Measuring Foundational Learning: Insights from ASER and EGRA/EGMA Comparisons in Botswana

November 2026

Learning gains of programs that teach foundational numeracy and literacy are often measured through different types of tests. Two of the most prominent tests are ASER and EGRA/EGMA. Recent studies have found that these tests can draw different conclusions on how much children learned for the same programs. We compared student learning in a Teaching at the Right Level (TaRL) program in Botswana across ASER, EGRA, and EGMA test items to understand how results differ across them. Pre-post learning gains measured in standard deviations are smaller on the EGRA and EGMA tests than on the ASER test. Yet, when we examine multiple alternative ways to relate assessments, the tests perform more similarly. We find stronger correlations across EGRA and ASER than for items within EGRA. When measured in equivalent years of schooling, the EGMA/EGRA-ASER gap decreases, in particular for numeracy. In a randomised A/B test of a literacy programme tweak, the estimated impact was similar on both tests in terms of standard-deviations but much larger on EGRA items when compared to control group gains. We conclude that standard deviations can be misleading, and that ASER and EGRA/EGMA are important complements.

Complementary tools to measure learning: comparing ASER and EGRA/EGMA

ASER-style assessments, developed for the Annual Status of Education Report, are short, one-on-one oral tests that check whether children can perform foundational literacy and numeracy tasks. They place a child on a ladder of skills: for literacy, from recognising letters to reading words, a paragraph and a story; for numeracy, from recognising numbers to addition, subtraction, multiplication and division. Developed to provide rapid snapshots of learning, they have been widely used across South Asia and sub-Saharan Africa to identify children's foundational skill levels and support remedial instruction.

EGRA and EGMA, developed by RTI International, are simple assessments also meant to provide rapid and reliable measures of learning. They are often more comprehensive than ASER, but can also be costlier and harder to run at scale. They run timed subtasks, such as oral reading fluency measured in words read correctly per minute, and produce more continuous scores that can be compared across systems (Dubeck & Gove, 2015; Platas et al., 2016). They are widely used to evaluate education interventions and inform policy.

Results on the two assessments can vary. A TaRL evaluation in Côte d'Ivoire found gains of 0.16 SD in literacy and 0.28 SD in numeracy on ASER-style measures, but smaller and statistically insignificant gains on EGRA and EGMA (0.10 and 0.08 SD) (Wolf et al., 2026). These gaps raise a practical question: whether these differences reflect real differences in what children learned, or differences in what each tool measures and how its scores are scaled.

Our study: assessing the same children with both the ASER and EGRA/EGMA tools

We studied how ASER and EGRA/EGMA compare in the context of a scaling TaRL program in Botswana in 2025. Children in standards 3 to 5 across 22 schools were assessed at baseline and again at endline, after one TaRL implementation cycle. In total, 1,327 children completed the numeracy assessments and 1,294 completed the literacy assessments. To enable direct comparisons, assessments designed to capture similar competencies were administered to the exact same students on the same day.

We built a combined assessment on top of the standard ASER test (Table 1). Children first completed the ASER levelling items as usual. We then added a set of items inspired by EGRA and EGMA: a timed addition fluency task and untimed word problems for numeracy, and a timed grade-level reading passage with comprehension questions for literacy. This design keeps the low cost of ASER while capturing more continuous, internationally recognisable measures from EGRA and EGMA. Examples of the tools are provided in the Appendix.

To assess alignment between tools, we compare the magnitude of the learning gains of the TaRL program across the assessments using different methods, and calculate Spearman's rank correlations (ρ), Cronbach's alpha, and elasticities between the assessments. We also calculate equivalent years of schooling gained, and compare impacts in an A/B test of a tweak in the literacy programme.

