



What Works Hub
for Global Education



**UK International
Development**

Partnership | Progress | Prosperity



What Works Hub
for Global Education

**Evaluating Education Interventions at Scale: Quasi-Experimental
and Experimental Results from India**

Motivation

Motivation

- **Challenges to scalability** - Growing evidence shows that many interventions that worked well in initial proof-of-concept studies fail to produce the same impacts when scaled
- **Most interventions remain small in scale** - across 186 studies, the average intervention reached just 120 schools, and 46% were implemented in only one region.
- **Practical application** - Governments often scale up interventions rapidly and on a large scale, making experimental designs infeasible in many contexts.
- **Feasible, low-cost evaluation tools** - there is a need for methodologies that can be easily and regularly deployed at scale and at low cost

This project

Approach:

- **We develop a low-cost difference-in-differences (DiD) application** that can be feasibly and regularly deployed by governments and implementers.
- **We validate this quasi-experimental approach against experimental data** from TaRL interventions in India.

Application:

- Large-scale education reform implemented across **100,000+ schools and 5 million children to** estimate impacts at scale in India's largest state.

Data sources

Data sources

- **Validation:** Randomized controlled trials of Teaching at the Right Level (TaRL) in **Bihar (2008–2010), Haryana (2012–2013), and Uttar Pradesh (2013–2014)**: approximately 11,000 students in around 1,000 schools
- **Application: Large-scale TaRL rollout in Uttar Pradesh (2018–2019)**: 100,000+ schools, 5–6 million students; monitoring data available for roughly 13,000 schools.

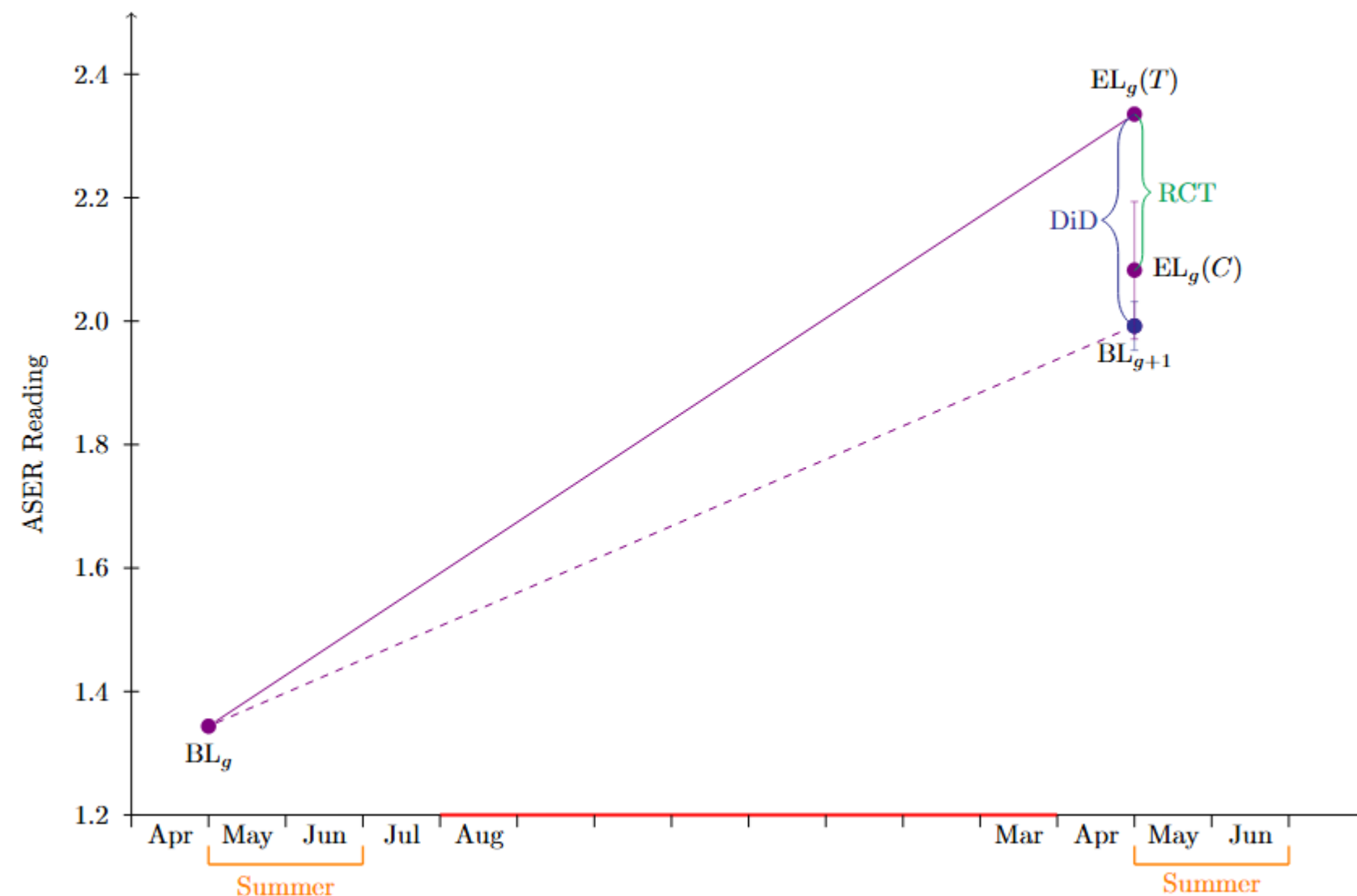
Data coverage
(Bihar, Haryana, Uttar Pradesh)



