

Knowledge Tracing for Early Childhood Learners with Minimal Telemetry from Low-Connectivity Environments: Evidence from India's *Anganwadi* Ecosystem

Badmavasan Kirouchenassamy¹ Rahul Singh¹

Vishnu Dev¹ Meenakshi Garg¹

Sukhna Sawhney¹

Sudeep Gowrishankar¹

¹Rocket Learning

{badmavasan.k, rahul.s, vishnu.d,
meenakshi.g, sukhna.s, sudeep.g}@rocketlearning.org

Abstract

Many of the students who could benefit most from distance learning lack the infrastructure it assumes. Most knowledge-tracing research takes personal devices, stable internet, and rich interaction logs for granted, and it has mostly been developed and evaluated with older learners. Early childhood in low-connectivity environments is therefore still underexplored. We ask whether two established Knowledge Tracing (KT) models, Bayesian Knowledge Tracing (BKT) and the Dynamic Key-Value Memory Network (DKVMN), can accurately model early-childhood Hindi literacy in India's *Anganwadi* ecosystem. Both were designed for high-resource settings, whereas here browser-based Learning Games are played on shared phones over slow, intermittent connections. Using only a timestamped sequence of correct and incorrect responses from 210,500 question attempts by 7,547 learners aged 3 to 6, we validate the models' mastery estimates against offline field assessments of 552 children. DKVMN predicts the next answer well (AUC 0.858), on par with high-resource benchmarks, and its mastery estimates correlate significantly with field assessments from as few as three attempts. BKT is less accurate overall, but gives more reliable per-skill estimates, interpretable diagnostics, and greater robustness to submissions lost to poor connectivity. These results show that meaningful learner modeling is feasible for early childhood education in low-connectivity environments, and that evaluating such models requires external validation against offline assessments and robustness to data loss.

1 Introduction

The *Anganwadi* system is the world's largest publicly funded early-childhood care and development program, providing nutrition, health,

daycare, and early-childhood education (ECE) through a nationwide network of *Anganwadi* Centers (AWCs). For many low-resource families, these centers are the primary point of access to ECE (ASER Centre, 2025). Rocket Learning (RL), an Indian nonprofit, partners with this system by training *Anganwadi* Workers (AWWs), building WhatsApp communities that connect workers and parents, and delivering play-based activities for AWCs and home. RL also sends hyperlinks to mobile browser-based Learning Games (LGs) over WhatsApp, letting children practice foundational skills without installing an app (Rocket Learning, 2024).

RL's model already drives large-scale engagement and learning gains across several Indian states, but evaluation at this scale still relies on periodic, sampled field measurements, which makes it hard to monitor every child's progress. Reliable learner-level estimates from LGs could instead flag children who need support, personalize learning, and speed up content iteration. We do not propose a new platform; we strengthen RL's intervention by enabling continuous learner-level measurement from minimal telemetry.

This is hard in RL's deployment environments. Dedicated devices and unrestricted data plans are rare, so games are delivered through lightweight web links that must work despite intermittent connectivity, shared phones, and limited bandwidth. As a result, only a little interaction data can be collected consistently. Yet most knowledge-tracing research assumes persistent connectivity, personal devices, and rich logs, so it is unclear whether these methods still help with minimal telemetry.

This gap motivates our **Research Question (RQ)**:

To what extent do standard Knowledge Tracing (KT) models, specifically

classical Bayesian Knowledge Tracing (BKT) and neural Dynamic Key-Value Memory Networks (DKVMN), remain effective in Hindi early-childhood literacy education when learner data are limited to sparse, intermittent, and shared-device telemetry, and how can their performance be rigorously evaluated in such low-resource settings?

Knowledge tracing (KT) estimates a student’s evolving mastery of skills from their sequence of right and wrong answers (Corbett and Anderson, 1994). It has been studied extensively in well-instrumented classrooms, universities, and online platforms (Piech et al., 2015; Zhang et al., 2017; Gervet et al., 2020), but it is relatively new in settings like ours. The closest precedents are adaptive learning deployments in South America and other low- and middle-income regions, especially for early childhood education (Araya and Diaz, 2020; Nye, 2015) (Section 2).

This paper makes three contributions.

