EQUATING STUDY: LARS IV, EGRA 2018, LEGRA 2021

USAID SOMA UMENYE

EQUATING STUDY: LARS IV, EGRA 2018, LEGRA 2021

USAID SOMA UMENYE

DISCLAIMER

CONTENTS

EXECUTIVE SUMMARY

USAID Soma Umenye has supported REB (and latterly the National Examination and School Inspection Authority - NESA) to improve the effectiveness of national and term assessment of Kinyarwanda early grade reading.

Following a collaborative approach to design and delivery, Soma Umenye facilitated REB development of a revised set of early grade reading assessment tools in 2018, which were used in the baseline assessment in August 2018. In 2019, the project supported the revision of the Kinyarwanda early grade reading learner benchmarks which were approved in 2019, and developed the 'local early grade reading assessment' (LEGRA) which was piloted in every primary school in five districts and is now a part of NESA's comprehensive assessment framework. In 2020, the project co-developed LARS IV P3 Kinyarwanda early grade reading assessment instruments for NESA. LARS IV P3 Kinyarwanda assessment was delivered in March 2021.

The LEGRA and LARS early grade reading assessment instruments were developed from the same revised set of early grade reading assessment tools developed in 2018. While LARS, LEGRA and EGRA are different assessments that are delivered differently (with LARS and EGRA having more in common as they are sample based assessments delivered by an external assessor whereas LEGRA is a census assessment delivered by teachers themselves), this study sets out to present a equating equation, allowing NESA to compare LARS and LEGRA data with each other and with the EGRA baseline.

As at December 2021, NESA has assessed Kinyarwanda early grade reading of all P1-P3 learners nationwide through LEGRA and also has a LARS assessment of P3 learners. We know that 4% of P3 learners met the new Kinyarwanda benchmark in the 2018 EGRA baseline. LARS IV suggests that 54% of learners achieved the benchmark. But can we compare this 4% with the 54%?

This study assesses the EGRA, LEGRA and LARS instruments and presents an 'equating table' which when applied, would allow data from the different assessments to be compared. For example, when the equating equation is applied to LARS IV, the EGRA P3 ORF benchmark of 40 correct words per minute is equivalent to 51 correct words per minute through LARS assessment. So, the LARS 54% of P3 learners meeting the ORF benchmark is equivalent to 29% meeting the benchmark when the EGRA/LARS equating is applied.

The purpose of the study is also to stimulate discussion about assessment of early grade reading. Rwanda's delivery of LEGRA allows teachers immediate access to assessment findings which they can reflect on and apply remediation. Schools, sectors, district and national level can utilize LEGRA data as diagnostics to target areas for improved performance. If there is a trusted way to compare LEGRA and LARS data, this could be an exciting opportunity to drive assessment for learning and potentially drive improvement in learner outcomes towards the Government of Rwanda's human capital goals.

INTRODUCTION AND CONTEXT

Soma Umenye is a five-year USAID-funded activity from 2016 through 2021 that aims to improve reading outcomes in Kinyarwanda for at least 1 million children in public and government-aided schools in Rwanda. Covering all 30 districts of Rwanda and working in close collaboration with the Rwanda Basic Education Board (REB), the National Examination and School Inspection Authority (NESA) and the Rwandan Ministry of Education (MINEDUC), Soma Umenye contributes to Rwanda's Education Sector Strategic Plan (ESSP) and the Government of Rwanda's national development priority of ensuring that Rwandan primary-grade students acquire the fundamental competency of literacy so that they can succeed in future schooling and, later, in the modern workplace.

Soma Umenye aims to achieve the following two goals.

- To improve reading outcomes in Kinyarwanda for at least 1 million children (unique direct beneficiaries) in public and government-aided schools in Rwanda by the end of P3; and
- To ensure that at least 70 percent of P1-P3 students are able to read grade-level text with fluency and comprehension.

CONTEXT OF THIS STUDY

In 2019, MINEDUC introduced a comprehensive assessment system with the main purpose of tracking student performance through formative and summative assessments so that struggling learners are identified early and provided with the necessary support through remediation activities. Building on the EGRA baseline work done in 2018 where a complete set of instruments for future early grade reading assessments was developed to be adopted and adapted for future studies (LARS IV, LEGRA, etc), USAID Soma Umenye supported REB and NESA with its implementation as follows:

- Development of the Local Early Grade Reading Assessment (LEGRA) where every P1-P3 student is assessed on key literacy subskills at the end of Terms 1, 2 and 3
- Development of end-of-term benchmarks against which performance on the LEGRA is measured
- Development of LEGRA assessment items
- Development of annual benchmarks against which performance on the LARS and LEGRA is measured
- Development of question items for teachers to draw from and support their students throughout the year. Parents can also access these resources and help their children.

Since the joint development of the EGRA baseline instruments in 2018, Rwanda has conducted five early grade Kinyarwanda reading assessments:

- September 2019: LEGRA pilot (sample: 176,000 P1-P3 students)

- March 2021: LARS (sample: 715 P4 students, tested on P3 items)
- March 2021: Term 1 LEGRA (all P1-P3 students)
- June 2021: Term 2 LEGRA (all P1-P3 students)
- August/September 2021: Term 3 LEGRA (all P1-P3 students)

Exhibit 1 describes the three different tools that will be used in this equating study. Equating will be conducted only on the reading fluency and reading comprehension subtasks.

Exhibit 1. Description of the Three Different Tools

	LEGRA Instrument	**2018 EGRA Instrument**	**LARS IV Instrument**
Administration	In schools, by teacher	Soma Umenye data collectors	Soma Umenye data collectors
Assessment items	Full assessment One ORF passage and questions, dictation, decoding, and recognizing letters/sounds (P1 term 1)	One ORF passage and comprehension questions	One ORF passage and comprehension questions
Frequency	End of every term (3 times a year)	Often dependent on USAID program cycles	Every 2 years
Primary users of the data	Teachers and school leaders	Donors and implementing partners	MINEDUC, NESA, REB
Primary goal of the assessment	To give teachers an understanding of where there students are at the end of each term so they can plan remediation for the following term.	To provide donors and education leaders an overview of learning achievement progress and often to benchmark impact of a specific early grade reading project.	To provide education leaders an overview of learning achievement progress toward national goals and support regional and international benchmarking.

The proposed equating process and reflection on the findings described in this report will give the Rwandan education system an opportunity to review its comprehensive assessment system and determine the extent to which its formative assessments, implemented through LEGRA for early grade reading in Kinyarwanda, produce data that can be compared to summative assessments (LARS). By ensuring the LEGRA and LARS assessments are comparable, the system players will be able to use them to understand trends, challenges and take necessary decisions to improve learner outcomes. We have included the EGRA administered in 2018 with Soma Umenye support in this report in order to enable comparison between the project's EGRA baseline and later LARS and LEGRA assessments.

OBJECTIVES OF THE EQUATING STUDY

The purpose of this study is to support Soma Umenye to equate (1) the LEGRA assessment (2) the LARS IV assessment, and (3)the 2018 EGRA .

As mentioned, the LEGRA tool is used at the end of Terms 1-3, 2 and 3 to help teachers understand whether students are making progress towards end of term performance benchmarks. The LARS assessment is used to measure student performance every two years. The EGRA (administered at the end of the year) was used in 2018 to set the baseline for Soma Umenye.

METHODOLOGY

In this section, the instruments, sample, and analysis approach are discussed.

INSTRUMENTS

As noted, Soma Umenye aims to equate the three assessments, LARS, EGRA, and LEGRA, for the following reasons:

- It would enable government of Rwanda officials and school personnel to understand how performance on the LEGRA (administered each term) translates to performance on the LARS, which helps them understand whether students are progressing towards end-of-term benchmarks measured by the LARS against annual benchmarks.

- It would enable Soma Umenye to construct a measure of end-of-project performance by enabling the measure of change between the baseline and an endline measure (which could be the LARS IV or the September 2021 LEGRA).

To develop this equating (comparing LARS, LEGRA, and EGRA), we draw on data from (1) a Soma Umenye-led administration of the fluency and comprehension subtests of the LARS and EGRA in August 2021 and (2) the school administration of the LEGRA in September 2021.

In addition, Soma Umenye data collectors collected data on student identifiers (name, age, and identification number) so that the LEGRA data could be matched with EGRA and LARS.

SAMPLE

In order to conduct an equating study and to be able to generalize to the population of Rwandan students, the best practice would be to select a sample that is representative of all provinces. Given the restrictions caused by COVID-19, it was not possible to collect data from every district of the country. Schools were closed in many districts when data needed to be collected. For this reason, data were collected only in three districts of Kigali City and some neighboring districts (Bugesera, Kamonyi, Rulindo and Rwamagana).

The students for this equating study came from 39 randomly selected schools. In each school, Soma Umenye data collectors collected data from 12 Grade 3 students (6 boys/6 girls). Grade 3 was selected for the this study because LARS was only administered at this grade. A total of 461 students participated in the EGRA/LARS assessment while data for LEGRA was collected for 448 students. The age of the students varied between 8 and 15 years old with an average age of 10.7 years old. The sample is composed of 229 (49.7%) boys and 232 (50.3%) girls.

TRAINING OF DATA COLLECTORS

From a pool of project staff, Soma Umenye selected data collectors to attend a three-day training for this study. EdIntersect provided remote facilitation of the training while the Soma Umenye team provided on-site facilitation and follow-up on practice sessions.

DATA COLLECTION

Data collectors worked in pairs in schools to collect data for EGRA and LARS passages. They collected data over 5 days (August 30 to September 3, 2021), visiting each school in the sample. Enumerators were trained to randomly sample students in the selected class and then administer the required items. In order to have one icebreaker activity before starting the subtasks for this study, it was decided to also include listening comprehension as the first subtask to be administered. As this subtask was not part of the equating study, the data from the listening comprehension subtask were not analyzed and are not part of this report.