Approach

Low-cost, at-scale impact evaluation difference-in-difference (DiD) application

Illustration of approach
(Using TaRL intervention in Haryana, India)



Key elements of the approach:

- Treatment group observed in grade g at the beginning of school year, BL_g
- Treatment group observed in grade g at the end of school year, $EL(T)_g$
- Treatment group observed in grade $g+1$ at the beginning of school year, BL_{g+1}

To validate it we also use data from the original RCT, namely:

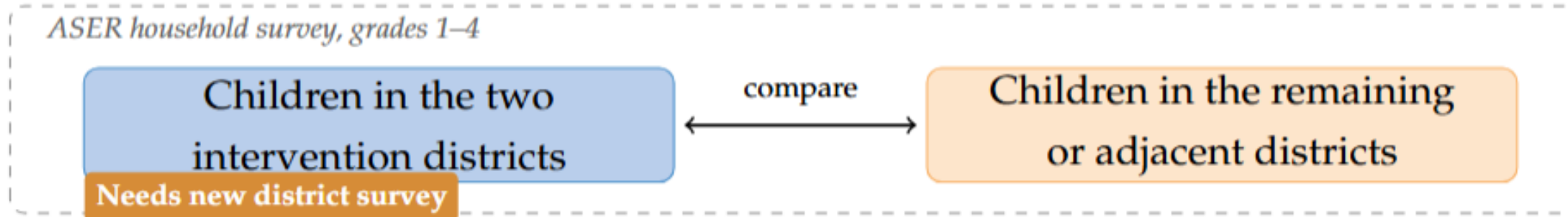
- Control group observed in grade g at the end of school year, $EL(C)_g$

Other visuals: (1) Uttar Pradesh; (2) Bihar summer camp ([appendix](#))

Comparing two DiD typologies: DiD-location and DiD-cohort

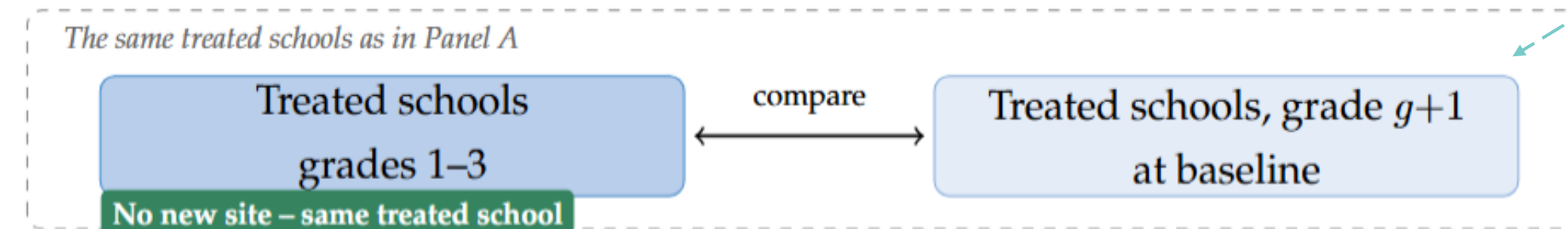
DiD-cohort: Compare the **learning trajectories at baseline** between two adjacent grades **within treated schools**

B. DiD-location: intervention districts against other districts, in survey data



Marginal cost of collecting data in an adjacent grade is lower than collecting data in another school