1. **Established KT models transfer to low connectivity:** We show that minimal telemetry from mobile browser-based games, a timestamped (learner, skill, correctness) sequence, is enough to train both a classical (BKT) and a neural (DKVMN) KT model and to produce learner-level estimates useful for monitoring and personalization. DKVMN reaches a next-step AUC of **0.858** on held-out students, comparable to standard high-resource benchmarks, and BKT stays the more robust of the two when poor connectivity drops submissions (Section 7.5).
2. **An evaluation methodology for such environments:** Next-step prediction accuracy alone can overestimate model quality, because it rewards models that exploit game difficulty rather than track learner mastery. We therefore pair internal evaluation (student-level splits, cross-validation) with *external field validation*: mastery estimates are correlated against offline assessments of **552 children**, administered on paper by trained evaluators. This channel needs no connectivity of its own (Section 6).
3. **Reusable infrastructure and a useful baseline:** We treat interpretable BKT as

a deliberate starting point rather than an exhaustive model comparison, use its learned parameters to diagnose game design and learning dynamics, and document where both models fail (ceiling effects, rare skills, and negative correlations). The same harness lets future work evaluate other BKT variants and newer KT models.

2 Background and Related Work

Educational technology in low-resource settings.

Nye (2015) reviews intelligent tutoring systems in the developing world and catalogs the barriers that shape our design space: shared, mobile-first devices, unreliable electricity, data costs, and intermittent internet. Despite these barriers, well-designed systems work. Mindspark produced large test-score gains for secondary students in urban India (Muralidharan et al., 2019); ConectaIdeas improved mathematics in randomized trials in Chilean primary schools and was adapted in Peru (Araya and Diaz, 2020); and tablet programs such as onebillion in Malawi (Pitchford, 2015; Pitchford et al., 2019) and E-Learning Sudan (Stubbé et al., 2016) showed early-grade children learn from adaptive software with no schooling infrastructure at all. These deployments show that such *interventions* can improve outcomes at the population level, but comparatively little work models the trajectories of *individual students* in these settings. That is the gap we address, and tracking individual progress is a prerequisite for instruction that adapts to each learner.

Knowledge tracing. BKT models each skill as a two-state hidden Markov model: the student either knows the skill or does not, and four probabilities govern initial knowledge, learning, guessing, and slipping (Corbett and Anderson, 1994). Its interpretability has kept it in production for three decades, and a rich family of variants exists, including individualized parameters (Yudelson et al., 2013; Pardos and Heffernan, 2010) and contextual guess and slip (Baker et al., 2008); see Pelánek (2017) for a survey. Deep learning models replaced the hand-built state with learned representations: Deep Knowledge Tracing (Piech et al., 2015), the Dynamic Key-Value Memory Network (DKVMN) that maintains explicit per-concept memory (Zhang et al., 2017), and attention-based successors (Pandey and Karypis,

2019; Ghosh et al., 2020). Khajah et al. (2016) and Gervet et al. (2020) showed that the gap between classical and deep neural models depends heavily on the size and structure of the dataset. This is directly relevant here, where some skills have thousands of learners and others have just four. Recent work treats *low-resource knowledge tracing* as a problem in its own right, using pre-training to cope with scarce interactions (Zhang et al., 2024), but still evaluates on standard online-platform datasets. To our knowledge, ours is the first KT study trained on mobile browser-game telemetry from a low-resource, low-connectivity population and validated against offline field assessments.

Early literacy in Hindi. Our target skills focus on early reading in Hindi, written in the Devanagari script. Unlike alphabetic systems, Devanagari is an *akshara*-based orthography with a large symbol inventory, so letter knowledge is acquired more gradually (Nag, 2007). Foundational literacy thus depends on both phonological awareness (e.g., identifying initial sounds) and letter-sound knowledge. India’s national priorities emphasize both (Government of India, 2020; ASER Centre, 2025), and both are reflected in the skills included in this study.