For LEGRA, since teachers are responsible for conducting the assessment for all schools across the country and for reporting results, data were not collected using the same method as for EGRA and LARS. Instead, when data for EGRA and LARS were collected by enumerators, the participating students were identified by their name and unique identification number on a data collection form. Once the LEGRA was administered and scored, teachers were then responsible for completing the reporting form, including recording the scores of the sampled students for the fluency and reading comprehension subtasks. Teachers sent the completed forms to Soma Umenye staff by WhatsApp. Soma Umenye compiled the scores, captured them in Excel files, and then merged the LEGRA scores with the EGRA and LARS datasets using the student unique identification number. Thus, the evaluation team writing this report did not have access to the name of the students assessed, only to their identification numbers.

ANALYSIS

Analysis of the data was conducted using a two-step approach. In the first step, a descriptive analysis of the performance of the students on each instrument was conducted. This first analysis served to explore the data and see differences in distribution of the scores between the three different instruments. At this step, we also compared the performance of the students by gender. Correlations between oral fluency scores on the different passages and performance on reading comprehension questions were also computed as well as cumulative distribution functions for fluency and reading comprehension scores.

Equating for the fluency scores and reading comprehension scores was conducted using the single group kernel approach (van Davier, *et al.*, 2004). This equating method is an observed-score test equating based on the percentile distribution of the fluency scores on each instrument and includes a pre-smoothing approach to correct for inconsistencies in score distribution on the different oral reading fluency or reading comprehension scores. Once satisfied by the equating method, tables were produced to transform scores on one instrument to the scale of value of the other. It was only possible to run equating with two instruments at a time, so equating was conducted pairwise (EGRA to LARS, EGRA to LEGRA, and LARS to LEGRA).

LIMITATIONS

As described, it was not possible to collect data for a representative sample of all provinces of Rwanda. In addition, due to pandemic restrictions on travel and gathering in a meeting space, there were logistical limitations on training of data collectors. Given these limitations, data were collected from 7 districts in 4 provinces: Kigali City (Gasabo, Kicukiro and Nyarugenge districts), Southern Province (Kamonyi district), Northern Province (Rulindo District), and Eastern Province (Bugesera and Rwamagana districts). The equating formula to transform scores on the scale of value of one instrument to another should be used with caution. Nonetheless, this equating study did provide interesting information on the comparability of the instruments and the possibility of reporting results from the different tools on a common scale.

It is important to remember, as referenced earlier in this report, that the LARS, EGRA, and LEGRA are different assessments with different purposes. While LARS and EGRA are more similar in that they are administered by external enumerators, the LEGRA is administered by P1-P3 teachers. Finally, given the logistical difficulties posed by COVID-19, enumerators for this equating study (who administered LARS and EGRA) had limited time for training and were required to engage virtually with core trainers.

RESULTS

This section presents all results obtained from the analysis of the data collected for this equating study. First, descriptive results for all instruments are fully presented and explained. Then, results from the equating analysis are presented. Those results focus on transforming scores collected on LARS to be put on the EGRA scale of values and also transforming scores collected on LEGRA to be put on the EGRA and LARS scales of values. This section ends with the table to transform scores from LARS to EGRA scale and scores from LEGRA to EGRA and LARS scales.

DESCRIPTIVE RESULTS ON THE PERFORMANCE OF THE STUDENTS ON EACH INSTRUMENT

Before presenting the results, it is important to be reminded of the approved ORF and reading comprehension annual benchmarks which were co-developed by REB and Soma Umenye in 2019.

Exhibit 2 contains approved national fluency benchmarks, P1 to P3, measured in correct words read per minute (CWPM).

Exhibit 2. Approved Oral Reading Fluency Benchmarks, P1 to P3 (End of Year)

	Below Categorization	Does not meet expectations	Partially meets expectations	Meets Expectations	Exceeds expectations	Benchmarks
P1	0 CWPM	1 to 6 CWPM	7 to 9 CWPM	10 to 20 CWPM	21+ CWPM	10 CWPM
P2	0 CWPM	1 to 9 CWPM	10 to 24 CWPM	25 to 35 CWPM	36+ CWPM	25 CWPM
P3	0 CWPM	1 to 17 CWPM	18 to 39 CWPM	40 to 50 CWPM	51+ CWPM	40 CWPM

Exhibit 3 contains approved national reading comprehension benchmarks for P1, P2 and P3, measured in percent of questions correctly answered.

Exhibit 3. Approved Reading Comprehension Benchmarks, P1 to P3 (End of Year)

	Below categorization	Does not meet expectations	Partially meets expectations	Meets expectations	Exceed expectations	Benchmarks
P1	0 questions answered correctly	1 question answered correctly	2 questions answered correctly	3 questions correctly answered	4 or 5 questions correctly answered	60%
P2	0 questions answered correctly	1 question answered correctly	2 questions answered correctly	3 questions correctly answered	4 or 5 questions correctly answered	60%
P3	0 questions answered correctly	1 or 2 questions answered correctly	3 questions answered correctly	4 questions correctly answered	5 questions correctly answered	80%

ORAL READING FLUENCY

The percentage of students with zero scores on the reading fluency subtask is similar for each instrument - from 4.3% for the LARS to 5.6% for LEGRA. On the other hand, average oral reading fluency scores show important differences between instruments. As Exhibits 4

and 5 show below, LARS had the lowest percentage of zero scores and EGRA was the instrument with the lowest average score with 26.6 correct words per minute and LEGRA the largest average with 45.8 correct words per minute. On average, students were reading almost twice the words in the LEGRA instrument compared to EGRA. Average scores on the LARS are between the performance on the EGRA and the LEGRA with an average of 33.9 correct words per minute.

Exhibit 4. Percentage of Students with Zero Scores on Oral Reading Fluency for Each Instrument

	EGRA 2018 (n=461)	LARS (n=461)	LEGRA (n=448)
Zero score	5.2%	4.3%	5.6%

Exhibit 5. Average Oral Reading Fluency Scores for Each Instrument

	N	Mean	Standard Deviation
EGRA 2018	461	26.6	13.3
LARS	461	33.9	17.1
LEGRA	448	45.8	24.2

In addition to important differences between average scores, we also observed differences between variability of the fluency scores for each instrument. As standard deviations suggest in Table 5 above and the histograms shown in Exhibits 6 to 8 below, there are more differences between the performance of the students for the LEGRA than for the two other instruments. The higher variability on the LEGRA can be explained in part by the fact that more students present high fluency scores on the LEGRA than what is observed on the EGRA and LARS.

Exhibit 6. Distribution of Oral Reading Fluency Scores for EGRA 2018
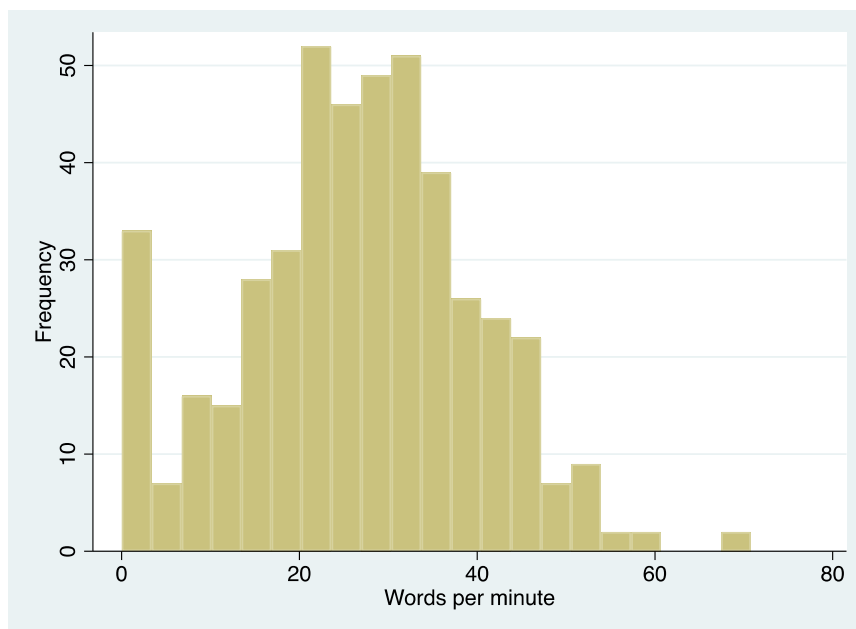
Exhibit 7. Distribution of Oral Reading Fluency Scores for LARS