C. DiD-cohort: treated against adjacent cohorts, within the very same schools



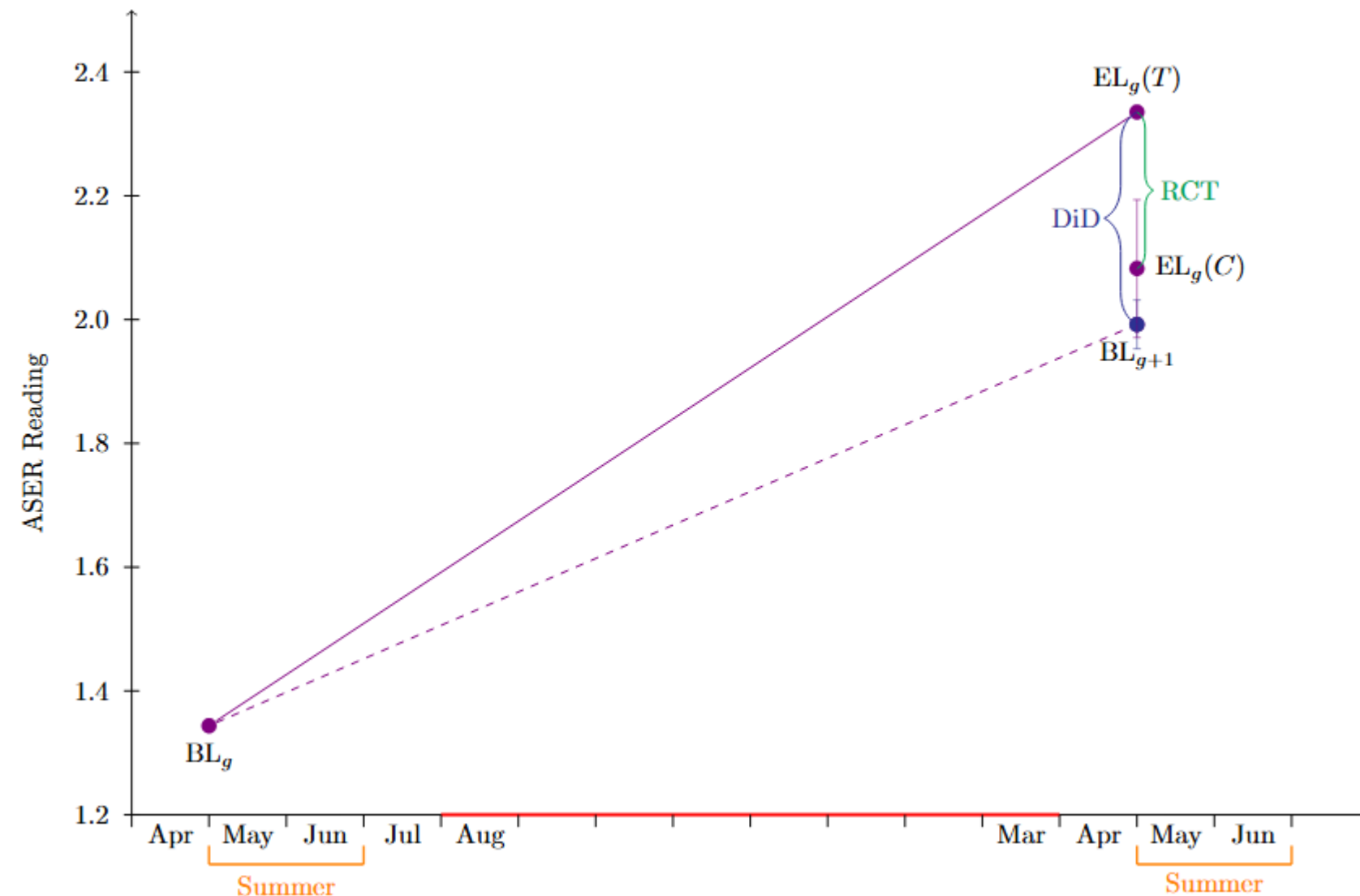
How often is DiD-cohort used?

- **Review of 159 papers** that use difference-in-differences (DiD) approaches, **published in the last five years** in top-five economics journals
- Papers classified **according to the identification strategy** used (location, sector/market) and the **respective field**
- Rare use of cohort DiD approaches – **fewer than 4% of all reviewed studies.**

	Total	Cohort DiD	Location DiD	Sector/Market DiD	Theory	Other
1. Labour, Human Capital & Education	54	6%	72%	0%	0%	22%
2. Public Economics, Taxation & Political Economy	33	9%	61%	0%	0%	30%
3. Finance, Trade, IO & Innovation	33	0%	61%	27%	0%	12%
4. Health, Environment, Family & Development	29	0%	83%	3%	3%	10%
5. Theory & Methodology	10	0%	0%	0%	100%	0%
Total	159	4%	65%	6%	7%	18%

Validate DiD-cohort relative to RCTs, then apply at scale

Illustration of approach
(Using TaRL intervention in Haryana, India)