3 The Learning Framework

Rocket Learning’s ECE content is layered. A **skill** in this study is defined as a (*learning module*, *learning competency*) pair such as FL-3.5.1_FL-LC49. A **Learning Domain (LD)** is a developmental area such as foundational literacy (FL); all our skills fall within the foundational literacy LD. A **Learning Module (LM)** is a themed content unit (e.g., FL-3.5.1), typically one game targeting a piece of curriculum. A **Learning Competency (LC)** is the observable ability it exercises (e.g., FL-LC49, “recognizes words that begin with the same sound”); there is a many-to-many mapping between LMs and LCs.

In this study, we focus on 32 skills that fall into four families: twelve modules of *beginning-sound recognition* (LC49), twelve modules of *beginning-sound differentiation* (LC50), and four modules of Devanagari letters (*ma*, *ka*, *cha*, *ba*), each at *basic* (LC34.1) and *advanced* (LC34.2) levels of letter-sound knowledge.

Learning Games (LGs) mainly reach children between 3 and 6 years of age, as personalized hyperlinks sent on WhatsApp to parents’ phones

Skill family	Skills	Attempts	Learners	% cor.
Beginning sounds: recognition	12	100,866	7,140	86.9
Beginning sounds: differentiation	12	91,944	6,599	82.6
Letters: basic	4	8,876	1,214	35.3
Letters: harder	4	8,814	1,208	36.8
Total	32	210,500	7,547	80.8

Table 1: Training data by skill family (December 2025). “% cor.” is the percentage of correct responses within each family.

(Rocket Learning, 2024). Each LG is a short sequence of questions answered through touch interactions, and each interaction is scored and logged. Crucially for this paper, the only telemetry we can rely on across devices and connection qualities is minimal: (*learner*, *timestamp*, *skill*, *isCorrect*). Everything that follows is built from this four-tuple.

4 Data

Training telemetry. The training set has 210,500 scored question attempts by 7,547 learners from December 2025 (Table 1); each attempt is binarized (a score above zero is correct). Where age is recorded (43% of attempts), learners are almost entirely 3- and 4-year-olds (61,859 and 27,652 attempts; only 146 from ages 5–6). Learners are identified by a persistent device-browser identifier. This is a constraint of the setting: digital literacy is low, and adding a login step causes large drop-off (Section 8). Engagement is short but meaningful: the median learner contributes 12 attempts (mean 27.9; top decile >60), and the median learner-skill sequence is 2 attempts (mean 3.1) across 68,356 sequences. About 27% of attempts fall on a later-day return. Any model used here must therefore say something useful from very few observations.

Assessed-cohort telemetry and field assessments. Separately, 9,275 attempts by 783 children were collected in April–May 2026 from a cohort that also took the *LCC field assessment*, a one-on-one, paper-based test administered offline by trained evaluators and covering 10 of the 32 skills in this study. For external validation, we matched child identifiers across the two datasets and kept only child-skill pairs with at least five question attempts. This yielded 552 overlapping children and 1,055 child-skill pairs.

5 Models

Why start with BKT. Our purpose is not to compare KT models exhaustively. It is to establish, with the most interpretable tool available, that student modeling is possible in this challenging environment, and to build infrastructure for richer comparisons in future work. We therefore begin with BKT, whose parameters have intuitive interpretations that curriculum designers can use when creating and revising content.

BKT. For each skill, we fit a two-state hidden Markov model (Corbett and Anderson, 1994) with four parameters, estimated by Baum–Welch expectation-maximization: $P(L_0)$, the probability that a child already knows the skill before playing; $P(T)$, the probability of *transitioning* to the known state after each practice opportunity (the learning rate); $P(G)$, the probability of *guessing* correctly while not knowing; and $P(S)$, the probability of *slipping* (answering wrongly despite knowing). After each observed attempt, the posterior probability of mastery $P(L_n)$ is updated by Bayes’ rule. We train 32 independent models, one per skill.

DKVMN. As a strong learned counterpart, we train a single Dynamic Key-Value Memory Network (Zhang et al., 2017) over all 32 skills jointly. It maintains a key matrix of latent concepts and a value matrix of mastery, read and written through attention. This shares statistical strength across skills, which helps skills with sparse data but, as we will see, occasionally hurts them.

6 Methodology : Evaluating in a Low-Connectivity Environment

Figure 1 summarizes the pipeline. Our second contribution is the evaluation design itself, which we argue is the right template for settings like ours.

