistent with the caution expressed at baseline re literacy findings relative to numeracy.

- Out of all the grade levels, Primary 5 girls made the most significant gains in terms of test score increases. By 2016, Primary 5 girls had closed the gap with boys in terms of numeracy test scores.
- The largest overall effects are at the Primary 5 level (coef: 0.040, p-value: <0.000). This is consistent with the pattern of enrolment growth over the same period, which has been stronger at Primary 5 through Primary 8 levels than in the first four years of Secondary school.

**Test Administration**

The GESS midterm learning assessments were administered in September and October 2016 using procedures that were nearly identical to that of the 2014 baseline. The 2014 assessments had been improved for validity, and no further changes were made to the question content in the 2016 assessments.

The test administration process improved in some areas but remains a challenge. For instance, unlike 2014, all missing and unused papers were returned. However, confirming full compliance with test administration procedures continued to be a challenge, as many invigilators failed to submit field reports.

# 1. Introduction

The GESS programme is an UK Aid-funded initiative aimed at increasing the participation of girls in primary and secondary education in South Sudan. This programme is monitored and evaluated on the basis of several tools, including a suite of learning assessments that are focused on literacy and numeracy.

The GESS learning assessments are administered in the South Sudan states participating in the programme. The learning assessments are confidential materials given that they may be used as 'live' tests in future years. Therefore, test security protocols should apply. Mindful of this, specific questions and text content (on literacy assessments) are not included in this report. These can be viewed separately on application to Charlie Goldsmith Associates.

## 1.1.    Background

The GESS programme was designed in 2012, shortly after South Sudan gained independence and officially launched in April 2013. Activities include local-language radio programmes and community mobilisation, school grants, cash transfers to girls, training for education managers and teachers, research studies and the establishment of the South Sudan School Attendance Monitoring System (SSAMS).[2]

In July 2014 *Forcier Consulting* were appointed to review the learning assessment materials, train invigilators and markers, manage the marking and coding of pupils' responses, analyse the data and provide a report on outcomes and recommendations. The 2014 assessment served as a baseline upon which the 2016 midterm results are compared.

The GESS learning assessments were originally developed and piloted in South Sudan by *Education for Change* (EfC) in 2013. The findings of these trials were incorporated into the 2014 learning assessments. The contribution of all earlier contributors is hereby acknowledged and appreciated. In particular, this midterm assessment draws heavily upon the initial insights as described in the 2014 baseline report prepared by Jacky Burnett.

Cash Transfers were not paid before the 2014 learning assessment baseline, but the 2014 and 2015 cash transfer instalments would have been received by girls before the 2016 midterm. The first instalment (tranche) of Capitations Grants were received by eligible schools in time for the 2014 baseline, but, as with the baseline sample frame, was limited to accessible areas. It must be noted that in some areas of South Sudan remain inaccessible and thus schools in these areas have not yet received their Capitation Grants or did not receive it in time by the time of the midterm.

---

[2] van der Meulen, Emma and Akuja de Garang. "Education development in a fragile environment: lessons from Girls' Education South Sudan. January 2017. http://odihpn.org/magazine/education-development-in-a-fragile-environment-lessons-from-girls-education-south-sudan/

# 2. Purpose of Assessment & Methodology

As described in the 2014 baseline report, the principal purpose of this learning assessment is to enable GESS to evaluate the success of their programme. The 2014 iterations of the assessments acted as the baseline by which to evaluate the effect of the interventions made. This 2016 midterm assessment enables the comparison of pupils' performance over time.

In the case of both the baseline and midterm GESS learning assessments, the principal purpose is for programme evaluation; it is linked to accountability with regards to the success of the overall GESS programme encouraging girls into school and to demonstrate an increase in learning outcomes. The need to include a sample of urban and rural schools may skew the sampling process. At a national level, for the purpose of programme evaluation this should not undermine the validity of the assessment. However, it must be noted that the sample size is not sufficient at the State level to use the statistical data for system monitoring on a regional basis; nor is it robust enough to report at a school level. State level data is not therefore discussed in the body of this report; however, selected state-specific and school-specific groupings have been included as an additional check on the aggregate results but are not intended to be considered as discrete outcomes.

The comparability of pupils' performance over time, linked to programme evaluation, is a subsidiary purpose of this midterm. The 2014 learning assessment served as a baseline by which to compare performance in future years of the GESS programme, including this 2016 midterm. This comparison works when the performance of the cohort overall is considered.[3] However, the assessment tool is not designed to report on the performance of individual pupils or schools. It is vital that rigorous test security procedures and practices are in place to ensure the integrity and validity of the learning assessments for use in future years. Remarks on test validity can be found in the 2014 baseline report.

It should be noted that curriculum in South Sudan has evolved over time, including over the period of the GESS programme. Previously, some areas of the country continued to teach Sudanese or Ugandan curricula, but this variance was eliminated following a 2015 MoGEI decision to phase out all foreign curricula.

## 2.1.    Methodology

The monitoring and evaluation of the GESS programme includes a summative assessment of learning, which requires that pupils' learning be measured at a set point in the school year. This consistent timing is particularly important for year-on-year comparisons; the timeline was generally adhered to for the administration of both the baseline and midterm learning assessments. For the GESS programme, the learning assessment 'standard' has been set at the Literacy and Numeracy curriculum at the Primary 5, Primary 8 and Secondary 2 grade levels.

There are 6 learning assessments in total that were designed to identify pupils' understanding of literacy and numeracy at the Primary 5, Primary 8, and Secondary 2 grade levels. The layout of the exam was changed into 2 configurations for each assessment, but the question content remained the same as the

---

[3] This comparison could evaluate improvements of specific grades (such as Primary 5 performance year by year) as well as the scores of a cohort's progression through grades over time.

2014 baseline assessments. A description of the review of the 2013 pilot materials, test development, and formatting can be found in the 2014 baseline report.

### 2.1.1. Test Administration and Invigilator Training

Training for 12 Lead Invigilators (11 Lead Invigilators and 1 CGA staff member) occurred on August 23 & 24 2016. Invigilators selected teachers from sampled schools and trained them to supervise the test administration. The 2014 test administration revealed failures on some of the administration procedures, such as the need to return all unused papers. These points were emphasized during trainings. Some Invigilators had also participated in the 2014 baseline test administration. With the exception of Warrap state, which had 2 Invigilators, there was 1 Invigilator per state.

*Table 1: Test Administration Dates*

| State | Field dates (test administration) |
|---|---|
| Central Equatoria | 27th September |
| Western Equatoria | 25th Sept- Tambura<br>5th Oct- Mvolo<br>12th Sept-Yambio<br>15th October-Maridi |
| Eastern Equatoria | 23rd September |
| Northern Bahr el Ghazal | 14th September |
| Western Bahr el Ghazal | 6th September |
| Lakes | 26th & 27th September |
| Upper Nile | 26th September |
| Unity | 26th September |
| Warrap | 15th September |
| Jonglei | 28th September |

As the learning assessment model is not designed to report on the performance of individual pupils, the papers are not 'marked', but rather a team of Markers are trained to 'code' pupils responses based on the answer options, unclear intent, or omissions. Forcier staff conducted specialized training in both data entry and marking/coding.

Data entry and quality control training occurred on 12 & 13 September for 2 Forcier staff. Training for the test markers/coders occurred on 29 & 30 September 2016 for 18 Forcier staff, involving one day of 'standardisation' to ensure consistency with marking decisions and a day of 'commentaries' to help markers internalise the decisions. The Lead Proctors were responsible for collecting the exams and bringing them to Juba. The first tests arrived from the field on 21 September, and the last tests arrived on 31 October, with one exception (see below). Tests were marked and coded as they arrived, including a quality assurance exercise on the first day of 'live' marking; overall coding occurred 3-24 October 2016.[4] Data entry was done alongside coding and occurred 14 October- 16 November. The coding process remained similar to the procedures in 2014 except that additional quality control steps, such as data entry tracking sheets, were integrated throughout the coding process.

## 2.1.2. Data Collection and Data Analysis

As mentioned, the findings presented in this report may vary slightly from those reported in a series of preliminary tables and, in some cases, the 2014 baseline report. There are three reasons for these discrepancies. First, the preliminary tables showing changes between 2016 and 2014 erroneously drew from the raw 2014 data, which included some duplicates. Secondly, it appears that the Marking Key used in 2014 for the Primary 5 literacy assessment had one incorrect mark. Both of these have been corrected in all references to the 2014 data in the present report.