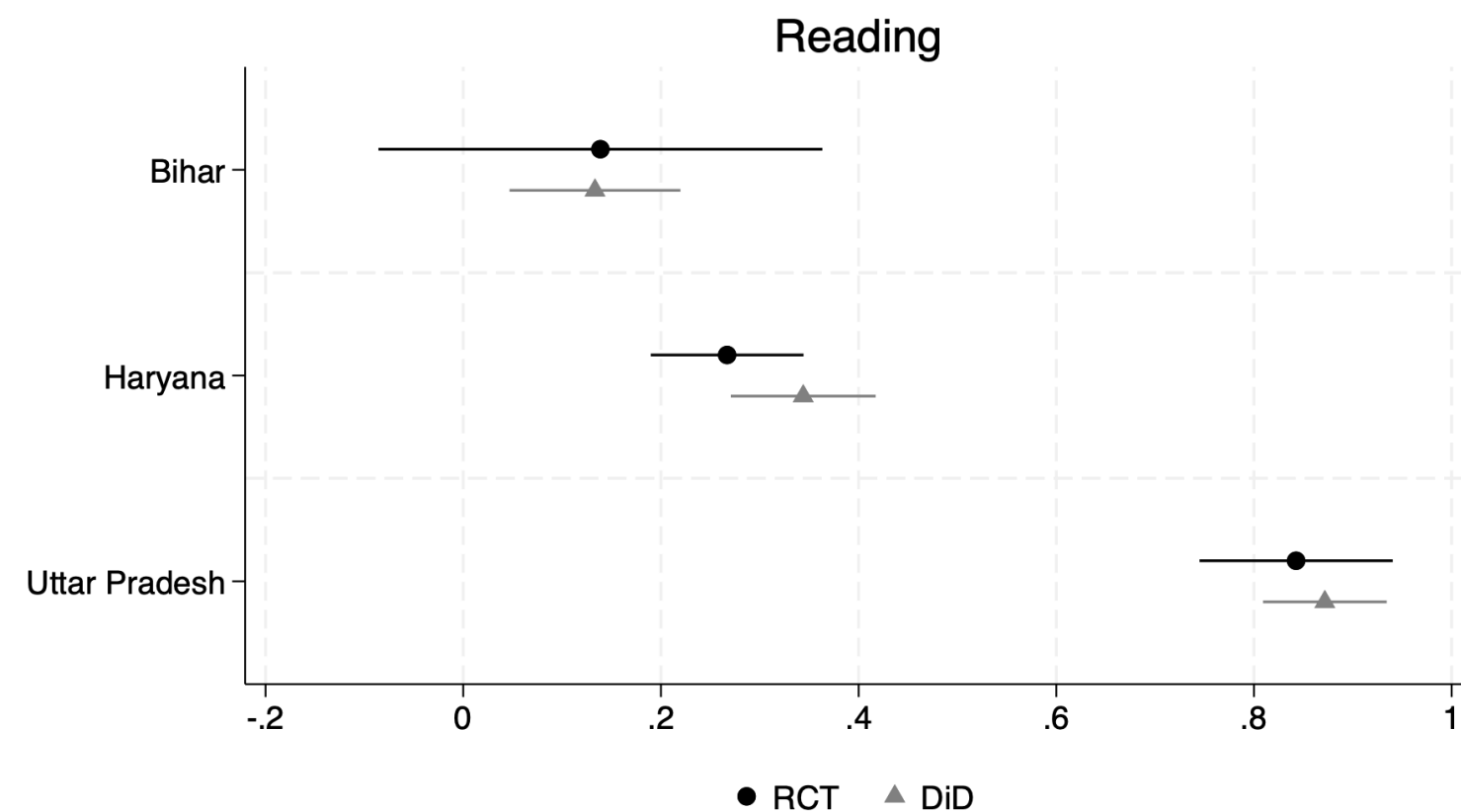
- **First, we validate the DiD-cohort approach relative RCT estimates in the same sample** using the high-profile Case of Teaching at the Right Level studies in India
- **Then, we apply it at low cost and at scale** with 5 million children in Uttar Pradesh

Other visuals: (1) Uttar Pradesh; (2) Bihar summer camp ([appendix](#))

Results

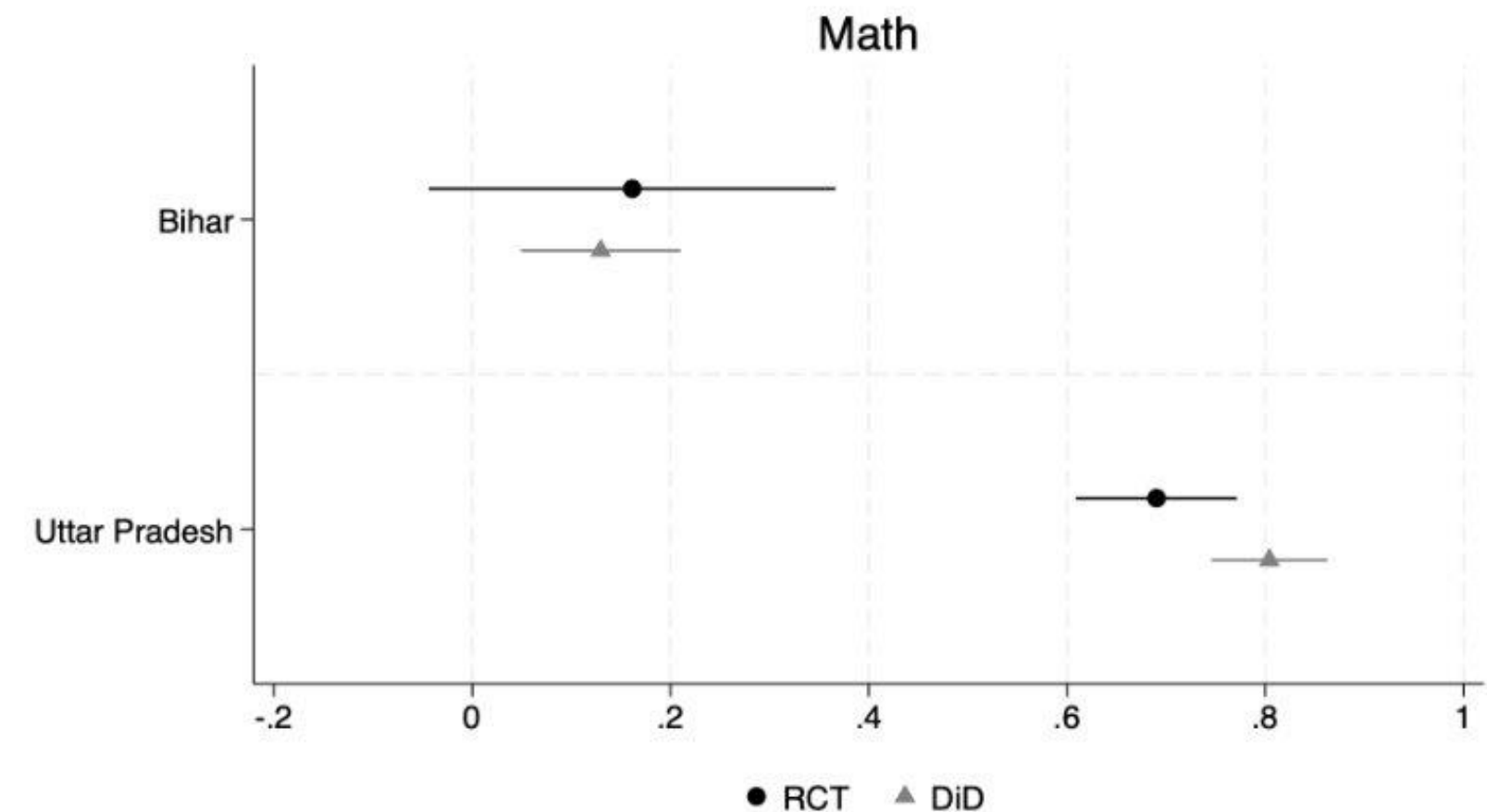
Estimation comparing DiD estimates to original RCT