Internal evaluation: next-attempt prediction. We split *students*, not interactions: 85% train+validation / 15% test (6,414 vs. 1,133 learners), stratified by each student’s most practiced skill, with stratified 5-fold cross-validation inside the training portion. Both models share identical splits. Splitting by student matters here: with shared devices and short sequences, splitting by interaction would leak a child’s own history into training.

Model	Eval	AUC	Acc	F1	LL
BKT	CV	.654 $\pm$.005	.598	.529	.701
BKT	Test	.656	.609	.532	.711
DKVMN	CV	.852 $\pm$.002	.845	.715	.352
DKVMN	Test	.858	.849	.716	.341

Table 2: Internal evaluation: 5-fold cross-validation (mean $\pm$ std) and held-out test students. LL = log-loss.

External validation: the field as ground truth.

Internal metrics such as next-step prediction cannot tell us whether the predicted “mastery” reflects what a child actually knows or merely patterns in the game. To check this, we train the models on December 2025 data and use them to estimate mastery for the April–May 2026 assessment cohort after 3, 5, 10, and all 12 attempts. We then compare these estimates with the independent offline assessments and report Pearson correlations (r , 95% CI) per skill and globally.

7 Results

7.1 Internal: predicting the next attempt

Table 2 reports next-attempt prediction on held-out students across all skills. DKVMN clearly outperforms BKT (AUC 0.858 vs. 0.656), with little variance across folds. For context, this matches the high end of state-of-the-art KT on high-connectivity benchmarks: DKVMN reports AUCs of roughly 0.73–0.83 on ASSISTments and Statics (Zhang et al., 2017), and Gervet et al. (2020) place most models between 0.70 and 0.85 across nine datasets. Minimal browser-game telemetry thus supports next-step prediction *on par with standard benchmarks*, our first main finding. Part of DKVMN’s edge is diagnostic rather than pedagogical: it learns that letter games are hard and sound games easy, something BKT’s per-skill structure already encodes. This is what motivates the external validation below.

7.2 BKT’s learned parameters, read aloud

Because BKT’s parameters mean something, we can read them (Figure 2). For the sound families, the model explains near-ceiling performance as *fast learning with generous guessing*: children start out not knowing the skill ($P(L_0) \approx .01$) but transition almost immediately ($P(T) \approx .8$ –.9) and guess often ($P(G) \approx .38$). For the letter families, the learning rate collapses to near zero. BKT assumes mastery can only increase, so a within-session accuracy *decline* cannot be reconciled with a

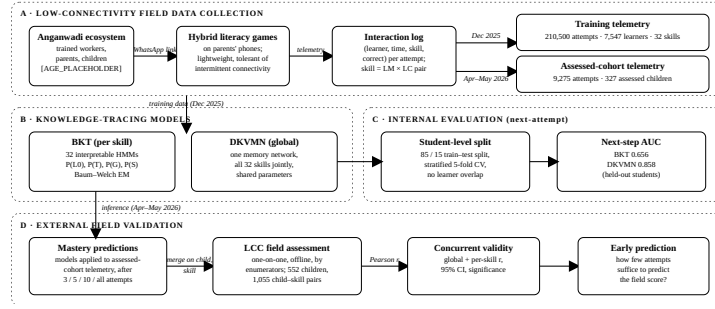


Figure 1: Evaluation pipeline. (A) Games reach families as web-links on WhatsApp; the only telemetry is a minimal four-tuple per attempt. (B) Two KT models train on December 2025 data. (C) Internal evaluation uses student-level splits (no child in both train and test). (D) External validation correlates predicted mastery with offline field assessments of the same children.

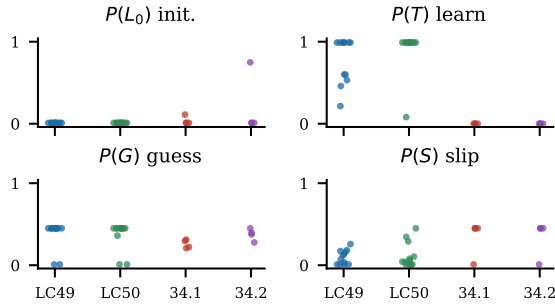


Figure 2: Learned BKT parameters for all 32 skills by family. Many sit at the bounds of their ranges (e.g. $P(T) = 0.99$, $P(G) = 0.45$), reflecting BKT’s known identifiability limitation, where many parameter combinations explain the data equally well.