Table 1: The combined ASER, EGRA and EGMA test items

Step	Numeracy items	Literacy items	Source
7	Word problems: 4 items, all four operations, untimed	Comprehension: 2 questions on a passage, untimed	Inspired by EGMA/EGRA
6	Addition fluency: 20 sums (1-by-1 to 2-by-2 digit) in 60 seconds	Oral reading fluency: grade-level passage, read for 60 seconds	
5	Division: 2-by-1 digit	Comprehension: 2 questions on the story	Standard ASER levelling items
4	Multiplication: 2-by-1 digit	Story: reads with fewer than 3 mistakes	
3	Subtraction: 2-by-2 digit	Paragraph: reads with fewer than 3 mistakes	
2	Addition: 2-by-2 digit	Word: reads 4 of 5 words	
1	Number recognition: 1 to 4 digits	Letter: recognises 4 of 5 letters	

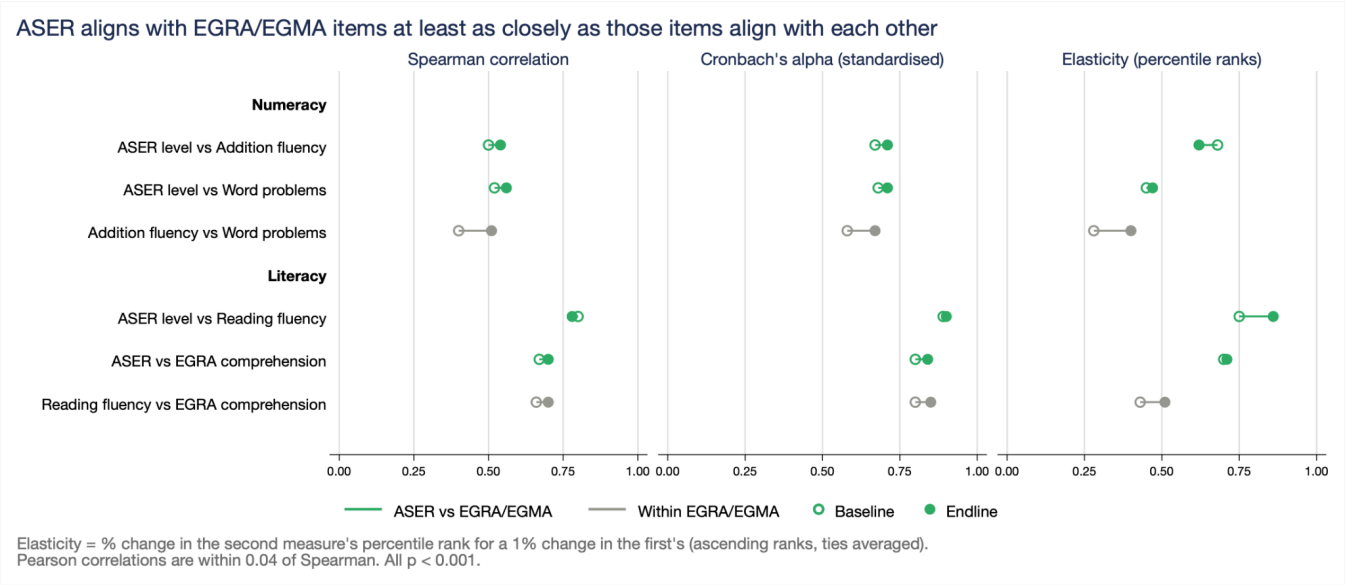
Note: ASER items are administered in sequence; children advance to the next level only after passing the one below.

Both tools rank children in a similar way

We find strong correlations between the ASER and EGMA and EGRA results. In literacy, ASER reading levels were strongly correlated with oral reading fluency: the coefficients are 0.80 at baseline and 0.78 at endline. In numeracy, ASER levels were moderately correlated with addition fluency and word problems: 0.50–0.52 at baseline and 0.54-0.56 at endline.

A telling comparison is within the EGRA/EGMA items themselves. In numeracy, ASER levels are more closely related to each EGMA-style item than the two EGMA-style items are to each other. In literacy,

ASER levels track oral reading fluency more closely than fluency tracks comprehension within the EGRA-style items. In other words, ASER is not a noisier stand-in for EGRA/EGMA: it lines up with those items at least as well as they line up with one another. We make comparisons using correlations, cronbach alpha statistics, and elasticities.



Assessment results scaling matters for interpreting the differences in gains

Every tool shows substantial learning over the TaRL cycle. On ASER, students moved up 1.42 levels on average in numeracy and 0.40 levels in literacy. On the EGRA/EGMA-style items, children answered 1.6 more addition sums per minute, solved 0.66 more word problems, and read 7.2 more words correctly per minute.