Effect sizes
(RCT vs DiD)



Notes: 95% confidence intervals using bootstrap SEs (reps. = 1000).

Effect sizes
(RCT vs DiD)



Notes: 95% confidence intervals using bootstrap SEs (reps. = 1000).

Findings show that quasi-experimental estimates align with RCTs in both direction and magnitude, and at scale in Uttar Pradesh effects are large and comparable to RCT efficacy trials.

Assumptions that need to hold

- 1 Learning trajectories – analogous to the parallel trends assumption for DiD
- 2 No sudden cohort-specific shocks other than those associated with the intervention

Learning trajectories (Slope treatment and control)			
	(1)	(2)	(3)
	Bihar	Haryana	Uttar Pradesh
Panel A: Reading			
Slope Treated BL_g - Treated BL_{g+1}	0.060	0.059	0.030
Slope Control BL_g - Control EL_g	0.087	0.071	0.038
Difference	-0.027 [0.339]	-0.013 [0.120]	-0.008 [0.201]
Panel B: Math			
Slope Treated BL_g - Treated BL_{g+1}	0.047	–	0.028
Slope Control BL_g - Control EL_g	0.049	–	0.037
Difference	-0.002 [0.925]	–	-0.009* [0.078]

Learning trajectories (Treated slopes across grades)			
	(1)	(2)	(3)
	Bihar	Haryana	Uttar Pradesh
Panel A: Reading			
Slope Treated BL_g - Treated BL_{g+1}	0.060	0.059	0.030
Slope Treated BL_{g+1} - Treated BL_{g+2}	0.027	0.069	0.038
Difference	0.033 [0.219]	-0.010 [0.444]	-0.008 [0.431]
Panel B: Math			
Slope Treated BL_g - Treated BL_{g+1}	0.047	–	0.028
Slope Treated BL_{g+1} - Treated BL_{g+2}	0.021	–	0.027
Difference	0.025 [0.243]	–	0.001 [0.932]

Notes: Standard errors for slope differences clustered at the school-grade level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
P-values of slope differences are reported in brackets.