Model	N	r	95% CI	p
BKT	1,055	.061	[.001, .121]	.048
DKVMN	1,055	.170	[.111, .228]	2.7×10^{-8}

Table 3: Global correlation between predicted mastery and offline field assessment scores (552 children, 10 skills, 1055 child-skill pairs).

positive $P(T)$; instead a high slip rate ($P(S) \approx .34$) absorbs the within-game difficulty ramp. The fits also expose the model’s limits: 30 of 32 skills sit at the $P(L_0)$ floor and many at the $P(T)$ or $P(G)$ ceiling (Figure 2), which is BKT’s classic identifiability and degeneracy problem (Beck and Chang, 2007; Baker et al., 2008). Still, these interpretable parameters give curriculum designers a useful lever: flatten the within-game difficulty ramp to better match learners’ abilities and improve model fit.

7.3 External: does predicted mastery match the field?

Globally (Table 3, Figure 3), both models correlate positively with the offline assessments, but

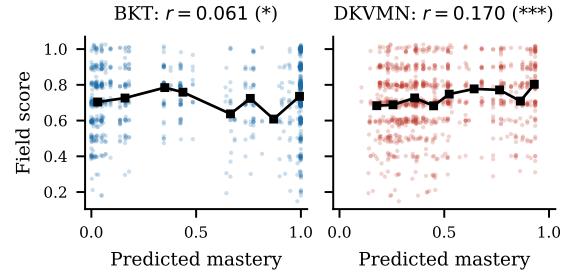


Figure 3: Predicted mastery vs. field assessment score for all 1,055 child-skill pairs (jittered), with binned means. Both relationships are positive but modest; DKVMN’s is three times stronger.

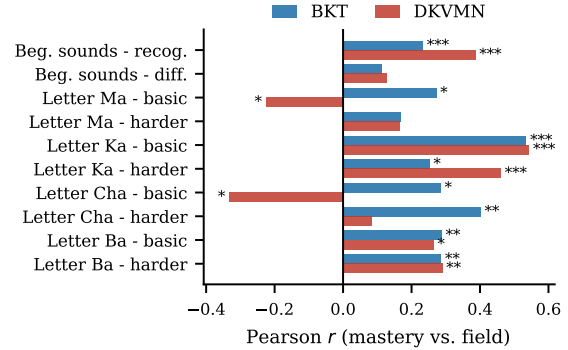


Figure 4: Per-skill correlation with the field assessment. BKT is significant on 8 of 10 skills and always positive; DKVMN is stronger where data is plentiful but significantly *negative* on two rare letter skills. Skills labelled “hard” are the advanced (LC34.2) variants.

modestly: BKT at $r = .061$ (barely significant) and DKVMN at $r = .170$ ($p < 10^{-7}$). Two constraints keep these numbers small. The assessment has to be brief given the children’s young ages, which limits the reliability of any single score, and several skills vary little across learners, which limits what a correlation can reveal.

The per-skill view (Figure 4) is more instructive than the global number, and this is where the two

models trade places.

Where both models work. On the letter *ka*, the best-represented letter skill, both models reach $r \approx .54$ ($p < 10^{-7}$), and DKVMN reaches $r = .46$ on its advanced variant. On beginning-sound recognition, DKVMN reaches $r = .39$ versus BKT’s $.23$. So for skills with enough data and enough variation, predicted mastery aligns well with field assessments.

Where BKT is safer. BKT is significant on 8 of 10 skills and *positive on all ten*. DKVMN is significant on 6 of 10, but two of those are significantly negative (letters *ma* and *cha*, basic; $r = -.23$ and $-.33$), so for these skills its predictions are actively misleading. These letters contribute only 1.7–4% of training attempts, while the sound games dominate (~80% of attempts). DKVMN’s shared parameters follow that majority and project its pattern, “children who play a lot succeed”, onto skills where it does not hold. BKT trains one small model per skill, so it cannot be misled by the statistics of another skill. In low-resource contexts, where some skills will often be rare, this failure mode argues for keeping an interpretable per-skill model in the loop, whatever the global AUC says. Behavior also varies across skills within a single model, which suggests that neither of the two formulations is optimal here and that different skill families may need different modeling assumptions.