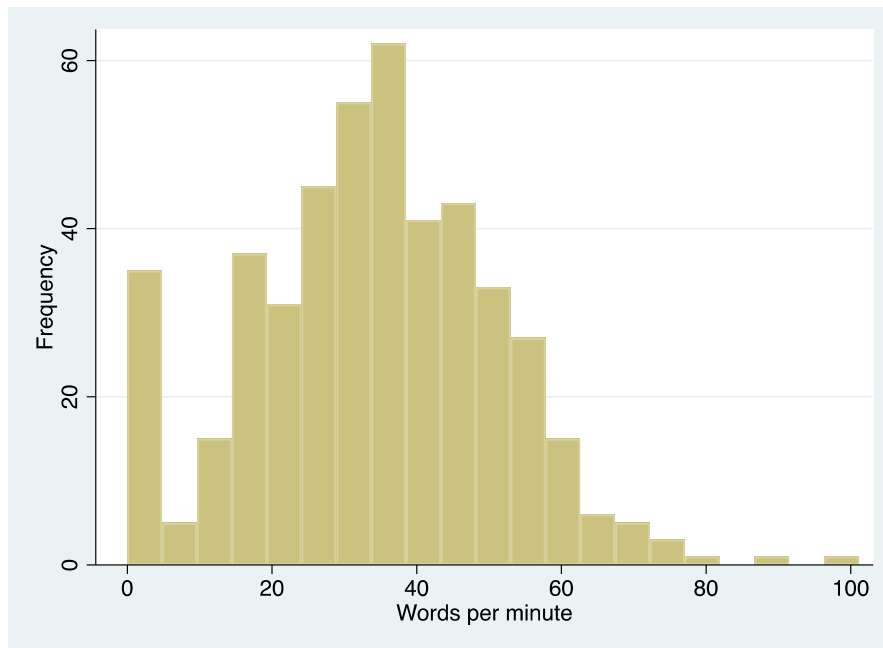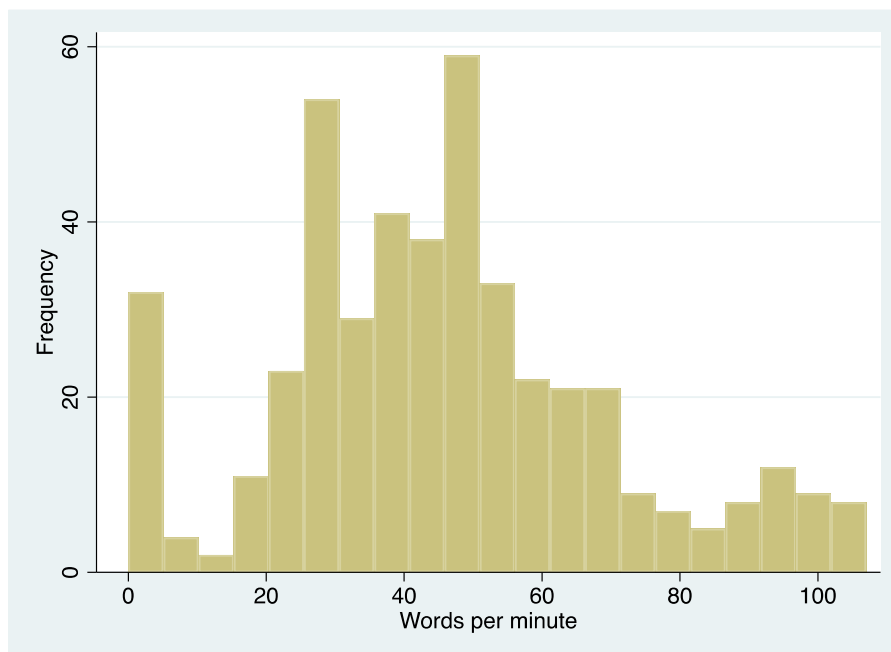**Exhibit 8. Distribution of Oral Reading Fluency Scores for LEGRA**

When comparing the performance of the students between genders as seen below in Exhibit 9, no significant difference is observed in the proportion of students with zero scores for all three instruments. The observed proportions of zero scores are higher for boys than girls, but those differences are not statistically significant.

Exhibit 9. Percentage of Students with Zero Scores by Gender for Each Instrument

	Boys	Girls	Total	p-value
EGRA 2018 (n=461)	7.0%	3.5%	5.2%	0.087
LARS (n=461)	6.1%	2.6%	4.3%	0.063
LEGRA (n=448)	7.7%	3,.5%	5.6%	0.058

Regarding the differences in the performance of boys and girls on average correct words per minute below in Exhibit 10, girls present higher average fluency scores than boys on the EGRA and LARS. For the LEGRA, while the girls also show a higher average score, the difference with the average score of the boys is not statistically significant.

Exhibit 10. Average Oral Reading Fluency Scores with Standard Deviation by Gender for Each Instrument

	Boys	Girls	Total	p-value
EGRA 2018 (n=461)	24.9 (13.9)	28.4 (12.4)	26.6 (13.3)	0.004
LARS (n=461)	31.2 (17.7)	36.6 (16.2)	33.9 (17.1)	0.001
LEGRA (n=448)	44.0 (25.1)	47.6 (23.1)	45.8 (24.2)	0.118

Finally, we examined the intraclass correlation (ICC) of student performance to assess the proportion of differences related to schools as shown in Exhibit 11 below. When the ICC is higher, this means that there is more difference between schools than between students within schools. When comparing the three ICC, we observed that this value is larger for LEGRA than for EGRA and LARS. This result suggests that the performance of the students on the LEGRA are more similar within each school and that there is more difference between the average performance of the schools. This difference could reflect, in part, the effect of the data collection mode used for the LEGRA (teacher administered).

Exhibit 11. Intraclass Correlation (% of Variance Related to School)

	ICC
EGRA 2018	0.15
LARS	0.15
LEGRA	0.45

RELATIONSHIPS BETWEEN PERFORMANCE ON THE DIFFERENT INSTRUMENTS

When comparing performance of the students on the EGRA and LARS instruments, we found that mostly all students who get zero scores on the EGRA also have zero scores on the LARS instrument as shown in Exhibit 12 below. The correlation between the distributions of the zero scores for the two instruments was computed using Cramer's V correlation; this value represents the degree to which students present the same value (non zero or zero scores) on the two instruments. The value of the Cramer's V between EGRA

and LARS is 0.76 (p<0.000), which suggest a very good level of association between the two instruments. This high level of association is also found between the performance on the fluency scores on the EGRA and LARS. As suggested by Exhibit 13 below, there is a strong association between the two instruments; the Pearson correlation between those two measures is 0.92 (p<0.000), which represents a very high value of correlation.

Exhibit 12. Distribution of Students with Zero Scores for EGRA 2018 and LARS

EGRA 2018	LARS		
	Non Zero	Zero score	Total
Non Zero	434 (99.3%)	3 (0.7%)	437 (100%)
Zero scores	7 (29.2%)	17 (70.8%)	24 (100%)
Total	441 (95.7%)	20 (4.3%)	461 (100%)

Exhibit 13. Scatterplot of the Performance of the Students on EGRA 2018 and LARS
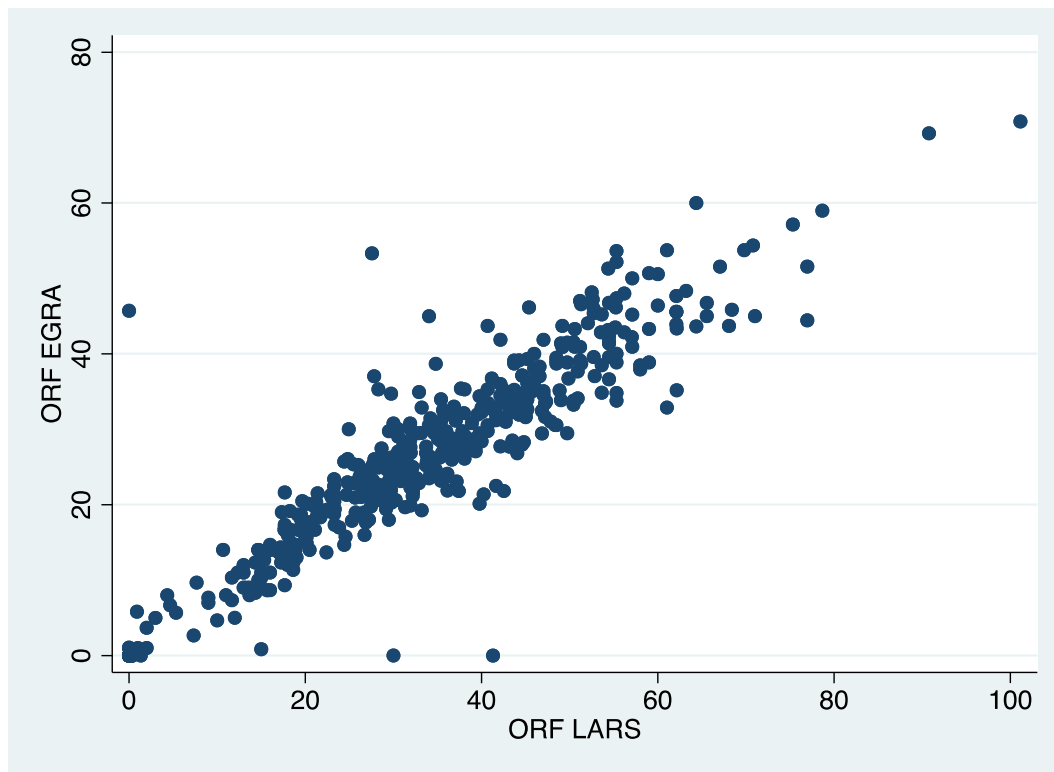
For the comparison between EGRA and LEGRA, we also found that most students who present zero scores on one instrument also present zero scores on the other instrument although the value of Cramer's V is lower with a value of 0.63. This suggests that the association is not as great between EGRA and LEGRA. The correlation between the performance of the students on both EGRA and LEGRA is 0.52, which is much lower thanthat when we compared performance on the EGRA and LARS. Exhibit 15 shows the lesser correspondence between the performance on EGRA and LEGRA.