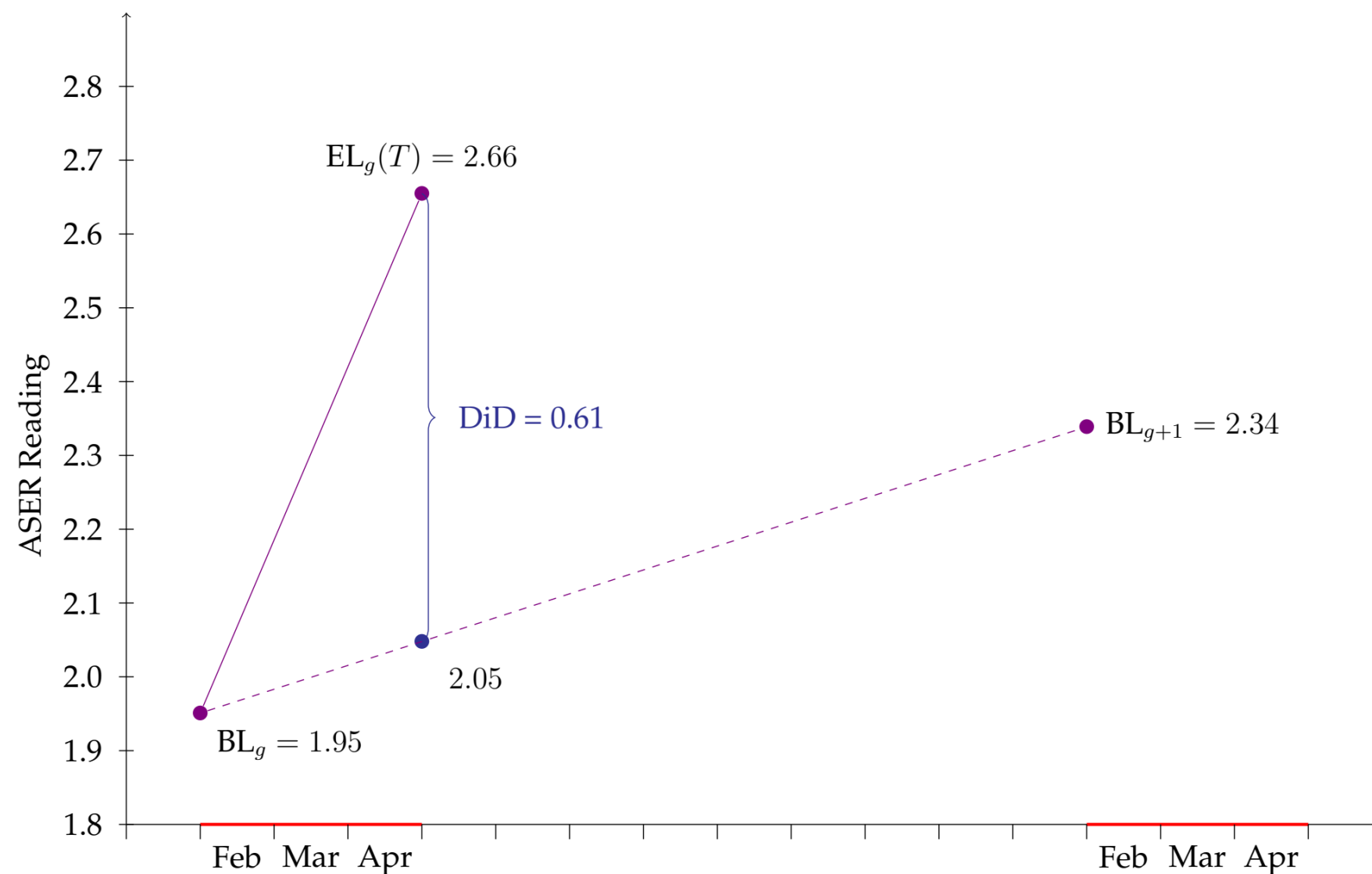
Evaluation at scale

- Statewide **scale-up of the TaRL model implemented in Uttar Pradesh.**
- The intervention was carried out during the **2018–2019 academic year and reached more than 100,000 schools and 5 million students.**
- **No randomized control group** in this scaled-up intervention — target grades: 3–5.

Evaluation at scale – Main learning results

Illustration of approach

Evaluation at scale: Uttar Pradesh, reading



- Average impacts of approximately **0.58 levels in math and 0.61 levels in reading**.
- **These effect sizes are substantial, particularly given the program's scale** and the implementation challenges associated with statewide delivery.
- **These effects are comparable to those observed in smaller-scale randomized trials.**