Where both fail: the ceiling. Neither model correlates on beginning-sound differentiation (LC50). Figure 2 gives the diagnosis: with 91–98% correct rates in those games, every child’s mastery estimate saturates near 1.0, and a constant cannot correlate. The fix is pedagogical, not statistical.

7.4 Early prediction: how many attempts are needed for reliable prediction?

A program that must detect mastery early, so it can adapt games and flag children who need support, cares less about whether a model is eventually right than about how soon it is right. Following the truncated-history analysis of the DKVMN literature (Zhang et al., 2017), we re-estimate each child–skill mastery from only the first k attempts and re-correlate with field scores (Figure 5). DKVMN is already significant after **three** attempts ($r = .141$, $p < 10^{-5}$), about a minute of play, and gains little beyond ten ($r = .170$, its full-history

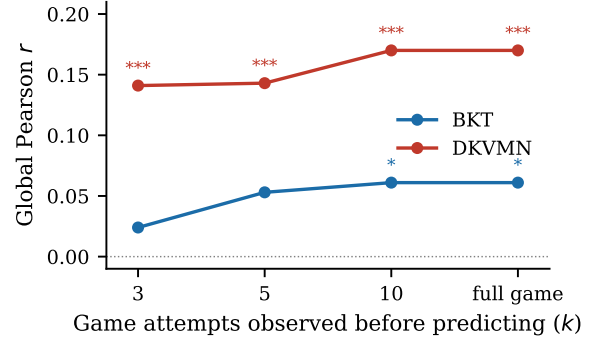


Figure 5: Global correlation with the field assessment when mastery is estimated after only the first k attempts per child–skill pair. DKVMN is significant from $k=3$; BKT needs the full game.

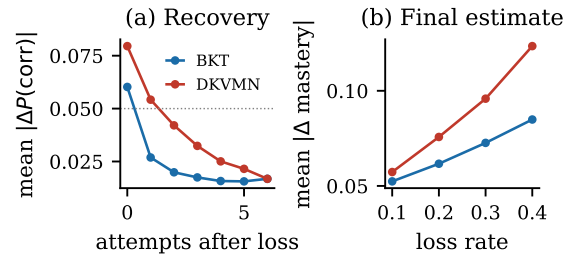


Figure 6: Robustness to submissions lost to poor connectivity. (a) Recovery after a single lost attempt. (b) Drift of the final mastery estimate under independent loss at rate ρ . BKT is more stable in both.

value). BKT becomes only marginally significant once the full game is seen ($r = .024$ at $k=3$, $.061$ at $k=12$). Since the median training learner–skill sequence is just 2–3 attempts (Section 4), this is perhaps the most deployment-relevant result: with the right model, the very short sessions this environment produces already carry a meaningful signal about a child’s mastery.

7.5 Robustness to lost submissions

Intermittent connectivity drops submissions before they reach the server, so a model updates on a partial sequence. On the assessed-cohort games, we take every child–skill sequence of at least five attempts, randomly drop attempts, and compare the model’s belief on the thinned stream against its full-data belief, using the predicted probability of the next answer as a common signal. Two patterns emerge (Figure 6):

- *Recovery*: a single lost attempt perturbs the prediction, but the effect fades as more attempts are observed. BKT returns close to its original estimate after one further attempt, while DKVMN needs two to three and starts

from a larger perturbation.

- *Graceful degradation*: when submissions are dropped at rate ρ , the final mastery estimate moves less for BKT than for DKVMN (mean $|\Delta|$ 0.068 vs. 0.088 over $\rho \in [0.1, 0.4]$) and flips fewer master/non-master decisions (6.3% vs. 8.6%), with the gap widening as ρ increases.

A plausible explanation is that BKT’s per-skill update is self-correcting and immune to other skills’ data, whereas DKVMN’s shared memory spreads each loss across concepts. This is another reason to keep the interpretable model in the loop where connectivity is unreliable.