Exhibit 14. Distribution of Students with Zero Scores for EGRA and LEGRA

EGRA 2018	LEGRA		Total
	Non Zero	Zero score	
Non Zero	415 (97.9%)	9 (2.1%)	424 (100%)
Zero scores	8 (33.3%)	16 (66.7%)	24 (100%)
Total	423 (94.4%)	25 (5.6%)	448 (100%)

Exhibit 15. Scatterplot of the Performance of the Students on EGRA 2018 and LEGRA 2021
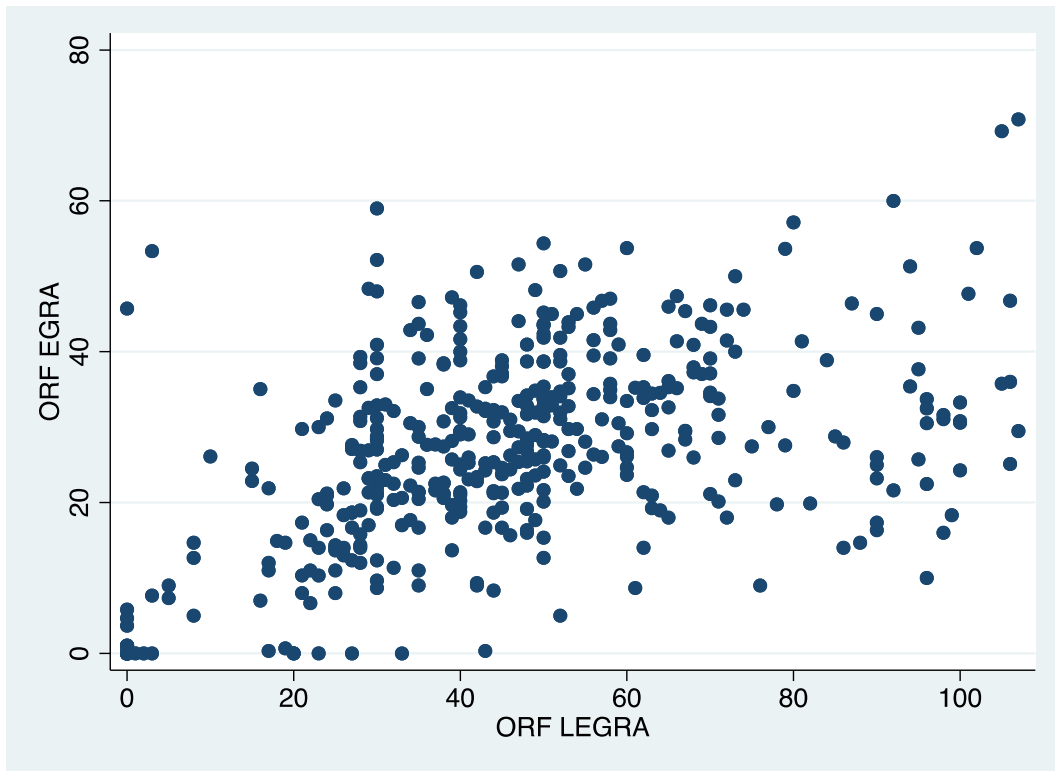
The association between zero scores on the LARS IV and LEGRA is very good. With the Cramer's V being 0.75 (p<0.000), that means that most of the students who got zero scores on one instrument also presented zero scores on the other instrument. However, the Pearson correlation between the performance on the oral reading fluency scores on both instruments is the same as what we observed between EGRA and LEGRA, with a correlation of 0.52 (p<0.000).

Exhibit 16. Distribution of Students with Zero Scores for LARS and LEGRA

LARS	LEGRA		Total
	Non Zero	Zero scores	Total
Non Zero	420 (98.1%)	8 (1.9%)	428 (100%)
Zero scores	3 (15.0%)	17 (85.0%)	20 (100%)
Total	423 (94.4%)	25 (5.6%)	448 (100%)

Exhibit 17. Scatterplot of the Performance of the Students on LEGRA and LARS
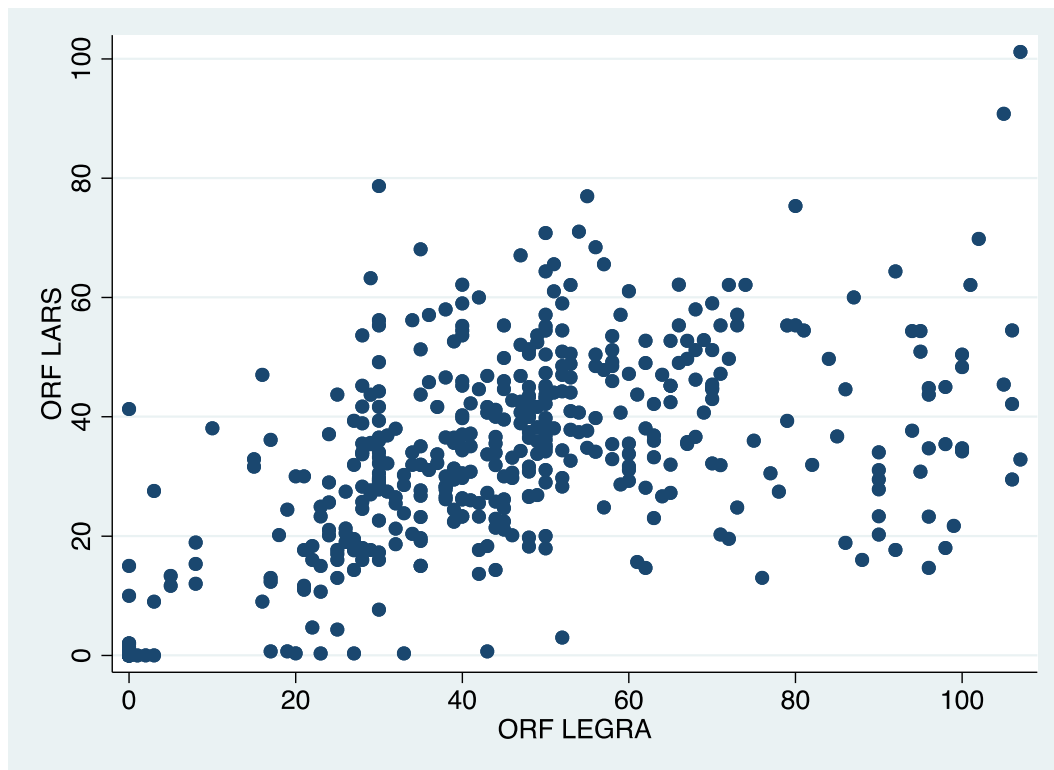
READING COMPREHENSION

Regarding the reading comprehension subtask, there is a bit more difference between the percentage of students with zero scores. The LEGRA showed the lowest percentage, with 5.7% of students having a score of zero, while for EGRA, 9.8% of the students have zero scores.

Exhibit 18. Percentage of Students with Zero Scores on Reading Comprehension

	EGRA 2018 (n=461)	LARS (n=461)	LEGRA (n=460)
Zero score	9.8%	8.5%	5.7%

The differences in average performance are smaller than what we observed for the oral reading fluency subtask as seen in Exhibit 19 below. LEGRA scores suggest that it is the easiest instrument for students, with students able to answer correctly an average of 78.4% of the reading comprehension questions. Meanwhile, EGRA appears to be the most difficult instrument as students provide on average 67.5% correct answers. The value of standard deviations are not as divergent, suggesting more similarities between performance on the different instruments. The format for the reading comprehension subtask, with only 5 comprehension questions, could also in part explain the observed similarities.

Exhibit 19. Average Reading Comprehension Scores for Each Instrument

	N	Mean	Standard deviation
EGRA 2018	461	67.5	32.8
LARS	461	75.7	29.7
LEGRA	460	78.4	27.4

As seen below in Exhibits 20 and 21, performance of boys and girls on all three instruments shows no significant differences regarding zero scores or average correct scores for all instruments. Globally, girls present lower proportions of zero scores and higher average scores, but those differences are small and are not statistically significant.

Exhibit 20. Percentage of Students with Zero Scores by Gender for Each Instrument

	Boys	Girls	Total	p-value
EGRA 2018 (n=461)	10.9%	8.6%	9.8%	0.406
LARS (n=461)	10.0%	6.9%	8.5%	0.225
LEGRA (n=460)	7.5%	3.9%	5.7%	0.097

Exhibit 21. Average Oral Reading Fluency Scores with Standard Deviation by Gender for Each Instrument

	Boys	Girls	Total	p-value
EGRA 2018 (n=461)	66.7	68.4	67.5	0.593
LARS (n=461)	74.3	77.1	75.5	0.321
LEGRA (n=460)	76.8	80.1	78.4	0.193

As seen in Exhibit 22 below, when looking at the different values of the ICC, the ICC for the LEGRA is lower than what was observed for the oral reading fluency scores and is more similar to the other two instruments. This result suggests a less important clustering effect of the school on the performance of the students for reading comprehension than for oral reading fluency.

Exhibit 22. Intraclass Correlation for the Reading Comprehension Subtask

	ICC
EGRA 2018	0.13
LARS	0.23
LEGRA	0.13

RELATIONSHIPS BETWEEN PERFORMANCE ON THE DIFFERENT INSTRUMENTS

As seen in the tables below, a very notable level of association is observed between the performance of the students on the EGRA and LARS reading comprehension subtasks. Cramer's V between percentages of students with zero scores is 0.87 (p<0.000). This results suggests that mostly the same students get a score of zero on both assessments. The Pearson correlation between scores on both subtasks assessments is 0.70 (p<0.000), suggesting a notable level of association between reading comprehension scores on both instruments.

Exhibit 23. Distribution of Students with Zero Scores for EGRA and LARS

EGRA 2018	LARS		
	Non Zero	Zero scores	Total
Non Zero	414 (99.5%)	2 (0.5%)	416 (100%)
Zero scores	8 (17.8%)	37 (82.2%)	45 (100%)
Total	422 (91.5%)	39 (8.5%)	461 (100%)

The observed relationships for zero scores and reading comprehension scores are a bit lower when considering EGRA and LEGRA performances. Cramer's V is a little bit lower with a value of 0.74 (p<0.000) and Pearson correlation between scores is 0.60 (p<0.000).

Exhibit 24. Distribution of Students with Zero Scores for EGRA and LEGRA

EGRA 2018	LEGRA		
	Non Zero	Zero scores	Total
Non Zero	415 (100%)	0 (0%)	415 (100%)
Zero scores	19 (42.2%)	26 (57.8%)	45 (100%)
Total	434 (94.3%)	26 (5.7%)	460 (100%)

Finally, LARS and LEGRA present similar results to the preceding relationships, the Cramer's V being 0.75 (p<0.000) and Pearson correlation 0.63 (p<0.000).

Exhibit 25. Distribution of Students with Zero Scores for LARS and LEGRA

| LARS	LEGRA		Total
	Non Zero	Zero scores	
Non Zero	420 (99.5%)	2 (0.5%)	422 (100%)
Zero scores	14 (36.8%)	24 (63.2%)	38 (100%)
Total	434 (94.3%)	26 (5.7%)	460 (100%)

EQUATING ANALYSIS BETWEEN INSTRUMENTS

Equating for this study is based on the equipercentile approach. This decision is based on the observed difference between the distribution of the scores on the three instruments. This section presents first the cumulative distribution functions for the three instruments together and then descriptive statistics and cumulative distribution functions for each original and equated instrument. This allows us to present how correct the equating method is in transforming scores on one instrument to the scale of values of a second instrument.