The last source of slight discrepancies is in approach to scoring of omitted and unclear answers. There are 3 possible approaches. The first includes every question in the test and rewards only the correct

---

[4] It should be noted that a few (<30) additional tests were retrieved from the field in early December 2016. These tests were marked in the same month that they were received.

answers. The second excludes questions in which the students' answer is unclear, thus changing the denominator when calculating a student's total score. The third excludes questions in which the students' answer is unclear as well as questions that the student failed to answer (omitted). The preliminary comparative tables assumed the first scoring approach for the 2014 data and reported scores as such. However, upon further investigation, it was shown that the 2014 analysis used the second and third scoring approaches for different tests. **For consistency, all scores of the 2014 and 2016 literacy assessments presented in this midterm report have been scored according to the second approach that does not penalize students' unclear answers but does penalize omitted questions.**

*Table 2: Maximum Marks Per Learning Assessment*

| Grade | Learning Assessment (maximum # marks) | |
|---|---|---|
| P5 | Literacy (16 marks) | Numeracy (30 marks) |
| P8 | Literacy (19 marks) | Numeracy (32 marks) |
| S2 | Literacy (30 mars) | Numeracy (32 marks) |

**Invigilator quality**
Poor compliance with test administration procedures or other procedures may have affected data quality even though they were emphasized during training. The majority of invigilators failed to use attendance sheets during testing. A lack of attendance sheets made it difficult to cross-check the number of tests received against the number of students who sat for the test. This also made it difficult to reconcile some cases of possible cheating. For instance, it was discovered during the data cleaning process that there were several duplicate observations in which the same full name was used on two or more exams, but each instance resulted in a different score. More than one student using the same name on their paper could have caused this, but as it could not be verified, this resulted in several removed observations. This points to a lack of attention on the part of the test administrators.

Despite emphasis during training, few Invigilator reports were received. Full compliance to the administration procedures was therefore difficult to confirm. In future assessments it would be advised to make Invigilator payment contingent upon submission of a field report and/or documents that confirm compliance such as attendance sheets.

## 2.2.     The 2016 School Sample

The sample was drawn among schools as the primary sampling unit, and then within every school relevant classes were tested. The sample design stipulated that 5 primary and 2 secondary schools be selected per state for a total of 70 schools. However, some schools were inaccessible at the time of fielding and were not replaced or substituted. In one instance, all observations from a primary school were excluded from the sample after test administration due to evidence of cheating. After the final data cleaning and quality control checks, a total of 59 schools have been included in the 2016 sample. The final sample size is ample for aggregate (national) analysis, but is not intended to be representative at the state level.

*Table 3: 2016 Sample- Schools Per State*

| State | Primary | | Secondary | |
|---|---|---|---|---|
| | Planned | Actual | Planned | Actual |
| Central Equatoria | 5 | 5 | 2 | 2 |
| Eastern Equatoria | 5 | 4 | 2 | 2 |
| Jonglei | 5 | 2 | 2 | 2 |
| Lakes | 5 | 4 | 2 | 2 |
| Northern Bahr El Ghazal | 5 | 4 | 2 | 2 |
| Unity | 5 | 3 | 2 | 2 |
| Upper Nile | 5 | 4 | 2 | 2 |
| Warrap | 5 | 4 | 2 | 2 |
| Western Bahr El Ghazal | 5 | 4 | 2 | 2 |
| Western Equatoria | 5 | 4 | 2 | 3 |
| Total | 50 | 38 | 20 | 21 |

The lower numbers of pupils at most grade levels in the 2016 sample must be considered when interpreting the data. In accordance with Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) recommendations, it would be hoped that roughly 2000 pupils would participate in each of the tests to allow for robust statistical analysis and comparison.[5] As the Learning Assessment sample was drawn among schools and not pupils, the total numbers of pupils tested was dependent on enrolment. No further sampling was conducted by grade level. In practice, this means that the numbers of pupils attained in the sample reflects the reality of primary and secondary enrolment in South Sudan. The simple lack of students may account for some of the larger number of students at the lower grades, which is in turn often a reflection of the fact that school enrolment generally decreases at each level. Students leave education for a variety of reasons, including pursuing secondary education abroad, typically in Uganda or Kenya. In other circumstances, education is not financially viable for families or social dynamics discourage girls from pursuing further education.[6]

Four primary schools included in the sample only have Primary 5 students, while one primary school has only Primary 8 students.

---

[5] SACMEQ is a collaborative network of 15 Ministries of Education that conducts educational policy research.
[6] This matter has been expanded upon in other GESS literature and has motivated the cash transfer and capitation grants.

*Table 4: 2016 Sample- Gender Ratios by Learning Assessment*

| Grade | Overall Total | Girls | | Boys | | Sex not indicated[7] | |
|---|---|---|---|---|---|---|---|
| | | Total | % | Total | % | Total | % |
| **Literacy** | | | | | | | |
| P5 Literacy | 2349 | 1115 | 47.47% | 1088 | 46.32% | 146 | 6.22% |
| P8 Literacy | 1473 | 613 | 41.62% | 850 | 57.71% | 10 | 0.68% |
| S2 Literacy | 1253 | 433 | 34.56% | 808 | 64.49% | 12 | 0.96% |
| Total | 5075 | 2161 | 42.58% | 2746 | 54.11% | 168 | 3.31% |
| **Numeracy** | | | | | | | |
| P5 Numeracy | 2224 | 863 | 38.80% | 1039 | 46.72% | 322 | 14.48% |
| P8 Numeracy | 1450 | 593 | 40.90% | 753 | 51.93% | 104 | 7.17% |
| S2 Numeracy | 1293 | 376 | 29.08% | 759 | 58.70% | 158 | 12.22% |
| Total | 4967 | 1832 | 36.88% | 2551 | 51.36% | 584 | 11.76% |
| | | | | | | | |
| Sample total | 10042 | 3993 | 39.76% | 5297 | 52.75% | 752 | 7.49% |

As with grade level, no further sampling was done to ensure equal gender representation. Therefore, the ratio of boys to girls presented in this report simply reflects who was in the classes tested, which tends to match the national enrolment pattern. Overall, girls make up roughly 40% of the 2016 sample, a ratio that has improved from the 2014 sample in which girls accounted for 32% of the sample.

The ability to make strong conclusions about the gender ratios present in these levels is limited by a large number of instances where no sex was listed. There are several grade levels with large omission rates, as high as 14.48% of the sample in the case of P5 Numeracy. These cases account for 7.5% of the total sample, a rate triple that of the 2014 sample in which 2.3% did not have sex listed.

Some of these omissions were caused by a failure to include sex as an item on the numeracy scripts, which is echoed by the higher instances of tests where sex was not listed among the numeracy tests. However, given the high instances with no sex listed among Primary 5 literacy, this also points to an error on the part of the test proctors, who were meant to double check that the basic information had been filled out properly. In addition, omission rates for numeracy tended to be higher for numeracy tests in 2014 as well, which might suggest that later test administration is tied to more lax proctors (i.e. proctors are more motivated to ensure all relevant elements of the form are completed in the morning than they are in the afternoon).

---

[7] The actual number of tests that were submitted without a sex listed is slightly higher than the rates here because Forcier staff was able to assign sex based on names or checking for whether the student listed their sex on the other test type for some instances. However, Forcier staff was unable to assign a sex for the majority of tests.

# 3. Findings

The differences in sample frame must be considered when investigating results at the aggregate (national) level. The sample in 2014 included 7 out of 10 possible states, while the 2016 sample includes schools from all 10 states. While the larger sampling frame still does not allow for comparison at the state level, it is more robust than the 2014 sample at a national level. However, it does pose the possibility that 2014 scores would have been different had the sample included the inaccessible states. In particular, GUN states, that are conflict-affected and likely to have weaker education systems, were excluded in the 2014 sample. It is possible that the 2014 results could have been lower in the aggregate had scores from these areas been included. State level scores have been included in the Annex as a matter of interest.

## 3.1.     Learning Assessment Performance

At the aggregate level, average test scores decreased from 45.74% in 2014 to 44.70% in 2016. However, when disaggregated by assessment type, average numeracy scores increased from 35.78% in 2016 to 40.23% in 2016.