To compare across tools, gains can be converted into standard deviations (SD) of the baseline distribution. In literacy, the gain was 0.34 SD on ASER and 0.22 SD on oral reading fluency. In numeracy, the gain was 1.08 SD on ASER, compared with 0.54 SD on addition fluency and 0.60 SD on word problems. These smaller standard deviations gains in EGRA/EGMA than ASER could be a function of either (a) smaller numerators or (b) larger denominators. We analyze these dynamics.

A gain in SD units depends on two things: how much children improved and how spread out scores were at baseline. The two tools differ in particular on the second. ASER sorts children into five broad levels, so a child who moves up a level registers a large jump relative to the spread. Timed fluency tasks produce finer-grained scores that have a more continuous distribution and more variation. That extra variation widens the baseline spread and shrinks the gain in SD terms. One way to check this is to express gains in a metric that does not depend on the baseline variance. We converted gains into equivalent years of schooling. We divide the level gains by the average difference in baseline scores between one grade and the next. On this metric, the numeracy gap reduces. Children gained 1.68 years of schooling on ASER, 1.71 years on word problems and 1.07 years on addition fluency. In literacy, we find gains of 1.48 years on ASER and 0.80 years on oral reading fluency.

Table 2: Level gains with different standardization methods

Measure	Baseline	Endline	Gain (SD)	Gain (equivalent years of schooling)
Numeracy				
ASER level (0-4)	1.62	3.03	1.08	1.68
EGMA addition fluency (sums per minute)	7.52	9.12	0.54	1.07
EGMA word problems (0-4)	1.16	1.83	0.60	1.71
Literacy				
ASER level (0-4)	2.92	3.32	0.34	1.48
EGRA oral reading fluency (words correct per minute)	47.55	54.74	0.22	0.80

We read this as evidence that scaling explains part of the gap, in particular for numeracy.

Moreover, each tool is fit-for-purpose. ASER literacy levels require crossing thresholds, such as reading a paragraph or a story accurately, which is exactly what TaRL interventions target. Reading speed tends to grow more gradually, and a child who has just learned to read a story may not yet read it quickly. This is consistent with evidence from Syria, where EGRA subtasks such as letter and sound recognition picked up large gains among emerging readers that ASER's items missed (Montes-de-Oca et al., 2022). This suggests that each tool is sensitive at a different stage of reading development.

In a randomised comparison, both tools point to a similar effect

Within the same study, we conducted an A/B test where schools were randomly assigned to two versions of a TaRL literacy programme: without a story toolkit (Group A) and with a story toolkit (Group B). This lets us compare how each tool measures a causal difference, not just a pre-post or correlational change. This randomised comparison is closer in design to the Côte d'Ivoire evaluation (Wolf et al., 2026). This tweak captures *additional* gains above and beyond the base program, rather than the full impact of the literacy program. The additional gains from the story toolkit on learning was 0.11 SD on ASER, and 0.12 SD on oral reading fluency (Table 3). Neither difference is statistically significant at conventional levels although p-values are on the margin of statistical significance (p=.138 and p =.154), so we treat this result as suggestive. Still, the two tools point to almost the same effect in terms of standard deviations, even though pre-post gains were about a third smaller on EGRA items.

In addition, we express impact relative to control group learning trajectories rather than in terms of standard deviations. We find that the same impact relative to the control group's gain shows causal learning gains of the story toolkit of 40% on ASER and nearly 80% on oral reading fluency EGRA items. These percentages illustrate how much the choice of scale matters. Whether ASER or EGRA "shows the bigger effect" depends as much on how gains are scaled as on the learning gains themselves.