8 Discussion and Limitations

What we claim. Minimal four-tuple telemetry from WhatsApp-delivered games supports (i) next-step prediction on par with high-connectivity benchmarks, (ii) significant agreement with offline field assessments on well-supported skills, (iii) usable signal within three attempts, and (iv) predictions that degrade gracefully when connectivity drops submissions, BKT most robustly. We do not claim the estimates are accurate for every skill or precise enough for individual personalization. We claim only that even sparse game telemetry carries enough signal to flag children who may need support. This establishes the feasibility of large-scale learner-level modeling in low-resource, low-connectivity early-childhood literacy, and lays a foundation for continuous progress monitoring, adaptive learning, and timely intervention.

Limitations. Learner identity is a device-browser identifier: siblings sharing a phone blur sequences, and a child who changes phones may appear as several learners. Scores are binarized, discarding partial credit. The assessed cohort is small for rare skills ($N=46-90$ per letter). Field ground-truth covers only 10 of the 32 skills; the rest are validated internally. Finally, games are played at home, where parental help varies between sessions and is unobserved.

Future work. The infrastructure now exists to make the next steps cheap: BKT variants designed for exactly our pathologies (individualized priors (Yudelson et al., 2013; Pardos and Heffernan, 2010), contextual guess/slip (Baker et al., 2008),

forgetting (Khajah et al., 2016)), attention-based models (Pandey and Karypis, 2019; Ghosh et al., 2020), and low-resource pre-training (Zhang et al., 2024) can all be evaluated against the same student splits and field-validation harness. On the program side, the gains are pedagogical: flatten within-game difficulty ramps, harden the ceiling-bound sound games, log partial credit, add question-difficulty information, and adjust for parental help.

More broadly, the long-term value of KT here may lie less in personalization than in system-level program improvement. Instead of periodic sampled evaluations, operators could monitor every learner, spot under-performing content, and refine pedagogy and game design continuously. Future models should also combine complementary evidence, such as worksheets, classroom activities, engagement, and educator observations, for more robust mastery estimates.

9 Conclusion

The children of India’s Anganwadi ecosystem will not soon have the devices, bandwidth, or rich logging that much knowledge-tracing research assumes. Our results show that meaningful student modeling can begin without them: a timestamped stream of right and wrong answers from lightweight browser-based games is enough to train models whose next-step accuracy matches high-connectivity benchmarks, whose mastery estimates agree with offline field assessments after as few as three attempts, and which stay usable even when connectivity drops submissions.

The interpretable BKT baseline also proved valuable as a diagnostic: it exposed steep within-game difficulty ramps and flagged skills where sparse data give unstable estimates. By showing both competitive predictive performance and measurable agreement with independent field assessments, this work establishes that learner-level modeling is feasible in low-resource, low-connectivity settings using only minimal telemetry, and it provides a practical foundation for future adaptive learning systems at scale.

Acknowledgements

We thank the Rocket Learning team including Azeez Gupta, Utsav Kheria, Namya Mahajan, and Siddhant Sachdeva for enabling this collaboration, and to Bilal Afroz, Pratik Purohit, Swati Pandey, Pratibha Rohal, Drishti Mrigwani, Simran Kaur,

and Asha Tiwari for their support in game design and field data collection. We also want to thank the Anganwadi Workers and parents who participated in the benchmarking study that enabled this research.