*Table 5: Average Scores by Sex*

| Assessment | Year | Overall (%) | Girls (%) | Boys (%) |
|---|---|---|---|---|
| Literacy | | | | |
| P5 Literacy | 2014 | 54.71 | 51.56 | 57.22 |
| P5 Literacy | 2016 | 48.75 | 47.62 | 50.33 |
| P8 Literacy | 2014 | 61.96 | 61.48 | 62.30 |
| P8 Literacy | 2016 | 63.53 | 62.36 | 64.40 |
| S2 Literacy | 2014 | 52.80 | 50.50 | 53.54 |
| S2 Literacy | 2016 | 32.63 | 30.51 | 33.72 |
| Numeracy | | | | |
| P5 Numeracy | 2014 | 40.50 | 37.47 | 43.53 |
| P5 Numeracy | 2016 | 47.05 | 48.58 | 49.78 |
| P8 Numeracy | 2014 | 39.10 | 37.15 | 40.27 |
| P8 Numeracy | 2016 | 41.79 | 43.43 | 41.65 |
| S2 Numeracy | 2014 | 30.16 | 27.74 | 30.90 |
| S2 Numeracy | 2016 | 26.91 | 27.45 | 26.91 |

When girls' and boys' scoring are considered in the aggregate, it is shown that girls' scores increased since 2014 while boys' scores decreased. Furthermore, while there was a statistically significant gender gap in terms of average scores in 2014, the gap has closed in the 2016 results.

*Table 6: Aggregate Average Scores by Sex*

| Year | Girls av. score % | Boys av. score % | P-value (Sig between gender) |
|---|---|---|---|
| 2014 | 44.20 | 46.67 | 0.000 |
| 2016 | 45.71 | 45.33 | 0.404 |
| P-value (sig between years) | 0.007 | 0.001 | |

It appears that because boys were still in a majority in the 2016 sample, the overall decrease in test scores is due to the decrease in scores among boys.

### 3.1.1. Difference-in-differences: National Picture

A difference-in-differences (DiD) statistical test can be applied to determine causal effects of the GESS programme. In this model, boys are considered the 'control group' and girls are considered the 'test group.' The 'treatment' is the difference in average test scores between girls and boys, as cash transfers are allotted to girls only.[8] This model defines 2014 as a baseline, with the assumption that GESS programming had not yet occurred, which puts the difference in years as the time ('post'). The difference-in-differences treatment (derived via an OLS regression) effect is equal to the difference between girls and boys multiplied by the difference between average scores (*DiD=post\*treatment*). Pre-existing differences between treatment and control groups are acceptable, but it should also be noted that the regression used to determine the DiD coefficient is subject to omitted variable bias. For example, consider that the recent uptick in insecurity following violence in the capital in the summer of 2016 may have resulted in a decrease in school attendance. However, in this model, there have been no additional variables added to the regression to account for external trends.[9] In other words, this model assumes that no other external factors influenced the results.

In regards to aggregate results, there is a positive relationship (0.029) that is statistically significant. In other words, when considering the level of girls' and boys' academic performance, the GESS programme had a positive effect on girls' performance relative to that of boys.

In order to substantiate this finding, the same regression has been run on two increasingly narrow subsamples. By narrowing the model to only states that were included in both the 2014 and 2016 samples and then only schools that were included in both the 2014 and 2016 samples, we are able to compare whether the aggregate effects hold true under more refinement.

As mentioned, state and school level data is not discussed in the body of this report as the sample design was not intended to be representative at the state or school level. The selected state-specific and school-specific groupings have been included as an additional check on the aggregate results but should not be considered as discrete outcomes.

---

[8] Other GESS activities such as capitation grants are not gender-specific, but it is still necessary to consider girls as the "treatment" in this case since the overall programme aims to improve girls' performance.
[9] Future studies and analyses could include conflict variables using data such as that maintained by the Armed Conflict Location and Data Project (ACLED).

*Table 7: Aggregate DiD Effects*

|  | DiD coef. | 95% Conf. Int. | St. error | t | p-value |
|---|---|---|---|---|---|
| All observations in 2014 & 2016 | 0.029 | (0.015, 0.042) | 0.007 | 4.14 | 0.000 |
| States that were sampled in both 2014 & 2016 | 0.037 | (0.022, 0.051) | 0.007 | 5.07 | 0.000 |
| Schools that were sampled in 2014 & 2016 | 0.047 | (0.311 , 0.063) | 0.008 | 5.76 | 0.000 |

As shown in Table 7, the positive and significant effect of the GESS programme holds true among the subsample of states (0.037) and schools (0.047) that were included in both the 2014 and 2016 samples. This reinforces the finding that the GESS programme has had an overall positive effect on girls' test scores. Average test scores by gender for these subpopulations are included in the Annex as a matter of interest.

**Considerations**

When considering the findings above, it must be noted that it is not always clear whether the results are a function of girls' improvement (mean test score) or a drop in boys' performance. The best assessment is that in the absence of the GESS program, girls' scores would have dropped along with boys' scores relative to the baseline. The fact that they did not is the effect we are assuming to be attributable to the program. An ideal design would allow for comparisons both within schools and across schools to eliminate with this limitation.

When considering the findings below, it must be noted that the comparison is within schools. If the program redistributed attention or resources from boys to girls, it could have improved girls' scores at the expense of boys' performance. We cannot say whether this is true or whether boys' scores fell because of environmental or non-programme factors. Another design for a more precise evaluation would have included control schools in order to compare performance in program schools to performance in non-programme schools for both boys and girls. However, schools that did not receive cash transfers or capitation grants, or received them late, are schools that have been deliberately excluded from the national CGA programming due to other influential factors, which would negate their utility as a control group.

## 3.2.    Numeracy

Overall, scores in the numeracy assessments are positive. Across all grades tested, average scores in 2014 were 35.78%; and increased to 40.23% in 2016. In addition, the DiD coefficient (0.042) suggests a positive and statistically significant effect of the GESS programme in numeracy test scores across all grades.

The numeracy scoring gains have been the most drastic among girls in Primacy 5 and Secondary 2, where girls closed the scoring gaps that existed in 2014.

**Primary 5 Numeracy**

Out of all the grade levels, Primary 5 girls made the most significant gains in terms of test score increases. By 2016, Primary 5 girls had closed the gap with boys in terms of numeracy test scores.

Average 2014 score: 40.50% (including no sex)
Average 2016 score: 47.05%

*Table 8: Average Scores in Primary 5 Numeracy*

| Year (P5 Numeracy) | Girls av. score % | Boys av. score % | Sig between males and females |
|---|---|---|---|
| 2014 | 37.47 | 43.53 | 0.000 |
| 2016 | 48.58 | 49.78 | 0.181 |
| Sig between 2014 & 2016 | 0.000 | 0.000 | |

*Table 9: DiD Effect in Primary 5 Numeracy Scores*

| DiD coef. | 95% Conf. Int. | Standard error | t | p-value |
|---|---|---|---|---|
| 0.049 | (0.021, 0.077) | 0.014 | 3.46 | 0.001 |

The DiD coefficient suggests that even though boys' scores also increased, girls' exhibited a larger increase. The GESS programme is determined to have a positive and significant effect. This gain in Primary 5 numeracy is consistent with the pattern of enrolment growth as documented by the SSAMS which shows stronger enrolment at the Primary levels, particular at Primary 5, than in the Secondary tiers.

**Primary 8 Numeracy**

Girls' scores increased greatly since 2014, to a level where they have almost surpassed boys' scores, in terms of a statistically significant gender gap. It follows then that the DiD coefficient shows a positive and significant effect of the GESS programme.

Average 2014 score: 39.10% (including no sex)
Average 2016 score: 41.79%

*Table 10: Average Scores in Primary 8 Numeracy*

| Year (P8 Numeracy) | Girls av. score % | Boys av. score % | Sig between males and females |
|---|---|---|---|
| 2014 | 37.15 | 40.27 | 0.011 |
| 2016 | 43.43 | 41.65 | 0.071 |
| Sig between 2014 & 2016 | 0.000 | 0.166 | |

Unlike boys' scores in Primary 5, the gains in average tests scores among boys is not statistically significant.

*Table 11: DiD Effect in Primary 8 Numeracy Scores*

| Diff-in-diff coeff. (score) | 95% Conf. Interval | Standard error | t | p-value |
|---|---|---|---|---|
| 0.049 | (0.017, 0.081) | 0.016 | 2.98 | 0.003 |

**Secondary 2 Numeracy**
While the DiD coefficient suggests a positive effect, this result is likely to be caused from the fact that boys' scores decreased significantly, rather than the intervention itself. Indeed, there was no change in girls' average scores between 2014 and 2016. Rather than an increase in girls' scores, it was the decrease in boys' scores that closed the gap.