Table 3: Impacts in a literacy A/B test with different standardization methods

	ASER literacy level		EGRA oral reading fluency (words per minute)	
	(1)	(2)	(3)	(4)
	Raw gain	SD gain	Raw gain	SD gain
Treatment (B vs A)	0.132	0.113	3.998	0.121
	(0.086)	(0.073)	(2.704)	(0.082)
	[0.138]	[0.138]	[0.154]	[0.154]
Control baseline mean	3.05	0	49.99	0
Control baseline SD	1.11		33.19	
Control mean gain	0.33	0.28	5.10	0.15
% more learning than control gains	39.9%		78.4%	
	(32.1%)		(88.0%)	
	[0.214]		[0.373]	
Controls for baseline values, gender and standard	Yes	Yes	Yes	Yes
Observations	1,274	1,274	1,274	1,274
Schools (clusters)	22	22	22	22

Note: Standard errors clustered at the school level in parentheses; p-values in brackets.
% more learning = (gain in B - gain in A) / gain in A. SD gains are divided by the baseline standard deviation. * p<0.10, ** p<0.05, *** p<0.01.

Implications for measurement and further research

ASER-style assessments remain a practical and pedagogically-aligned tool for tracking progress in the specific skills that Teaching at the Right Level (TaRL) is designed to teach: helping children move from one specific learning level to the next. Their strong correlation with EGRA and EGMA confirms that ASER captures similar underlying learning concepts while offering a fast and lower-cost option for ongoing program monitoring.

At the same time, EGRA and EGMA complement ASER by providing additional granularity and provide continuous distributions with wider variation. Placing EGRA/EGMA-style items on top of the standard ASER assessment keeps the assessment practical while producing measures with more variance and continuous distributions, as well as measures that funders and governments can compare with more international benchmarks. Our results show how each test is fit-for-purpose and how standard deviations can indeed be misleading (Rossiter et al. 2026), emphasizing the importance of reporting effects on multiple score scales.

Youth Impact gratefully acknowledges the support of the Gates Foundation, whose funding made this research possible. We also thank the participating schools, teachers, and learners in Botswana, the Ministry of Basic Education, as well as the Youth Impact team members who contributed to this work.

REFERENCES

Dubeck, M. M., & Gove, A. (2015). The early grade reading assessment (EGRA): Its theoretical foundation, purpose, and limitations. *International Journal of Educational Development, 40*, 315–322. https://doi.org/10.1016/j.ijedudev.2014.11.004

Montes de Oca, M., Hill, J., Aber, L., Tubbs Dolan, C., & Gjicali, K. (2022). *The impact of attending a remedial support program on Syrian children's reading skills: Using BART for causal inference* (arXiv:2208.13906). arXiv. https://doi.org/10.48550/arXiv.2208.13906

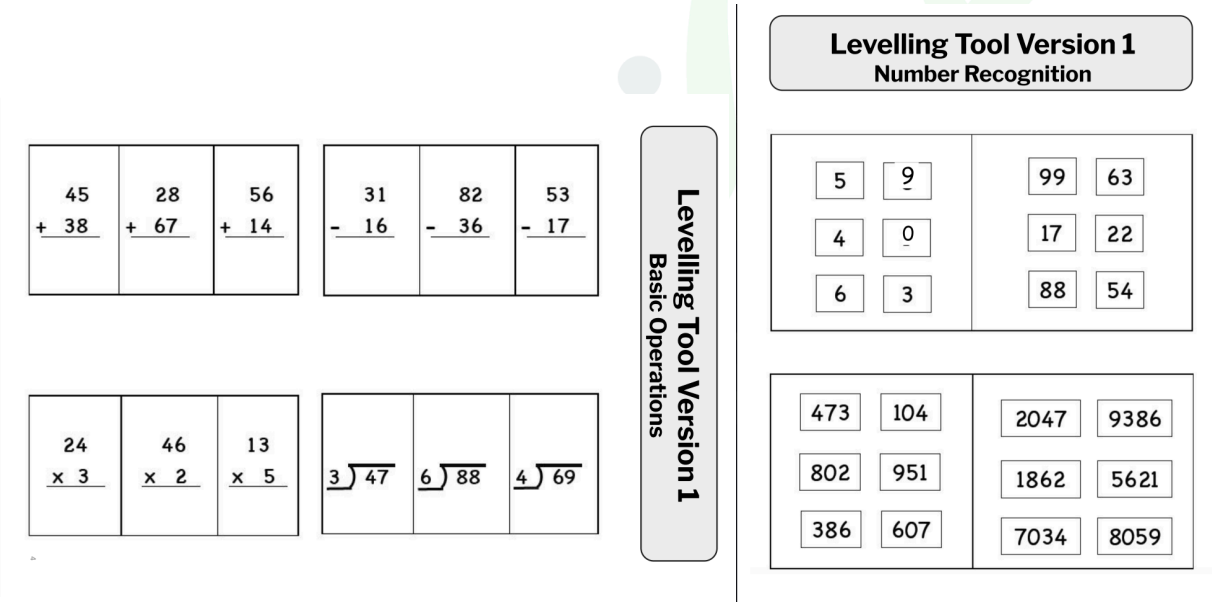
Platas, L. M., Ketterlin-Geller, L. R., & Sitabkhan, Y. (2016). Using an assessment of early mathematical knowledge and skills to inform policy and practice: Examples from the Early Grade Mathematics Assessment. *International Journal of Education in Mathematics, Science and Technology, 4*(3), 163–173.