ORAL READING FLUENCY

Exhibit 26 shows the cumulative distribution function for each of the three instruments. Those curves present the relation between the percentile score (Y-axis) and the observed raw oral reading fluency scores. It shows that at similar values of percentile, the performance of the students on the three instruments are rather different. For example, students at the 20th percentile read about 18 correct words per minute for the EGRA, 20 correct words for the LARS, and 28 correct words for the LEGRA. Students at the 80th percentile read about 30 words for the EGRA, 40 words for LARS, and 60 words for LEGRA. This shows that the difference between passages changes, depending on the percentile of the students.

Exhibit 26. Cumulative Distribution Function for Oral Reading Fluency on Each Instrument
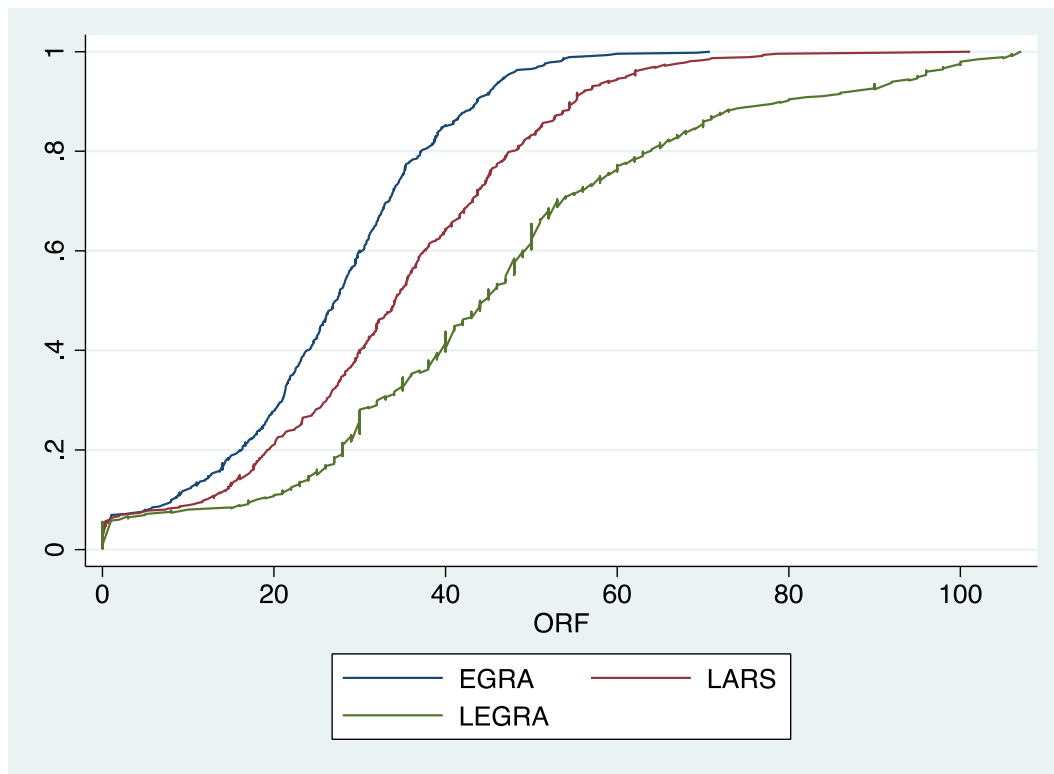
The first equipercentile is conducted between LARS and LEGRA. Descriptive statistics presented in Exhibit 27 show that the average scores and standard deviations of the equated LARS scores are very near to the values observed for EGRA 2018. As shown in Exhibit 28, the cumulative distributions function of the equated LARS scores is very similar to the curve for the EGRA 2018 scores.

Exhibit 27. Descriptive Statistics of EGRA 2018, LARS and Equated LARS Scores

	N	Mean	SD
EGRA 2018	461	26.6	13.3
LARS	461	33.9	17.1
Equated LARS	461	26.8	13.2

Exhibit 28. Cumulative Distribution Function for EGRA 2018, LARS and Equated LARS Scores
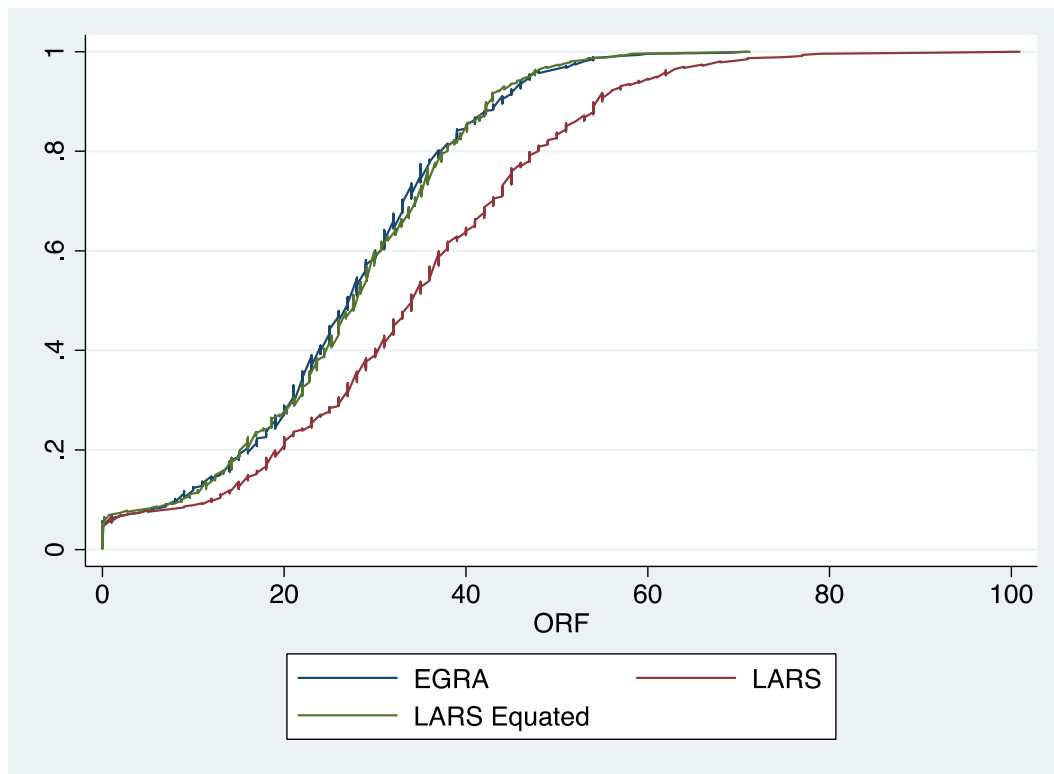
Descriptive statistics for the equated LEGRA scores on the EGRA 2018 scale of values for oral reading fluency scores are presented in Exhibit 29. We observed that the descriptive statistics of the equated LEGRA scores are similar and the cumulative distribution function presents a similar curve to the EGRA 2018 scores.

Exhibit 29. Descriptive Statistics of EGRA 2018, LEGRA and Equated LEGRA Scores

	N	Mean	SD
EGRA 2018	461	26.6	13.3
LEGRA	448	45.8	24.2
Equated LEGRA	461	26.9	13.4

Exhibit 30. Cumulative Distribution Function for EGRA 2018, LEGRA and Equated LEGRA Scores
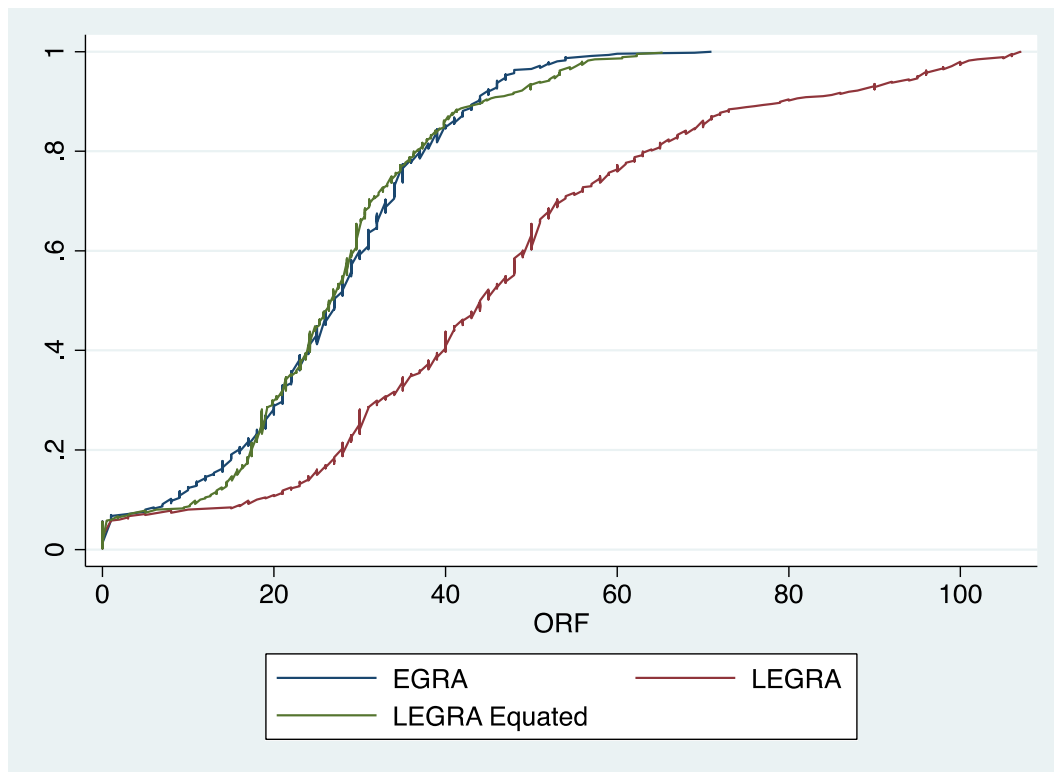
Finally, for the equating of the LEGRA scores on the scale of values of the LARS scores, we found similar average scores and standard deviations as for the LARS scale. The cumulative distribution function of the equated LEGRA scores is also similar to the LARS curve.