Average 2014 score: 30.16% (including no sex)
Average 2016 score: 26.91%

*Table 12: Average Scores in Secondary 2 Numeracy*

| Year (S2 Numeracy) | Girls av. score % | Boys av. score % | Sig between males and females |
|---|---|---|---|
| 2014 | 27.74 | 30.90 | 0.000 |
| 2016 | 27.45 | 26.91 | 0.499 |
| Sig between 2014 & 2016 | 0.759 | 0.000 | |

*Table 13: DiD Effect in Secondary 2 Numeracy Scores*

| DiD coef. | 95% Conf. Int. | Standard error | t | p-value |
|---|---|---|---|---|
| 0.037 | (0.015, 0.059) | 0.011 | 3.30 | 0.001 |

## 3.3. Literacy

While Primary 5 literacy scores showed a positive and slightly significant result, other literacy results showed less clear evidence of effect of the programme. The overall trend shows that the effects for literacy are weaker in every grade level when compared to numeracy.

Across all grade groups, average scores of literacy assessments were 55.51% in 2014 and dropped to 49.06% in 2016. Overall, the DiD coefficient for literacy (across all grades) is 0.015, and is not statistically significant (0.105).

**Primary 5 Literacy**
There was a significant drop in literacy test scores at the Primary 5 level, but the DiD shows a positive and slightly significant effect.

Average 2014 score: 54.71 (including no sex)
Average 2016 score: 48.75

*Table 14: Average Scores in Primary 5 Literacy*

| Year (P5 Literacy) | Girls av. score % | Boys av. score % | Sig between males and females |
|---|---|---|---|
| 2014 | 51.56 | 57.22 | 0.000 |
| 2016 | 47.62 | 50.33 | 0.001 |
| Sig between 2014 & 2016 | 0.001 | 0.000 | |

*Table 15: DiD Effect in Primary 5 Literacy Scores*

| DiD coeff. | 95% Conf. Int. | Standard error | t | p-value |
|---|---|---|---|---|
| 0.029 (2.9%) | (-0.0003, 0.059 ) | 0.015 | 1.94 | 0.053 |

The positive DiD coefficient is still a promising finding in terms of programme impact for girls, although in this case it also holds that if boys scores had dropped more than girls', there would be a more positive DiD coefficient. This could support a case that some of the gains in average scores for girls have come at the expense of boys, but is a difficult assertion to make at a national level.

**Primary 8 Literacy**
Although test scores for Primary 8 literacy are higher on average than any other grade or assessment type, and scores improved overall, the gap for girls in Primary 8 appears to have widened since 2014.

Average 2014 score: 61.96 (including no sex)
Average 2016 score: 63.53

*Table 16: Average Scores in Primary 8 Literacy*

| Year (P8 Literacy) | Girls av. score % | Boys av. score % | Sig between males and females |
|---|---|---|---|
| **2014** | 61.48 | 62.30 | 0.508 |
| **2016** | 62.36 | 64.40 | 0.025 |
| **Sig between 2014 & 2016** | 0.511 | 0.022 | |

Girls' scores were not significantly different from that of boys in 2014. By 2016, the gap had widened: girls' scores fell significantly below those of boys. The DiD test did not show a significant programme effect on P8 literacy.

*Table 17: DiD Effect in Primary 8 Literacy Scores*

| Diff-in-diff coeff. (score) | 95% Conf. Interval | Standard error | T | p-value |
|---|---|---|---|---|
| -0.012 | (-0.043, 0.018) | 0.016 | -0.78 | 0.437 |

**Secondary 2 Literacy**
Test scores for both girls and boys in Secondary 2 decreased since 2014 and the gender gap between girls' and boys' test scores remained. There is also no DiD effect.

Average 2014 score: 52.80% (including no sex)
Average 2016 score: 32.63%

*Table 18: Average Scores in Secondary 2 Literacy*

| Year (S2 Literacy) | Girls av. score % | Boys av. score % | Sig between gender (p-value) |
|---|---|---|---|
| **2014** | 50.50 | 53.54 | 0.000 |
| **2016** | 30.51 | 33.72 | 0.002 |
| **Sig between years** | 0.000 | 0.000 | |

*Table 19: DiD Effect in Secondary 2 Literacy Scores*

| DiD coef. | 95% Conf. Interval | St. error | t | p-value |
|---|---|---|---|---|
| -0.002 | (-0.027, 0.023) | 0.013 | -0.14 | 0.887 |

The DiD coefficient is small and not statistically significant.

## 3.1.    GESS Indicators

An increase in the mean test score by 0.1 standard deviations (of the mean of the baseline test) is a key indicator of the GESS programme. Results can be found in the following table.

*Table 20: GESS Indicator Progress*

| | Girls | | | Boys | | |
|---|---|---|---|---|---|---|
| | 2014 *mean score (SD)* | 2016 *mean score (SD)* | met target (+.1SD of 2014 mean) | 2014 *mean score (SD)* | 2016 *mean score (SD)* | met target (+.1SD of 2014 mean) |
| P5 Literacy | 51.6% (23.1) | 47.6% (20.1) | NO | 57.2% (22.1) | 50.3% (19.5) | NO |
| P8 Literacy | 61.5% (16.8) | 62.3% (18.4) | NO | 62.3% (15.4) | 64.4% (16.2) | YES |
| S2 Literacy | 50.5% (13.2) | 30.5% (16.7) | NO | 53.5% (11.4) | 33.7% (17.9) | NO |
| P5 Numeracy | 37.5% (15.8) | 48.6% (20.0) | YES | 43.5% (17.2) | 49.7% (19.0) | YES |
| P8 Numeracy | 37.1% (15.5) | 43.4% (18.5) | YES | 40.2% (15.6) | 41.6% (17.6) | NO |
| S2 Numeracy | 27.7% (11.3) | 27.5% (13.5) | NO | 30.9% (12.8) | 26.9% (12.2) | NO |

As mentioned in the Methodology section, this report used a single scoring approach for all assessments in both 2014 and 2016. One effect of implementing a consistent scoring approach has been that our results for 2014 scores for girls' numeracy are slightly lower than previously reported. However, the overall finding of the target outcome is unlikely to have changed if the original scores (as reported in the 2014 report using a slightly different scoring approach) had been used.

# 4. Conclusions and Recommendations

| | Conclusion | Recommendation |
|---|---|---|
| **Overall Results** | Aggregate results show positive effects for girls. This appears to be driven largely by girls' improvements in numeracy at Primary 5 and Primary 8 levels.<br><br>Overall, the largest gains are at the Primary 5 level. Further assessments could monitor whether this is related to lower attrition rates at the higher grades. | • Targeted support, and more of it, needed for Secondary grade levels.<br>• Include conflict-related variable(s) in the DiD model in the endterm analysis (to account for external factor of conflict/insecurity)<br>• Build in comparisons with SSSAMS enrolment data into endterm analyses |
| **Numeracy Scores** | Overall, scores in the numeracy assessments are positive. Regardless of grade, average scores increased to 40.23% in 2016. There is a statistically significant effect of the GESS programme in numeracy test scores across all grades. | • Work with MoGEI to assess whether the emphases in curriculum may have resulted in positive gains in numeracy but not in literacy<br>• Expand SSAMS monitoring to include notes on whether students leave school early in a given day (for instance, if mathematics is taught in the second half of a school session, students leaving early would miss this curriculum item) |
| **Literacy Scores** | Overall literacy scores remain higher than numeracy scores. However, there was no statistical effect on girls' literacy test scores, while numeracy scores did show a statistically positive change.<br><br>Average scores of literacy assessments dropped to 49.06% in 2016. | |
| **Test Administration** | While Invigilators insist at trainings that they understand their responsibilities, there are several areas in which data quality has suffered as a result of Invigilators and/or the Proctors, who are trained by the Invigilators, not following procedures. | • Develop measures to ensure compliance with testing administration procedures such as: mandating reporting; making Invigilator payment based on submitting reports; providing incentives for returning tests with all demographic items, especially sex, filled<br>• Increase number supervision at test sites |

# 5. Annex

## 5.1.    Subsamples of DiD Testing

*The following tables show the aggregate test results among **only states that were in both the 2014 and 2016 samples, and among only schools that were in both the 2014 and 2016 samples. In other words, the following tables do not include schools in Jonglei, Unity, or Upper Nile in the 2016 results.***

*Table 21: Average Scores of States Included in 2014 & 2016 Samples*

|      | Girls  | Boys   |
|------|--------|--------|
| 2014 | 44.2%  | 46.7%  |
| 2016 | 46.7%  | 45.5%  |

*Table 22: Average Scores at Schools Included in 2014 & 2016 Samples*

|      | Girls  | Boys   |
|------|--------|--------|
| 2014 | 45.2%  | 47.1%  |
| 2016 | 47.7%  | 45.0%  |

*Approx. 30 schools were in both the 2014 and 2016 samples

## 5.2.    Facility and Discrimination Rates

This section includes facility and discrimination rates for each question in the learning assessments. Question numbers refer to their order of appearance in the 2014 assessments.