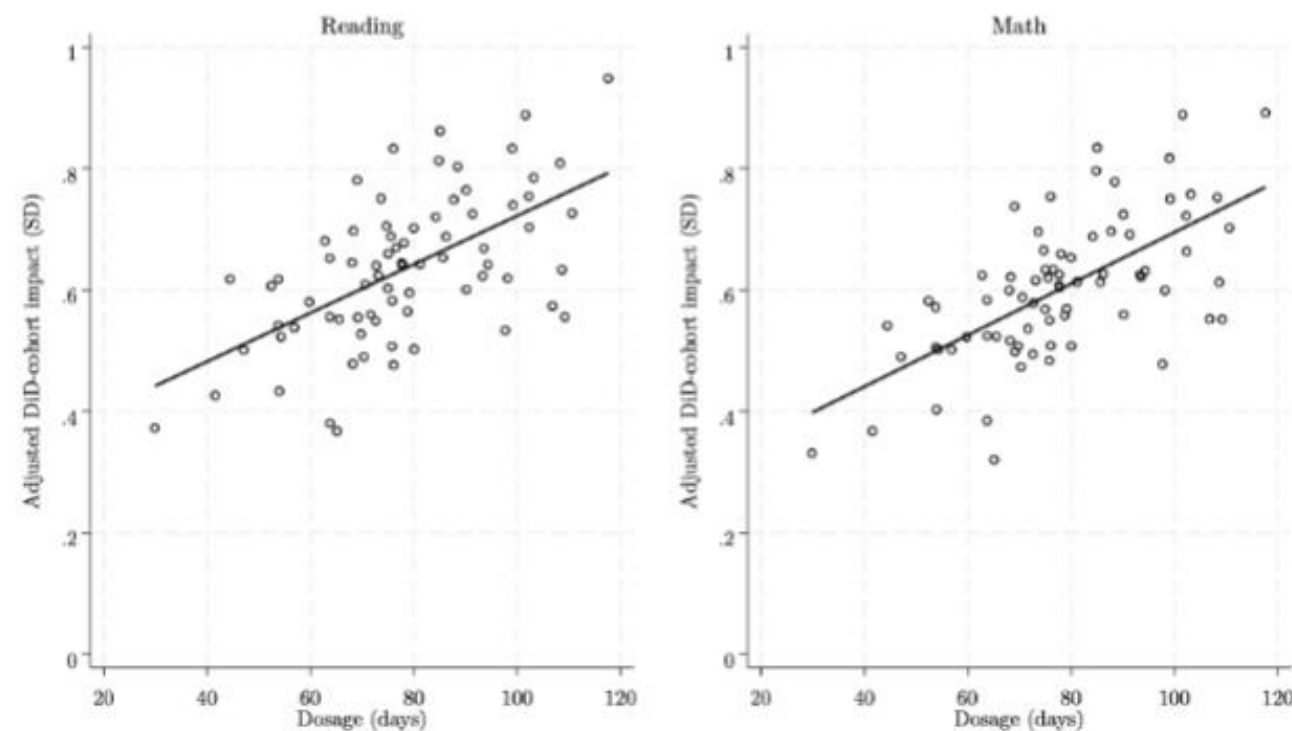
Source: What Works Hub for Global Education. Evaluation at scale (N. Angrist, P. Freitas, J. Rodriguez) (upcoming publication)

Evaluation at scale – Implementation results

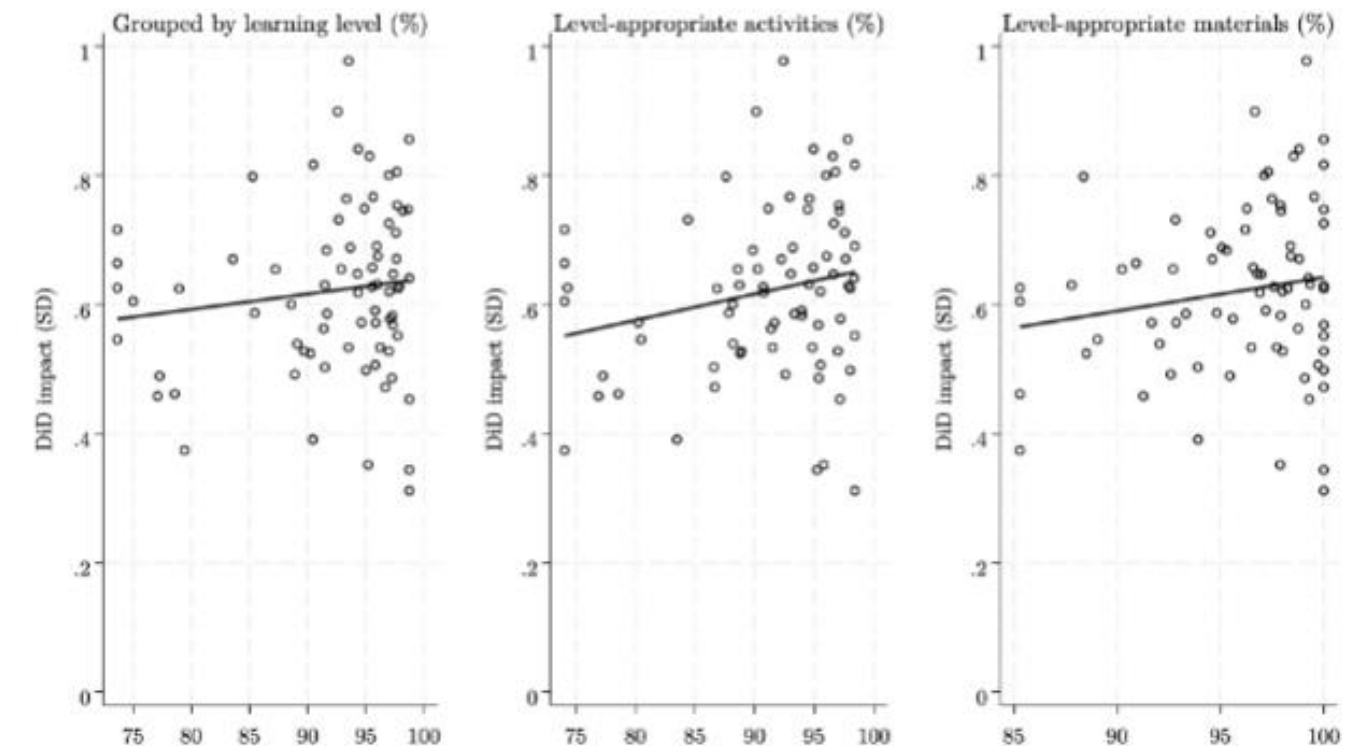
- At the district level, **results correlated with several implementation variables**

DiD-cohort effect size showed the **strongest positive correlation with intervention dosage (in days)**

Quantity measures (e.g. dosage)



Quality measures



Conclusion

Conclusion

- **Rigorous evaluation at scale is particularly constrained** in settings where randomized assignment is not feasible.
- We have extended the canonical **DiD approach that exploits learning trajectories between grades** to be able to evaluate education interventions in a more cost-effective manner.
- Our estimates **show a high degree of alignment with randomized trial results** (Banerjee et al. 2007, 2017).