Rossiter, J., Evans, D. K., Hares, S., & Henny, C. (2026). *The illusion of comparability among standardised effect sizes: Why education evaluations should report raw effects* (CGD Working Paper No. 748). Center for Global Development.

Wolf, S., Ogan, A., Jasińska, K. K., Kembou, S., Cannanure, V. K., Whitehead, H., Corbett, J., Tanoh, F., Wortsman, B., Lim, J., Zhao, H., & Kumar, A. (2026). Impacts of targeted instruction with mobile-based support on teachers and students in rural Côte d'Ivoire. *American Educational Research Journal, 63*(4). https://doi.org/10.3102/00028312261430679

APPENDIX

ASER TOOLS



LEVELLING TOOL V1

5

Letter

c	q	k
x	a	
t	u	j
b	p	

Word

boy	want
home	sun
cap	river
paint	make
tell	dig

Paragraph

I am not feeling well. My throat is sore. I will go to the clinic. A kind doctor will help me.

Story

Kago and Pearl were good friends. They liked to play together. One day Kago bought a sweet for herself but forgot to buy one for Pearl. This made Pearl sad and she started to cry. Kago did not want to make Pearl upset and apologized. They returned to the shop to buy another sweet for Pearl.

1) What did Kago buy

a) pencils **b)** fruit
c) juice **d)** sweets

2) Why was Pearl sad?

a) Kago did not play with her.
b) She could not eat sweets.
c) She did not have a sweet.
d) Kago was sad.

EGMA AND EGRA ITEMS

Additional Numeracy Assessment Tool, May 2025: Student Copy

1) Fluency in Addition. Timed: 60 seconds.

1) **1 + 2**	5) **10 + 5**	9) **13 + 9**	13) **14 + 12**	17) **16 + 15**
2) **3 + 3**	6) **21 + 6**	10) **18 + 8**	14) **23 + 11**	18) **19 + 13**
3) **2 + 7**	7) **2 + 27**	11) **26 + 7**	15) **34 + 15**	19) **25 + 17**
4) **4 + 4**	8) **33 + 5**	12) **35 + 6**	16) **24 + 22**	20) **27 + 18**

2) Word Problems. You may use paper and pencil to help you solve these problems

(1)
Neo has **P50**. She spends **P12**. How much is she left with?

(2)
Kago's village has **23** boys and **16** girls. How many children live in Kago's village?

(3)
Thato planted **a total of 30 seedlings** in **3 rows**. He planted the same number of seedlings in each row. How many seedlings did he plant in each row?

(4)
A class is made up of **5 groups** of pupils. There are **6 pupils in each group**. How many pupils are there altogether?

Additional Literacy Assessment Tool, May 2025: Student Copy

1) Oral Reading Fluency. Timed: 60 seconds.

Tubu is a small village in Botswana near a river. Mr Sengadi owned a bus that the villagers used every day. He played music on the bus, and people loved it. They called it *King of the Road*.

One morning, the bus was late. Black smoke came from the back. “What is wrong?” asked Bame. Mr Sengadi did not know.

Bame ran to find Fred, the mechanic. Fred came quickly and looked at the engine. He worked for a while and then fixed the bus. Everyone cheered as the bus started again.

2) Reading Comprehension. Untimed.

(1) What did the villagers call Mr Sengadi’s bus?

(2) Why do you think the villagers liked Mr Sengadi’s bus?