As mentioned, all figures and percentages cited in this report were calculated with the most common approach used in the 2014 Baseline. This approach excludes questions in which the students' answer is unclear from the denominator, but includes those that were left unanswered (omitted) when calculating a students' total score. However, it is also possible to penalize both omitted and unclear answers, effectively including every question in the test while only rewarding correct answers. Scores and rates from this approach ("Method 2- Alt. Method") have been included in the following tables for comparison.

It should also be noted that the discrimination rates vary slightly from those presented in the 2014 Baseline Report due to cut-offs for selecting high and low-scoring groups within a given population. The percentage used in the 2014 Baseline Report cannot be confirmed by the author; the present tables grouped the highest and lowest 27% of each grade level.

## Literacy items discussion

**Primary 5 Literacy**
**Facility Rate**                                                    **Discrimination Rates**

| Q | Method 1 - (2014 Baseline) | | Combined | Method 2 - Alt. Method | | Combined | Method 1 - (2014 Baseline) | | Method 2 - Alt. Method | |
|---|---|---|---|---|---|---|---|---|---|---|
|   | 2014 | 2016 | ned | 2014 | 2016 | ned | 2014 | 2016 | 2014 | 2016 |
| 1 | 0.8358714 | 0.7609265 | 0.7864553 | 0.8233333 | 0.7411665 | 0.768949 | 0.4259259 | 0.5885168 | 0.4259259 | 0.6324921 |
| 2 | 0.7706577 | 0.737489 | 0.7488453 | 0.7616667 | 0.715198 | 0.7309101 | 0.4722222 | 0.6570973 | 0.4783951 | 0.6798107 |
| 3 | 0.3477157 | 0.3346457 | 0.3391004 | 0.3425 | 0.3256705 | 0.3313609 | 0.5030864 | 0.6299841 | 0.5030864 | 0.6041009 |
| 4 | 0.597479 | 0.1940559 | 0.3320874 | 0.5925 | 0.1890166 | 0.3254438 | 0.6018519 | -0.2424242 | 0.6172839 | -0.2129337 |
| 5 | 0.5226891 | 0.5378225 | 0.5326431 | 0.5183333 | 0.5236271 | 0.5218371 | 0.6419753 | 0.7591707 | 0.6574074 | 0.7539432 |
| 6 | 0.5092749 | 0.3970976 | 0.4355491 | 0.5033333 | 0.3844189 | 0.4246266 | 0.6018519 | 0.3317384 | 0.6172839 | 0.351735 |
| 7 | 0.6576955 | 0.659071 | 0.6585999 | 0.6516666 | 0.6402724 | 0.6441251 | 0.5648148 | 0.437016 | 0.558642 | 0.4873817 |
| 8 | 0.5994962 | 0.6215979 | 0.6140098 | 0.595 | 0.6028097 | 0.6001691 | 0.5030864 | 0.6028708 | 0.5123457 | 0.6293375 |
| 9 | 0.4625736 | 0.5513158 | 0.5208994 | 0.4583333 | 0.5351213 | 0.5091575 | 0.5833333 | 0.7878788 | 0.5895061 | 0.7870662 |
| 10 | 0.4760303 | 0.5503502 | 0.5249064 | 0.4716667 | 0.5351213 | 0.5136658 | 0.5370307 | 0.6842105 | 0.5246913 | 0.6861199 |
| 11 | 0.4369748 | 0.4226398 | 0.4275446 | 0.4333333 | 0.4116645 | 0.4189913 | 0.7283915 | 0.7304625 | 0.7253087 | 0.7429022 |
| 12 | 0.3920742 | 0.4811569 | 0.450692 | 0.3875 | 0.4674329 | 0.4404058 | 0.5987654 | 0.5821372 | 0.6018519 | 0.6182965 |
| 13 | 0.6647009 | 0.638549 | 0.647482 | 0.6575 | 0.6219668 | 0.6339814 | 0.5956709 | 0.6937799 | 0.6018519 | 0.6971609 |
| 14 | 0.5956192 | 0.1728665 | 0.3173963 | 0.5891666 | 0.1681567 | 0.31051 | 0.5462963 | -0.0733652 | 0.5709876 | -0.0883281 |
| 15 | 0.541876 | 0.5126859 | 0.5227011 | 0.5391667 | 0.4989357 | 0.5125387 | 0.7808642 | 0.7288677 | 0.7808642 | 0.7397476 |
| 16 | 0.3738397 | 0.2657005 | 0.3077428 | 0.3691667 | 0.210728 | 0.2642998 | 0.4320988 | 0.0590112 | 0.4290124 | 0.0883281 |

**Primary 8 Literacy**

| | **Facility Rate** | | | | | | **Discrimination Rates** | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Method 1 - 2014 Baseline Method | | | Method 2 - Alt. Method | | | Method 1 - 2014 Baseline Method | | Method 2 - Alt. Method | |
| **Q** | **2014** | **2016** | **Combined** | **2014** | **2016** | **Combined** | **2014** | **2016** | **2014** | **2016** |
| 1 | 0.5816 | 0.6512 | 0.6279 | 0.5801 | 0.6490 | 0.6259 | 0.5423 | 0.6357 | 0.5174 | 0.6281 |
| 2 | 0.9001 | 0.8319 | 0.8548 | 0.8977 | 0.8296 | 0.8524 | 0.2338 | 0.3367 | 0.2537 | 0.3442 |
| 3 | 0.8070 | 0.7886 | 0.7948 | 0.8048 | 0.7875 | 0.7933 | 0.3731 | 0.4171 | 0.3731 | 0.4146 |
| 4 | 0.7642 | 0.7655 | 0.7650 | 0.7631 | 0.7644 | 0.7640 | 0.4925 | 0.5151 | 0.4925 | 0.5075 |
| 5 | 0.4953 | 0.4458 | 0.4624 | 0.4926 | 0.4440 | 0.4603 | 0.4975 | 0.5352 | 0.4826 | 0.5276 |
| 6 | 0.2534 | 0.3849 | 0.3409 | 0.2517 | 0.3836 | 0.3394 | 0.3383 | 0.4246 | 0.3184 | 0.4221 |
| 7 | 0.5425 | 0.6560 | 0.6179 | 0.5410 | 0.6524 | 0.6151 | 0.6468 | 0.6030 | 0.6169 | 0.5804 |
| 8 | 0.7598 | 0.8206 | 0.8002 | 0.7577 | 0.8167 | 0.7969 | 0.5174 | 0.3844 | 0.5224 | 0.4020 |
| 9 | 0.8057 | 0.8769 | 0.8530 | 0.8035 | 0.8751 | 0.8511 | 0.4975 | 0.3291 | 0.4677 | 0.3342 |
| 10 | 0.8448 | 0.8799 | 0.8681 | 0.8425 | 0.8758 | 0.8646 | 0.4179 | 0.3442 | 0.3980 | 0.3543 |
| 11 | 0.6248 | 0.6005 | 0.6087 | 0.6231 | 0.5940 | 0.6038 | 0.3881 | 0.3920 | 0.3930 | 0.3769 |
| 12 | 0.2267 | 0.2580 | 0.2475 | 0.2261 | 0.2566 | 0.2464 | 0.2239 | 0.2789 | 0.2338 | 0.2764 |
| 13 | 0.2291 | 0.2842 | 0.2651 | 0.2288 | 0.2695 | 0.2559 | 0.1045 | 0.2312 | 0.1095 | 0.2362 |
| 14 | 0.8473 | 0.8023 | 0.8174 | 0.8439 | 0.7990 | 0.8141 | 0.3483 | 0.4497 | 0.3731 | 0.4548 |
| 15 | 0.6781 | 0.6322 | 0.6477 | 0.6662 | 0.6137 | 0.6313 | 0.3682 | 0.3492 | 0.3881 | 0.3417 |
| 16 | 0.2542 | 0.2697 | 0.2645 | 0.2463 | 0.2600 | 0.2554 | 0.1990 | 0.2965 | 0.2090 | 0.2789 |
| 17 | 0.8222 | 0.8276 | 0.8258 | 0.7968 | 0.7950 | 0.7956 | 0.3582 | 0.3970 | 0.3980 | 0.4548 |
| 18 | 0.4730 | 0.4605 | 0.4647 | 0.4603 | 0.4467 | 0.4513 | 0.5771 | 0.5427 | 0.6070 | 0.5603 |
| 19 | 0.8738 | 0.8316 | 0.8459 | 0.8668 | 0.8147 | 0.8321 | 0.3532 | 0.4472 | 0.3582 | 0.4799 |