Exhibit 31. Descriptive statistics of LARS, LEGRA and Equated LEGRA scores

	N	Mean	SD
LARS	461	33.9	17.1
LEGRA	448	45.8	24.2
Equated LEGRA	461	34.1	17.4

Exhibit 32. Cumulative Distribution Function for LARS, LEGRA and Equated LEGRA Scores
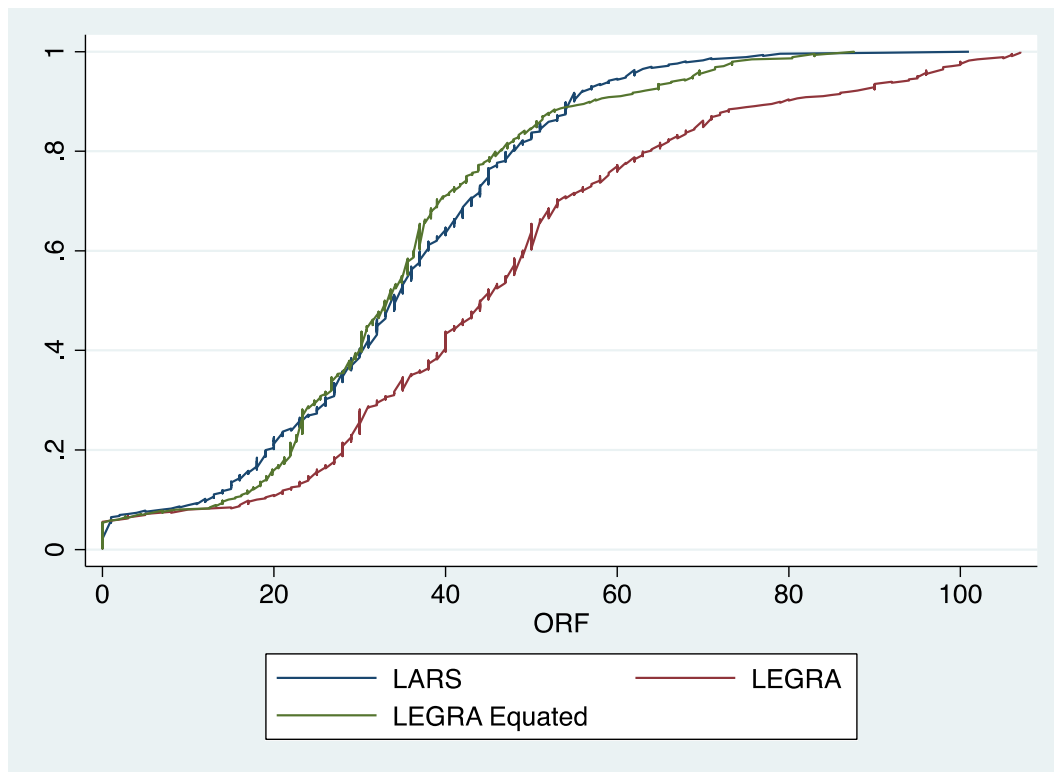
READING COMPREHENSION

For reading comprehension, the cumulative distribution curves are a bit harder to interpret because there are less possible scores on each instrument. While the oral reading fluency is a continuous scale, the reading comprehension is a discrete variable with possible scores of 0%, 20%, 40%, 60%, 80%, and 100%. The interpretation ends up being similar as we observe in Exhibit 33 that students at the 20th percentile present a score of about 40% for the EGRA 2018 instrument, 50% for LARS, and 60% for LEGRA.

Exhibit 33. Cumulative Distribution Function for Reading Comprehension on Each Instrument
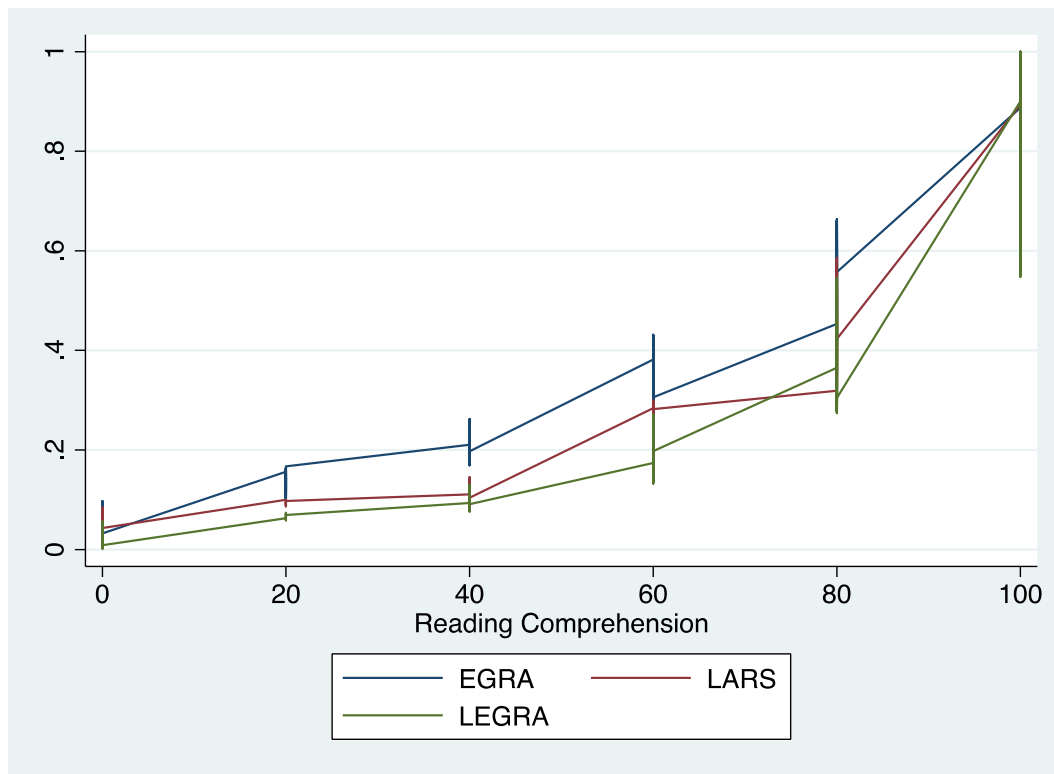
Equating was perfomed to transform scores on the LARS scale of values to the EGRA scale. Descriptive statistics of the equated LARS scores suggest that the results are similar with the EGRA 2018 scores. The cumulative distribution function graphic shows the difference between the curve of the equated LARS score and EGRA 2018 scores. Those are expected given the discrete nature of the reading comprehension score variable.

Exhibit 34. Descriptive Statistics of EGRA, LARS and Equated LARS Scores

	N	Mean	SD
EGRA 2018	461	67.5	32.8
LARS	461	75.7	29.7
Equated LARS	461	68.3	31.6

Exhibit 35. Cumulative Distribution Function for EGRA 2018, LARS and Equated LARS Scores
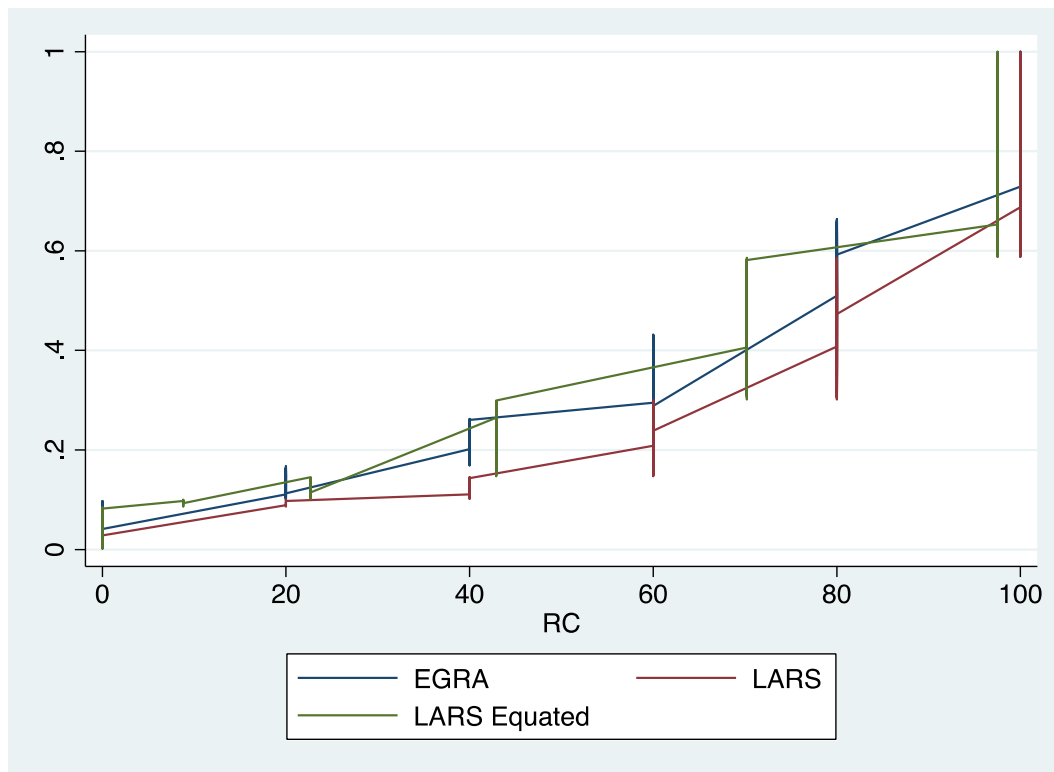
Similar results are observed when looking at the results of the equating of the LEGRA reading comprehension scores and the EGRA 2018 scores. The same kind of differences are also observed between the curves for the cumulative distribution functions.