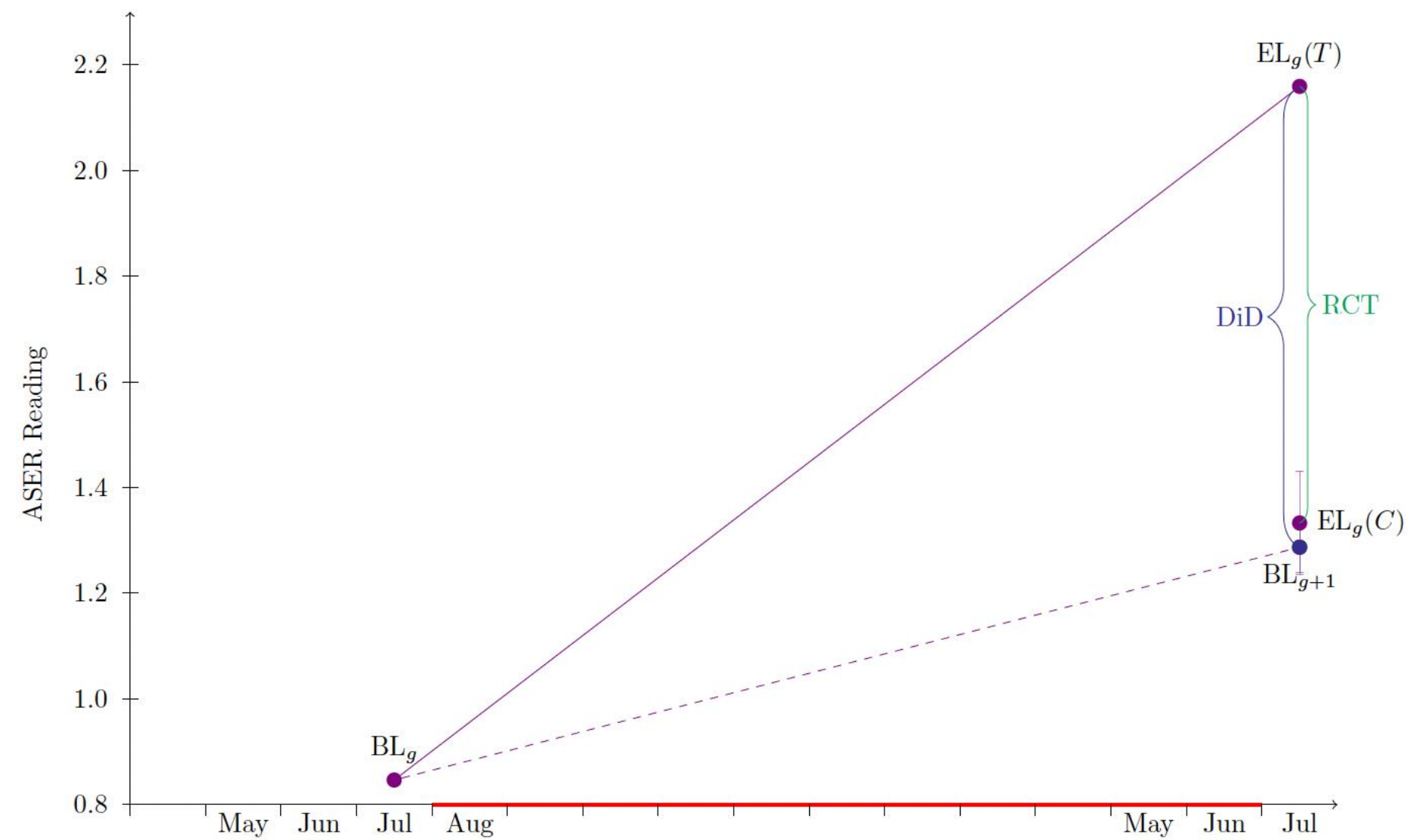
Conclusion

- Application to a **large-scale TaRL roll-out** shows **large, sustained impacts state-wide**.
- The approach **requires only within-school cohort data, substantially reducing data collection costs** compared with designs that require information from multiple regions.
- This evaluation-at-scale approach **is part of the WWGHE TA offer** and can be readily deployed in large-scale programs implemented by funders, implementers, or governments

Appendix

How the approach works (visually)

Illustration of approach
(Using TaRL intervention in Uttar Pradesh, India)



How the approach works (visually)

Illustration of approach
(Using TaRL intervention in Bihar, India)