**Secondary 2**
**Literacy**

| | **Facility Rate** | | | | | | **Discrimination Rates** | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Method 1 - 2014 Baseline Method | | | Method 2 - Alt. Method | | | Method 1 - 2014 Baseline Method | | Method 2 - Alt. Method | |
| **Q** | **2014** | **2016** | **Combined** | **2014** | **2016** | **Combined** | **2014** | **2016** | **2014** | **2016** |
| 1 | 0.591 | 0.333 | 0.470 | 0.587 | 0.332 | 0.467 | 0.332 | 0.530 | 0.337 | 0.533 |
| 2 | 0.234 | 0.226 | 0.230 | 0.233 | 0.224 | 0.229 | 0.193 | 0.163 | 0.185 | 0.163 |
| 3 | 0.257 | 0.161 | 0.212 | 0.256 | 0.160 | 0.211 | 0.248 | 0.281 | 0.243 | 0.281 |
| 4 | 0.881 | 0.789 | 0.838 | 0.877 | 0.786 | 0.834 | 0.219 | 0.266 | 0.232 | 0.272 |
| 5 | 0.704 | 0.315 | 0.522 | 0.687 | 0.305 | 0.508 | 0.282 | 0.666 | 0.279 | 0.669 |
| 6 | 0.285 | 0.277 | 0.281 | 0.277 | 0.272 | 0.275 | 0.358 | 0.571 | 0.345 | 0.568 |
| 7 | 0.830 | 0.371 | 0.615 | 0.803 | 0.358 | 0.594 | 0.366 | 0.896 | 0.389 | 0.902 |
| 8 | 0.504 | 0.419 | 0.464 | 0.494 | 0.412 | 0.455 | 0.535 | 0.272 | 0.533 | 0.266 |
| 9 | 0.863 | 0.062 | 0.488 | 0.850 | 0.061 | 0.480 | 0.334 | 0.044 | 0.339 | 0.041 |
| 10 | 0.982 | 0.434 | 0.725 | 0.980 | 0.431 | 0.723 | 0.050 | 0.964 | 0.052 | 0.967 |
| 11 | 0.926 | 0.408 | 0.685 | 0.924 | 0.401 | 0.679 | 0.185 | 0.882 | 0.172 | 0.876 |
| 12 | 0.761 | 0.344 | 0.565 | 0.754 | 0.342 | 0.560 | 0.352 | 0.766 | 0.358 | 0.769 |
| 13 | 0.874 | 0.398 | 0.651 | 0.870 | 0.395 | 0.647 | 0.245 | 0.864 | 0.245 | 0.870 |
| 14 | 0.483 | 0.322 | 0.407 | 0.481 | 0.319 | 0.405 | 0.211 | 0.609 | 0.188 | 0.598 |
| 15 | 0.349 | 0.167 | 0.264 | 0.348 | 0.166 | 0.262 | 0.149 | 0.334 | 0.154 | 0.322 |
| 16 | 0.966 | 0.419 | 0.709 | 0.963 | 0.418 | 0.708 | 0.115 | 0.947 | 0.117 | 0.947 |
| 17 | 0.539 | 0.227 | 0.393 | 0.537 | 0.226 | 0.391 | 0.540 | 0.476 | 0.551 | 0.482 |
| 18 | 0.293 | 0.268 | 0.281 | 0.292 | 0.267 | 0.280 | 0.274 | 0.021 | 0.290 | 0.027 |
| 19 | 0.434 | 0.352 | 0.395 | 0.432 | 0.350 | 0.393 | 0.232 | 0.198 | 0.238 | 0.213 |
| 20 | 0.318 | 0.386 | 0.350 | 0.317 | 0.383 | 0.348 | 0.444 | 0.160 | 0.441 | 0.186 |
| 21 | 0.346 | 0.3 | 0.362 | 0.345 | 0.377 | 0.360 | 0.389 | 0.157 | 0.389 | 0.148 |

| Q | 2014 | 2016 | Combined | 2014 | 2016 | Combined | 2014 | 2016 | 2014 | 2016 |
|---|---|---|---|---|---|---|---|---|---|---|
| 22 | 0.433 | 0.806 | 0.373 | 0.431 | 0.305 | 0.372 | 0.170 | 0.473 | 0.164 | 0.459 |
| 23 | 0.522 | 0.370 | 0.451 | 0.520 | 0.368 | 0.449 | 0.350 | 0.367 | 0.368 | 0.373 |
| 24 | 0.569 | 0.586 | 0.577 | 0.566 | 0.583 | 0.574 | 0.488 | 0.204 | 0.483 | 0.216 |
| 25 | 0.413 | 0.278 | 0.350 | 0.411 | 0.276 | 0.348 | 0.418 | 0.598 | 0.423 | 0.604 |
| 26 | 0.280 | 0.329 | 0.303 | 0.279 | 0.327 | 0.302 | 0.368 | 0.163 | 0.342 | 0.166 |
| 27 | 0.427 | 0.262 | 0.349 | 0.425 | 0.260 | 0.348 | 0.313 | 0.373 | 0.332 | 0.361 |
| 28 | 0.417 | 0.281 | 0.351 | 0.354 | 0.255 | 0.307 | 0.298 | 0.340 | 0.337 | 0.334 |
| 29 | 0.267 | 0.216 | 0.243 | 0.245 | 0.201 | 0.225 | 0.251 | 0.154 | 0.266 | 0.160 |
| 30 | 0.083 | 0.114 | 0.097 | 0.083 | 0.110 | 0.095 | 0.034 | 0.074 | 0.031 | 0.065 |