Exhibit 36. Descriptive Statistics of EGRA 2018, LEGRA and Equated LEGRA Scores

	N	Mean	SD
EGRA 2018	461	67.5	32.8
LEGRA	460	78.4	27.4
Equated LEGRA	460	68.7	31.1

Exhibit 37. Cumulative Distribution Function for EGRA 2018, LEGRA and Equated LEGRA Scores
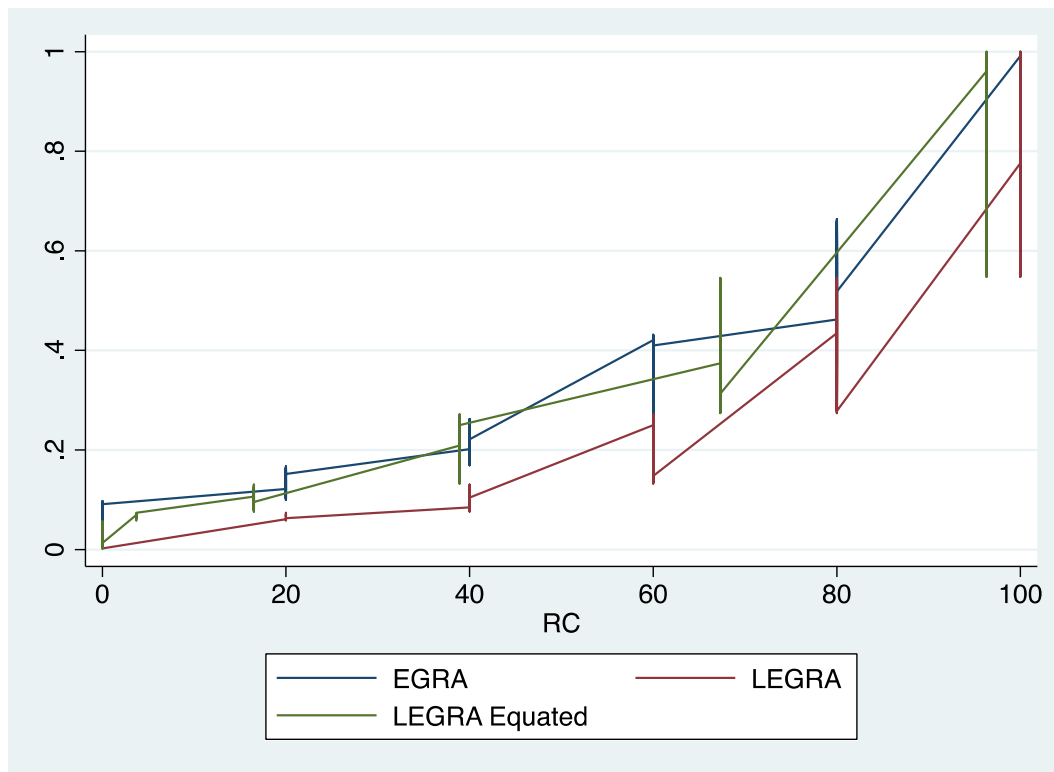
Finally, when running equating with LEGRA comprehension scores on the scale of values of the LARS scores, we also observed similar descriptive statistics between equated LEGRA scores and LARS scores, but some differences in the cumulative distribution function curves.

Exhibit 38. Descriptive Statistics of LARS, LEGRA and Equated LEGRA Scores

	N	Mean	SD
LARS	461	75.7	29.7
LEGRA	460	78.4	27.4
LARS on LEGRA	461	73.8	30.2

Exhibit 39. Cumulative Distribution Function for LARS, LEGRA and Equated LEGRA Scores
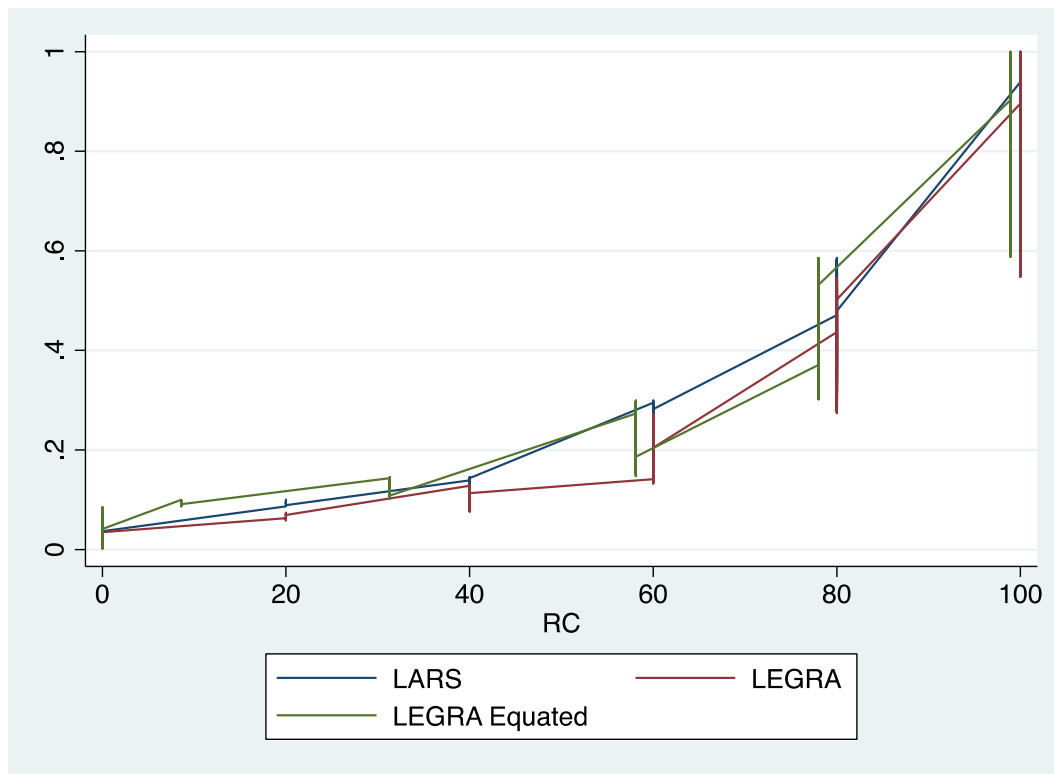
TABLES TO TRANSFORM SCORES FROM ONE INSTRUMENT TO THE OTHER INSTRUMENTS

Overall, results obtained for all equated scores suggest that the equipercentile equating approach used for this study performed well. Therefore, equated scores could be used to compare performance of future students who will take the LARS or the LEGRA with performance on the EGRA or to compare performance of students who take the LEGRA with what their performance could be on the LARS. The following tables present the correspondence between the LARS or LEGRA scores and their equated values when equating was perform with EGRA or LARS.

FLUENCY SCORES

Exhibit 40. LARS to EGRA Scale Scores

Observed ORF scores on LARS	Equated ORF scores on LARS to EGRA scale scores
0	0
1	0.2
2	0.7
3	1.3
4	2.0
5	2.7
6	3.5
7	4.3
8	5.1

Observed ORF scores on LARS	Equated ORF scores on LARS to EGRA scale scores
9	6.0
10	6.8
11	7.7
12	8.7
13	9.6
14	10.5
15	11.4
16	12.4
17	13.3
18	14.2
19	15.1
20	16.0
21	16.9
22	17.7
23	18.6
24	19.5
25	20.3
26	21.1
27	22.0
28	22.8
29	23.6
30	24.4
31	25.2
32	26.0
33	26.8
34	27.6
35	28.4
36	29.1
37	29.9
38	30.7
39	31.4
40	32.2
41	32.9
42	33.7
43	34.4
44	35.1
45	35.8
46	36.6
47	37.3
48	38.0
49	38.7
50	39.4
51	40.1
52	40.8
53	41.5
54	42.2
55	42.9
56	43.6
57	44.2
58	44.9

Observed ORF scores on LARS	Equated ORF scores on LARS to EGRA scale scores
59	45.6
60	46.2
61	46.9
62	47.6
63	48.2
64	48.9
65	49.5
66	50.2
67	50.8
68	51.5
69	52.1
70	52.7
71	53.4
72	54.0
73	54.6
74	55.3
75	55.9
76	56.5
77	57.1
78	57.7
79	58.3
80	59.0
81	59.6
82	60.2
83	60.8
84	61.4
85	62.0
86	62.5
87	63.1
88	63.7
89	64.3
90	64.9
91	65.5
92	66.1
93	66.6
94	67.2
95	67.8
96	68.4
97	68.9
98	69.5
99	70.1
100	70.7
101	71.3

Observed ORF scores on LARS	Equated ORF scores on LARS to EGRA scale scores

Exhibit 41. LEGRA to EGRA Scale Scores

Observed ORF scores on LEGRA	Equated ORF scores on LEGRA to EGRA scale scores
0	0.0
1	0.5
2	1.1
3	1.8
4	2.4
5	3.1
6	3.7
7	4.4
8	5.0
9	5.7
10	6.3
11	7.0
12	7.6
13	8.2
14	8.9
15	9.5
16	10.2
17	10.8
18	11.4
19	12.0
20	12.6
21	13.3
22	13.9
23	14.5
24	15.1
25	15.7
26	16.3
27	16.9
28	17.4
29	18.0
30	18.6
31	19.2
32	19.8
33	20.3
34	20.9
35	21.4
36	22.0
37	22.6
38	23.1
39	23.7
40	24.2
41	24.8
42	25.3
43	25.8
44	26.4
45	26.9
46	27.5
47	28.0

Observed ORF scores on LEGRA	Equated ORF scores on LEGRA to EGRA scale scores
48	28.5
49	29.0
50	29.6
51	30.1
52	30.6
53	31.1
54	31.7
55	32.2
56	32.7
57	33.2
58	33.7
59	34.2
60	34.7
61	35.3
62	35.8
63	36.3
64	36.8
65	37.3
66	37.8
67	38.3
68	38.8
69	39.3
70	39.8
71	40.3
72	40.8
73	41.3
74	41.8
75	42.3
76	42.8
77	43.3
78	43.8
79	44.3
80	44.8
81	45.3
82	45.8
83	46.3
84	46.8
85	47.3
86	47.9
87	48.4
88	48.9
89	49.4
90	49.9
91	50.5
92	51.0
93	51.6
94	52.1
95	52.7
96	53.3
97	53.9