References

- Roberto Araya and Karina Diaz. 2020. Implementing government elementary math exercises online: Positive effects found in RCT under social turmoil in Chile. *Education Sciences*, 10(9):244.
- ASER Centre. 2025. Annual status of education report (ASER) 2024: National findings. Technical report, ASER Centre, Pratham, New Delhi, India.
- Ryan S. J. d. Baker, Albert T. Corbett, and Vincent Aleven. 2008. More accurate student modeling through contextual estimation of slip and guess probabilities in Bayesian knowledge tracing. In *Proceedings of the 9th International Conference on Intelligent Tutoring Systems (ITS 2008)*, pages 406–415.
- Joseph E. Beck and Kai-min Chang. 2007. Identifiability: A fundamental problem of student modeling. In *Proceedings of the 11th International Conference on User Modeling (UM 2007)*, pages 137–146.
- Albert T. Corbett and John R. Anderson. 1994. Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4(4):253–278.
- Theophile Gervet, Ken Koedinger, Jeff Schneider, and Tom Mitchell. 2020. When is deep learning the best approach to knowledge tracing? *Journal of Educational Data Mining*, 12(3):31–54.
- Aritra Ghosh, Neil Heffernan, and Andrew S. Lan. 2020. Context-aware attentive knowledge tracing. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD 2020)*, pages 2330–2339.
- Government of India. 2020. National education policy 2020. Technical report, Ministry of Human Resource Development, Government of India.
- Mohammad Khajah, Robert V. Lindsey, and Michael C. Mozer. 2016. How deep is knowledge tracing? In *Proceedings of the 9th International Conference on Educational Data Mining (EDM 2016)*, pages 94–101.
- Karthik Muralidharan, Abhijeet Singh, and Alejandro J. Ganimian. 2019. Disrupting education? Experimental evidence on technology-aided instruction in India. *American Economic Review*, 109(4):1426–1460.
- Sonali Nag. 2007. Early reading in Kannada: The pace of acquisition of orthographic knowledge and phonemic awareness. *Journal of Research in Reading*, 30(1):7–22.
- Benjamin D. Nye. 2015. Intelligent tutoring systems by and for the developing world: A review of trends and approaches for educational technology in a global context. *International Journal of Artificial Intelligence in Education*, 25(2):177–203.
- Shalini Pandey and George Karypis. 2019. A self-attentive model for knowledge tracing. In *Proceedings of the 12th International Conference on Educational Data Mining (EDM 2019)*, pages 384–389.
- Zachary A. Pardos and Neil T. Heffernan. 2010. Modeling individualization in a Bayesian networks implementation of knowledge tracing. In *Proceedings of the 18th International Conference on User Modeling, Adaptation, and Personalization (UMAP 2010)*, pages 255–266.
- Radek Pelánek. 2017. Bayesian knowledge tracing, logistic models, and beyond: An overview of learner modeling techniques. *User Modeling and User-Adapted Interaction*, 27(3):313–350.
- Chris Piech, Jonathan Bassen, Jonathan Huang, Surya Ganguli, Mehran Sahami, Leonidas J. Guibas, and Jascha Sohl-Dickstein. 2015. Deep knowledge tracing. In *Advances in Neural Information Processing Systems 28 (NIPS 2015)*, pages 505–513.
- Nicola J. Pitchford. 2015. Development of early mathematical skills with a tablet intervention: A randomized control trial in Malawi. *Frontiers in Psychology*, 6:485.
- Nicola J. Pitchford, Antonie Chigeda, and Paula J. Hubber. 2019. Interactive apps prevent gender discrepancies in early-grade mathematics in a low-income country in sub-Saharan Africa. *Developmental Science*, 22(5):e12864.
- Rocket Learning. 2024. How Rocket Learning is using WhatsApp to end learning poverty. <https://rocketlearning.org/how-rocket-learning-is-using-whatsapp-to-end-learning-poverty/>. Accessed June 2026.
- Hester Stubbé, Aiman Badri, Rebecca Telford, Anja van der Hulst, and Wouter van Joolingen. 2016. E-learning Sudan: Formal learning for out-of-school children. *Electronic Journal of e-Learning*, 14(2):136–149.
- Michael V. Yudelson, Kenneth R. Koedinger, and Geoffrey J. Gordon. 2013. Individualized Bayesian knowledge tracing models. In *Proceedings of the 16th International Conference on Artificial Intelligence in Education (AIED 2013)*, pages 171–180.
- Hengyuan Zhang, Zitao Liu, Shuyan Huang, Chenming Shang, Bojun Zhan, and Yong Jiang. 2024. Improving low-resource knowledge tracing tasks by supervised pre-training and importance mechanism fine-tuning. arXiv preprint arXiv:2403.06725.
- Jiani Zhang, Xingjian Shi, Irwin King, and Dit-Yan Yeung. 2017. Dynamic key-value memory networks for knowledge tracing. In *Proceedings of the 26th International Conference on World Wide Web (WWW 2017)*, pages 765–774.