## Numeracy items discussion

**Primary 5 Numeracy**

| | Facility Rate | | | | | | Discrimination Rates | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Method 1 - 2014 Baseline Method | | | Method 2 - Alt. Method | | | Method 1 - 2014 Baseline Method | | Method 2 - Alt. Method | |
| Q | 2014 | 2016 | Combined | 2014 | 2016 | Combined | 2014 | 2016 | 2014 | 2016 |
| 1 | 0.876 | 0.828 | 0.845 | 0.855 | 0.802 | 0.820 | 0.252 | 0.393 | 0.290 | 0.440 |
| 2 | 0.796 | 0.737 | 0.757 | 0.777 | 0.713 | 0.735 | 0.408 | 0.509 | 0.404 | 0.530 |
| 3 | 0.659 | 0.673 | 0.669 | 0.645 | 0.653 | 0.650 | 0.503 | 0.641 | 0.500 | 0.660 |
| 4 | 0.604 | 0.646 | 0.631 | 0.591 | 0.623 | 0.612 | 0.516 | 0.562 | 0.503 | 0.582 |
| 5 | 0.106 | 0.275 | 0.217 | 0.102 | 0.264 | 0.209 | 0.159 | 0.444 | 0.156 | 0.452 |
| 6 | 0.717 | 0.814 | 0.780 | 0.662 | 0.720 | 0.700 | 0.589 | 0.609 | 0.627 | 0.638 |
| 7 | 0.585 | 0.542 | 0.557 | 0.571 | 0.526 | 0.541 | 0.484 | 0.524 | 0.503 | 0.550 |
| 8 | 0.546 | 0.622 | 0.596 | 0.537 | 0.605 | 0.582 | 0.529 | 0.685 | 0.545 | 0.698 |
| 9 | 0.596 | 0.619 | 0.611 | 0.584 | 0.600 | 0.595 | 0.570 | 0.715 | 0.589 | 0.730 |
| 10 | 0.498 | 0.611 | 0.571 | 0.491 | 0.590 | 0.556 | 0.656 | 0.742 | 0.653 | 0.758 |
| 11 | 0.256 | 0.418 | 0.362 | 0.251 | 0.404 | 0.351 | 0.325 | 0.570 | 0.328 | 0.565 |
| 12 | 0.171 | 0.398 | 0.319 | 0.168 | 0.384 | 0.310 | 0.239 | 0.791 | 0.236 | 0.792 |
| 13 | 0.489 | 0.593 | 0.557 | 0.481 | 0.571 | 0.540 | 0.471 | 0.514 | 0.481 | 0.537 |
| 14 | 0.318 | 0.525 | 0.453 | 0.313 | 0.507 | 0.441 | 0.490 | 0.788 | 0.490 | 0.795 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 15 | 0.268 | 0.504 | 0.422 | 0.262 | 0.485 | 0.409 | 0.315 | 0.675 | 0.309 | 0.698 |
| 16 | 0.263 | 0.270 | 0.268 | 0.259 | 0.261 | 0.260 | 0.239 | 0.162 | 0.239 | 0.167 |
| 17 | 0.222 | 0.334 | 0.296 | 0.218 | 0.324 | 0.288 | 0.398 | 0.391 | 0.408 | 0.413 |
| 18 | 0.564 | 0.674 | 0.635 | 0.556 | 0.650 | 0.618 | 0.596 | 0.664 | 0.602 | 0.688 |
| 19 | 0.330 | 0.419 | 0.388 | 0.324 | 0.404 | 0.376 | 0.478 | 0.460 | 0.484 | 0.492 |
| 20 | 0.175 | 0.194 | 0.187 | 0.172 | 0.186 | 0.181 | -0.108 | 0.044 | -0.108 | 0.050 |
| 21 | 0.145 | 0.307 | 0.250 | 0.143 | 0.295 | 0.242 | 0.172 | 0.565 | 0.166 | 0.548 |
| 22 | 0.207 | 0.345 | 0.297 | 0.205 | 0.332 | 0.288 | 0.475 | 0.616 | 0.475 | 0.613 |
| 23 | 0.475 | 0.578 | 0.542 | 0.467 | 0.555 | 0.525 | 0.516 | 0.641 | 0.535 | 0.662 |
| 24 | 0.297 | 0.221 | 0.248 | 0.293 | 0.212 | 0.240 | 0.248 | -0.140 | 0.242 | -0.120 |
| 25 | 0.417 | 0.606 | 0.540 | 0.412 | 0.585 | 0.525 | 0.398 | 0.718 | 0.436 | 0.757 |
| 26 | 0.444 | 0.260 | 0.324 | 0.436 | 0.251 | 0.315 | 0.653 | 0.256 | 0.662 | 0.268 |
| 27 | 0.393 | 0.348 | 0.364 | 0.387 | 0.335 | 0.353 | 0.500 | 0.265 | 0.490 | 0.288 |
| 28 | 0.271 | 0.441 | 0.382 | 0.265 | 0.424 | 0.369 | 0.471 | 0.745 | 0.471 | 0.747 |
| 29 | 0.297 | 0.292 | 0.294 | 0.289 | 0.280 | 0.283 | 0.280 | 0.260 | 0.261 | 0.273 |
| 30 | 0.271 | 0.232 | 0.246 | 0.263 | 0.222 | 0.236 | 0.271 | 0.110 | 0.296 | 0.120 |

**Primary 8 Numeracy**

| | Facility Rate | | | | | | Discrimination Rates | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Method 1 - 2014 Baseline Method | | | Method 2 - Alt. Method | | | Method 1 - 2014 Baseline Method | | Method 2 - Alt. Method | |
| Q | 2014 | 2016 | Combined | 2014 | 2016 | Combined | 2014 | 2016 | 2014 | 2016 |
| 1 | 0.910 | 0.888 | 0.895 | 0.908 | 0.886 | 0.893 | 0.157 | 0.217 | 0.173 | 0.224 |
| 2 | 0.501 | 0.432 | 0.455 | 0.496 | 0.430 | 0.452 | 0.655 | 0.561 | 0.670 | 0.561 |
| 3 | 0.598 | 0.587 | 0.590 | 0.592 | 0.570 | 0.578 | 0.523 | 0.413 | 0.513 | 0.413 |
| 4 | 0.287 | 0.427 | 0.381 | 0.286 | 0.426 | 0.379 | 0.355 | 0.589 | 0.355 | 0.579 |
| 5 | 0.352 | 0.447 | 0.415 | 0.352 | 0.444 | 0.413 | 0.457 | 0.640 | 0.437 | 0.622 |
| 6 | 0.722 | 0.012 | 0.250 | 0.721 | 0.012 | 0.249 | 0.396 | -0.010 | 0.406 | -0.013 |
| 7 | 0.882 | 0.830 | 0.847 | 0.879 | 0.802 | 0.828 | 0.299 | 0.403 | 0.289 | 0.411 |
| 8 | 0.545 | 0.620 | 0.595 | 0.544 | 0.617 | 0.592 | 0.411 | 0.487 | 0.401 | 0.505 |
| 9 | 0.340 | 0.419 | 0.393 | 0.339 | 0.417 | 0.391 | 0.376 | 0.485 | 0.376 | 0.474 |
| 10 | 0.758 | 0.777 | 0.771 | 0.757 | 0.774 | 0.768 | 0.157 | 0.362 | 0.157 | 0.360 |
| 11 | 0.296 | 0.404 | 0.368 | 0.295 | 0.402 | 0.366 | 0.665 | 0.658 | 0.650 | 0.648 |
| 12 | 0.405 | 0.337 | 0.360 | 0.402 | 0.329 | 0.354 | 0.315 | 0.349 | 0.310 | 0.357 |
| 13 | 0.202 | 0.259 | 0.240 | 0.202 | 0.257 | 0.239 | 0.223 | 0.319 | 0.213 | 0.321 |
| 14 | 0.591 | 0.689 | 0.657 | 0.588 | 0.688 | 0.654 | 0.589 | 0.607 | 0.599 | 0.605 |
| 15 | 0.220 | 0.282 | 0.261 | 0.218 | 0.273 | 0.255 | 0.315 | 0.569 | 0.330 | 0.571 |
| 16 | 0.278 | 0.474 | 0.409 | 0.277 | 0.472 | 0.407 | 0.457 | 0.617 | 0.442 | 0.602 |
| 17 | 0.240 | 0.298 | 0.279 | 0.239 | 0.297 | 0.277 | 0.254 | 0.444 | 0.254 | 0.436 |
| 18 | 0.218 | 0.302 | 0.274 | 0.217 | 0.300 | 0.272 | 0.376 | 0.500 | 0.360 | 0.492 |
| 19 | 0.269 | 0.389 | 0.349 | 0.268 | 0.388 | 0.348 | 0.574 | 0.673 | 0.589 | 0.671 |
| 20 | 0.444 | 0.396 | 0.412 | 0.442 | 0.389 | 0.407 | 0.660 | 0.648 | 0.685 | 0.635 |
| 21 | 0.226 | 0.299 | 0.275 | 0.225 | 0.298 | 0.274 | 0.355 | 0.426 | 0.360 | 0.423 |
| 22 | 0.434 | 0.554 | 0.514 | 0.433 | 0.551 | 0.511 | 0.482 | 0.582 | 0.477 | 0.577 |
| 23 | 0.096 | 0.117 | 0.110 | 0.096 | 0.116 | 0.109 | 0.091 | 0.115 | 0.091 | 0.115 |
| 24 | 0.381 | 0.423 | 0.409 | 0.380 | 0.421 | 0.407 | 0.345 | 0.390 | 0.330 | 0.408 |
| 25 | 0.583 | 0.529 | 0.547 | 0.581 | 0.528 | 0.545 | 0.629 | 0.696 | 0.640 | 0.717 |

| Q | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 25 | | | | | | | | | | |
| 26 | 0.094 | 0.149 | 0.130 | 0.093 | 0.148 | 0.130 | 0.061 | 0.143 | 0.066 | 0.151 |
| 27 | 0.362 | 0.374 | 0.370 | 0.361 | 0.372 | 0.369 | 0.553 | 0.574 | 0.553 | 0.574 |
| 28 | 0.210 | 0.324 | 0.286 | 0.209 | 0.321 | 0.283 | 0.213 | 0.477 | 0.203 | 0.482 |
| 29 | 0.326 | 0.317 | 0.320 | 0.320 | 0.314 | 0.316 | 0.208 | 0.408 | 0.223 | 0.411 |
| 30 | 0.332 | 0.366 | 0.355 | 0.331 | 0.363 | 0.353 | 0.629 | 0.589 | 0.635 | 0.605 |
| 31 | 0.220 | 0.226 | 0.224 | 0.220 | 0.224 | 0.223 | 0.198 | 0.191 | 0.198 | 0.191 |
| 32 | 0.214 | 0.457 | 0.376 | 0.213 | 0.457 | 0.375 | 0.239 | 0.077 | 0.234 | 0.084 |