Observed ORF scores on LEGRA	Equated ORF scores on LEGRA to EGRA scale scores
98	54.5
99	55.2
100	55.9
101	56.6
102	57.4
103	58.3
104	59.3
105	60.6
106	62.3
107	65.2

Exhibit 42. LEGRA to LARS Scale Scores

Observed ORF scores on LEGRA	Equated ORF scores on LEGRA to LARS scale scores
0	0.0
1	0.9
2	1.8
3	2.7
4	3.6
5	4.5
6	5.3
7	6.2
8	7.0
9	7.8
10	8.6
11	9.4
12	10.2
13	11.0
14	11.7
15	12.5
16	13.2
17	14.0
18	14.7
19	15.5
20	16.2
21	16.9
22	17.6
23	18.4
24	19.1
25	19.8
26	20.5
27	21.2
28	21.9
29	22.6
30	23.3
31	24.0
32	24.7
33	25.4

Observed ORF scores on LEGRA	Equated ORF scores on LEGRA to LARS scale scores
34	26.0
35	26.7
36	27.4
37	28.1
38	28.8
39	29.5
40	30.2
41	30.8
42	31.5
43	32.2
44	32.9
45	33.6
46	34.2
47	34.9
48	35.6
49	36.3
50	37.0
51	37.6
52	38.3
53	39.0
54	39.7
55	40.4
56	41.0
57	41.7
58	42.4
59	43.1
60	43.8
61	44.4
62	45.1
63	45.8
64	46.5
65	47.2
66	47.9
67	48.5
68	49.2
69	49.9
70	50.6
71	51.3
72	52.0
73	52.7
74	53.4
75	54.0
76	54.7
77	55.4
78	56.1
79	56.8
80	57.5
81	58.2
82	59.0
83	59.7

Observed ORF scores on LEGRA	Equated ORF scores on LEGRA to LARS scale scores
84	60.4
85	61.1
86	61.8
87	62.6
88	63.3
89	64.0
90	64.8
91	65.6
92	66.3
93	67.1
94	67.9
95	68.8
96	69.6
97	70.5
98	71.4
99	72.4
100	73.4
101	74.5
102	75.7
103	77.0
104	78.5
105	80.4
106	83.0
107	87.6

READING COMPREHENSION

Exhibit 43. LARS to EGRA Scale Scores

Observed reading comprehension scores on LARS	Equated reading comprehension scores on LARS to EGRA scale scores
0%	0%
20%	8.8%
40%	22.7%
60%	42.9%
80%	70.2%
100%	97.5%

Exhibit 44. LEGRA to EGRA Scale Scores

Observed reading comprehension scores on LEGRA	Equated reading comprehension scores on LEGRA to EGRA scale scores
0%	0%
20%	3.7%
40%	16.5%
60%	38.9%
80%	67.3%
100%	96.3%

Exhibit 45. LEGRA to LARS Scale Scores

Observed reading comprehension scores on LEGRA	Equated reading comprehension scores on LEGRA to LARS scale scores
0%	0%
20%	8.6%
40%	31.3%
60%	58.1%
80%	78.0%
100%	98.9%

CONCLUSION

Thus, the main objective of this study was to support Soma Umenye to equate the (1) LEGRA assessment (2) the 2018 EGRA, and the (3) LARS IV assessment. As mentioned above, those tools were already used in different contexts and at different times. The LEGRA tool was used at the end of Term 1 and Term 2 to help teachers understand whether students were making progress towards end-of-term performance benchmarks. The LARS assessment was used to measure student performance every two years. The EGRA was used in 2018, at the end of the school year, to set the baseline for Soma Umenye. Equating was performed on the oral reading fluency and reading comprehension subtasks.

The equating study results presented in this report give the Rwandan education system an opportunity to review its comprehensive assessment system and determine the extent to which its formative assessments, implemented through LEGRA for early grade reading in Kinyarwanda, produce data that is confirmed by summative assessments (LARS, EGRA etc.). By ensuring the LEGRA, LARS, and EGRA assessments are aligned, the system players will be able to use them to understand trends, challenges, and take necessary decisions to improve learner outcomes. Results from this study could also be used to look at performance of the students on the 2021 assessment compared to the performance of the students on the EGRA used at baseline in 2018.

With the equating of the LARS, EGRA, and LEGRA, Soma Umenye can now compare its baseline (conducted with the EGRA) with an endline measure from either the LARS or the LEGRA.

During the five-year Soma Umenye project, since the joint development of the EGRA baseline instruments in 2018 for evaluating Soma Umenye, Rwanda has conducted five early grade Kinyarwanda reading assessments. An EGRA study was administered in 2018 with a sample of P1 to P3 students using jointly developed instruments. This EGRA was planned to be a baseline measure for the performance assessment of Soma Umenye. LARS was nationally administered in 2021 with a sample of P4 students who were tested on P3 items. The LARS assessment includes oral reading fluency and reading comprehension subtasks. Finally, LEGRA was developed by Soma Umenye as a literacy assessment administered by teachers in their classrooms for the purpose of formative assessment.

This design necessitated that we administer the three instruments to be equated to the same group of students. Data were collected from a total of 461 P3 students, with almost 50% of those students being girls and a bit more than 50%, boys. A team of enumerators collected data for the EGRA and LARS subtasks while the LEGRA was administered by teachers to respect the mode of administration of this instrument. Descriptive analysis was first conducted on the collected data before running equating analysis that produced tables to transform scores on one instrument to the scale of values of the other instrument.

MAIN RESULTS FROM THE EQUATING

Descriptive analysis showed that the oral reading fluency subtask was easier on the LEGRA than for the two other passages. Since LEGRA differs in both the content and mode of administration, it is not possible at this juncture to definitively conclude why results are higher for the students on the LEGRA. The mode of administration can have an impact on the performance of the students, with students being more comfortable with reading a passage to their teacher instead of to an unfamiliar and external data collector. Teachers could also be more lenient in the administration of the LEGRA than a trained enumerator. It

is also possible that content differences existed and that the passage developed for the LEGRA was simply easier. The evaluation team did not have sufficient information to engage in a review of the content of the LEGRA. Unfortunately, the design of the study does not permit a clear answer to the question of why the results of the three assessments varied, but since this is an equating study, differences in difficulty of the assessment were controlled when using tables to equate scores from LEGRA to EGRA or LARS. Equating helps to control differences in difficulty of the instruments no matter what explains this difference. It should be noted that the differences in difficulty of the passages were smaller when looking at the performance on reading comprehension.

Very few differences were observed between the performance of girls and boys. The only significant differences were observed for oral reading fluency for the LEGRA and LARS instruments. For both of those instruments, the girls outperformed the boys. No significant differences were found on reading comprehension.

It is interesting to note that the association between the oral reading fluency performance was higher for EGRA and LARS than for LEGRA, while the associations are similar between all instruments for the reading comprehension. These results also suggest that the mode of administration and passage content interact differently with the students for the oral reading fluency assessment in the LEGRA. It could be interesting to investigate why such differences are observed.

The equipercentile method used to put all scores on the same scale of values seems to have worked well considering that the descriptive statistics of the equated scores are similar to the descriptive statistics of the instrument for which we want the scores to be equated. Also, the similarity in the cumulative distribution functions for the equated scores suggest the effectiveness of the equating. Given those results, we are assured in saying that new data collected on the LARS or LEGRA can be put on the same scale of values and directly compared (and also can be directly compared with performance of the P3 students on EGRA in 2018). However, since the equating method that we use for this study based the transformation of the scores on the observed scores in the data collected, it is only possible to produce transformation tables for different instruments for the observed scores in the present study. This constitutes the main limitation with the chosen approach and should affect only a very limited number of students on any future study. Also, transformation is only possible with rounded scores while fluency scores sometimes are computed using decimals. The rounded scores can easily be obtained and should not have a big impact on the transformation. Furthermore, those limitations only apply to the oral reading fluency scores.

WHAT'S NEXT?

The present study allowed for conducting equating between EGRA 2018, LARS, and LEGRA (term 3) instruments, thus allowing for the transformation of future students' performance on the LEGRA on the scale of the LARS (or the LARS or LEGRA on the scale of the EGRA 2018 to allow direct comparison with performance of P3 students in 2018). The performed equating also enables the LEGRA scores of students to be transformed to the LARS scale. Equating takes into account differences in difficulty of the instruments to transform the scores, so equated scores can be directly compared and hence any differences will represent real differences.

We need to note that the equating study only reviewed the LEGRA term 3 instrument. There could be benefit from further review of the LEGRA term 1 and term 2 instruments to inform potential re-calibration of the term benchmarks to provide a better projection to teachers when they complete their end of term assessments. It is likely that the relative difficulty of the term 1 and term 2 LEGRA assessments may be different to that assessed for term 3.

By comparing the performance of the same students on EGRA 2018, LARS IV and LEGRA (term 3), we can conclude that there are differences in the difficulty of the instruments. The observed oral reading fluency scores of the LEGRA seem to be different than the EGRA 2018 and LARS. Nonetheless, equating allowed us to take the differences in performance into account and control for them. This is demonstrated by the descriptive statistics of the equated LEGRA scores that are similar to those of the EGRA and of the LARS.

With the transformation tables now available, it will be possible to conduct future comparisons of the performance of students who were assessed with the different instruments without having to administer all these instruments to the same students. LEGRA results can be collected and directly compared to LARS, allowing an additional datapoint to monitor the performance of the students. In addition, the 2018 EGRA study based on a nationally representative sample of P3 students can be compared with future LEGRA or LARS results to measure the development of the competency of Rwandan students in literacy.

We would recommend a future calibration exercise is carried out when LARS V instruments are being prepared, so that LARS V / LEGRA 2022 instruments are aligned, allowing results to be compared.