**Secondary 2 Numeracy**

| **Facility Rate** | | | | | | **Discrimination Rates** | | | |
|---|---|---|---|---|---|---|---|---|---|
| Method 1 - 2014 Baseline Method | | | Method 2 - Alt. Method | | | Method 1 - 2014 Baseline Method | | Method 2 - Alt. Method | |
| Q | 2014 | 2016 | Combined | 2014 | 2016 | Combined | 2014 | 2016 | 2014 | 2016 |
| 1 | 0.530 | 0.420 | 0.478 | 0.527 | 0.417 | 0.474 | 0.596 | 0.524 | 0.596 | 0.527 |
| 2 | 0.513 | 0.464 | 0.489 | 0.510 | 0.460 | 0.486 | 0.496 | 0.544 | 0.522 | 0.550 |
| 3 | 0.280 | 0.275 | 0.277 | 0.278 | 0.272 | 0.275 | 0.467 | 0.501 | 0.464 | 0.510 |
| 4 | 0.291 | 0.258 | 0.276 | 0.290 | 0.257 | 0.274 | 0.464 | 0.490 | 0.451 | 0.484 |
| 5 | 0.251 | 0.274 | 0.262 | 0.251 | 0.274 | 0.262 | 0.354 | 0.453 | 0.351 | 0.450 |
| 6 | 0.330 | 0.295 | 0.313 | 0.328 | 0.294 | 0.312 | 0.478 | 0.556 | 0.485 | 0.567 |
| 7 | 0.665 | 0.583 | 0.626 | 0.664 | 0.578 | 0.623 | 0.501 | 0.547 | 0.507 | 0.550 |
| 8 | 0.306 | 0.267 | 0.287 | 0.305 | 0.265 | 0.286 | 0.375 | 0.332 | 0.377 | 0.338 |
| 9 | 0.161 | 0.143 | 0.152 | 0.160 | 0.142 | 0.152 | 0.259 | 0.298 | 0.259 | 0.301 |
| 10 | 0.190 | 0.194 | 0.192 | 0.189 | 0.188 | 0.189 | 0.214 | 0.252 | 0.219 | 0.261 |
| 11 | 0.256 | 0.244 | 0.250 | 0.255 | 0.241 | 0.248 | 0.385 | 0.484 | 0.388 | 0.484 |
| 12 | 0.339 | 0.384 | 0.361 | 0.337 | 0.382 | 0.359 | 0.464 | 0.519 | 0.467 | 0.536 |
| 13 | 0.310 | 0.254 | 0.283 | 0.308 | 0.252 | 0.281 | 0.293 | 0.281 | 0.296 | 0.278 |
| 1 | 0.344 | 0.357 | 0.350 | 0.342 | 0.356 | 0.349 | 0.343 | 0.421 | 0.338 | 0.421 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 4 | | | | | | | | | | |
| 15 | 0.182 | 0.235 | 0.207 | 0.181 | 0.234 | 0.206 | 0.303 | 0.436 | 0.309 | 0.427 |
| 16 | 0.419 | 0.408 | 0.414 | 0.419 | 0.405 | 0.412 | 0.224 | 0.427 | 0.227 | 0.421 |
| 17 | 0.612 | 0.539 | 0.577 | 0.611 | 0.534 | 0.574 | 0.536 | 0.556 | 0.528 | 0.562 |
| 18 | 0.402 | 0.307 | 0.357 | 0.400 | 0.303 | 0.354 | 0.293 | 0.252 | 0.303 | 0.241 |
| 19 | 0.168 | 0.173 | 0.171 | 0.168 | 0.171 | 0.169 | 0.127 | 0.195 | 0.132 | 0.183 |
| 20 | 0.030 | 0.023 | 0.027 | 0.030 | 0.023 | 0.027 | 0.037 | 0.020 | 0.037 | 0.014 |
| 21 | 0.149 | 0.160 | 0.154 | 0.148 | 0.159 | 0.153 | 0.124 | 0.223 | 0.137 | 0.215 |
| 22 | 0.246 | 0.266 | 0.256 | 0.245 | 0.264 | 0.254 | 0.219 | 0.289 | 0.230 | 0.284 |
| 23 | 0.297 | 0.223 | 0.261 | 0.295 | 0.221 | 0.260 | 0.198 | 0.112 | 0.179 | 0.123 |
| 24 | 0.186 | 0.184 | 0.185 | 0.185 | 0.183 | 0.184 | 0.351 | 0.112 | 0.346 | 0.112 |
| 25 | 0.121 | 0.108 | 0.115 | 0.120 | 0.107 | 0.114 | 0.092 | 0.106 | 0.084 | 0.109 |
| 26 | 0.470 | 0.257 | 0.368 | 0.468 | 0.256 | 0.366 | 0.158 | 0.089 | 0.153 | 0.077 |
| 27 | 0.412 | 0.254 | 0.337 | 0.411 | 0.253 | 0.335 | 0.588 | 0.330 | 0.586 | 0.330 |
| 28 | 0.159 | 0.153 | 0.156 | 0.158 | 0.152 | 0.155 | 0.177 | 0.138 | 0.172 | 0.135 |
| 29 | 0.270 | 0.241 | 0.256 | 0.269 | 0.238 | 0.254 | 0.222 | 0.238 | 0.237 | 0.238 |
| 30 | 0.309 | 0.234 | 0.273 | 0.307 | 0.231 | 0.271 | 0.053 | 0.052 | 0.047 | 0.060 |
| 31 | 0.248 | 0.206 | 0.228 | 0.248 | 0.205 | 0.227 | 0.061 | 0.037 | 0.042 | 0.034 |
| 32 | 0.219 | 0.257 | 0.237 | 0.218 | 0.255 | 0.236 | 0.000 | 0.163 | 0.000 | 0.175 |

## 5.3. State-level Disaggregation

The learning assessment sample was not designed to be representative at the state level. The following data is therefore not robust enough to be indicative of any significant differences between individual states but has been included as a matter of interest.

*Table 23: Average Test Scores by State*

|  | Girls | | Boys | |
|---|---|---|---|---|
|  | 2014 (% score) | 2016 (% score) | 2014 (% score) | 2016 (% score) |
| Central Equatoria | 48.1 | 48.8 | 51.2 | 49.2 |
| Western Equatoria | 33.0 | 42.4 | 39.0 | 45.7 |
| Eastern Equatoria | 41.3 | 36.1 | 43.5 | 42.3 |
| Northern Bahr el Ghazal | 36.6 | 45.6 | 39.6 | 43.0 |
| Western Bahr el Ghazal | 37.2 | 31.4 | 47.9 | 31.3 |
| Lakes | 56.5 | 52.5 | 48.7 | 49.9 |
| Warrap | 41.6 | 54.5 | 46.0 | 49.4 |
| Upper Nile | N/A | 29.4 | N/A | 29.7 |
| Unity | N/A | 51.2 | N/A | 51.4 |
| Jonglei | N/A | 51.6 | N/A | 50.8 |

*The Lakes observations included the exact same schools in 2014 & 2016.

## Forcier

Forcier is a development research firm that operates in challenging post-conflict environments. Established in 2011 in South Sudan, Forcier has invested in developing methodologies and approaches to research that are contextually appropriate and feasible, whilst adhering to international standards for social science research and utilising the latest data collection technology available. Our core services include population and social science research, project evaluations, market assessments for livelihoods and vocational trainings, private sector and market research for feasibility studies, strategic planning and representation, and training and capacity building workshops.
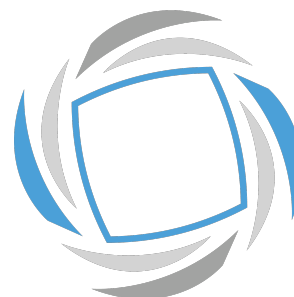
For further information, please visit www.forcierconsulting.com.