

NIETE-UET Baseline Report

At a Glance: Key Learning Trends

Establishing a baseline of student learning is the first step in identifying where the greatest impact can be made through the program's interventions. The baseline data doesn't just set the stage—it sends a clear and urgent message: children in early grades in Balochistan are facing a foundational learning crisis.

The baseline data reveals a critical learning crisis across early grades in Balochistan. Despite comparable student demographics—each arm comprising approximately 700 students, with an average age of 10.5 years and a slightly higher proportion of girls—academic performance remains alarmingly low across core subjects:

- English is the most underperforming subject. Fewer than 1% of students across both control and treatment groups meet grade-level expectations. Grade 2 stands out as the most disadvantaged, with almost no students on target. This points to a foundational literacy gap that demands urgent and system-wide attention.
- Urdu, while comparatively better, still leaves a significant majority behind. The highest on-target performance (24% in Grade 4 treatment groups) indicates that even the strongest cohorts are far from acceptable proficiency levels.
- Math performance is consistently low and uneven. On-target percentages hover between 4% and 11%, with some promise in Grade 2 treatment students, but a noticeable drop-off in higher grades. This reflects a need for foundational reinforcement and curriculum-aligned support.

Across all subjects, Grade 2 students exhibit the lowest baseline learning levels, underscoring the need for intensive early-grade support. Meanwhile, Variations 12 of the intervention show slightly stronger starting points in English and Urdu, while Variation 4 marginally leads in Math. These initial differences will help guide the interpretation of future impact results and highlight the potential influence of specific intervention components.

Together, these trends form a clear call to action: if not addressed immediately, early-grade learning deficits will deepen over time. The findings reinforce the urgency of a robust, outcomes-based teacher professional development program—one that is capable of tackling foundational gaps and systematically improving student learning outcomes across Balochistan's public schools.

Background and Rationale

Balochistan's education system continues to struggle with low student learning outcomes, as evidenced by NAT 2023 results, where Grade 4 students scored just 49% in mathematics, 56% in English, and 68% in Urdu—significantly below national averages. These poor results are linked to systemic challenges such as the lack of continuous professional development (CPD) for teachers, a shortage of qualified educators, ghost schools and teachers, and inadequate school infrastructure. To address these issues, a comprehensive blended CPD program is proposed, utilizing high-quality video training modules aligned with the Single National Curriculum, user-friendly digital platforms for remote access, scripted lesson plans to standardize and improve classroom instruction, and on-site coaching by trained academic mentors. The program emphasizes outcomes-based design, ensuring every activity is tied to improvements in student learning, and will be implemented in phases to test multiple approaches for effectiveness and cost-efficiency. With strong monitoring and analytics guiding personalized mentorship, the initiative aims to build teacher capacity, foster consistency in instructional delivery, and ultimately boost foundational literacy and numeracy across the province. The end goal is to generate evidence on the most impactful and scalable model for long-term education reform in Balochistan.

This document serves as the baseline report, detailing the student learning outcomes at the initiation of the interventions. At this point, we are primarily showcasing the current academic levels of the students before the full impact of the program can be observed.

The baseline survey's primary objective was to evaluate student learning levels. The study's population consisted of all students enrolled in Grades 2 – 4 across the 77 selected schools in two districts of Balochistan. These schools were chosen because the Government of Balochistan's educational authorities granted permission for evaluation in Zarghoon and Chiltan districts. This selection ensured that the findings were relevant and generalizable to similar contexts within Balochistan.

Program Interventions and Design

At the start of the program, we tested a range of interventions aimed at improving student learning outcomes. These interventions, including teacher trainings, lesson plans, textbooks, coaching, assessments, and at-home learning, were systematically deployed across different intervention arms.

The randomized controlled trial is structured to evaluate the effectiveness and contribution of these different components of a blended teacher professional

development (CPD) model. The sample is divided into a control group and four treatment groups, each receiving a distinct combination of intervention components to allow comparative analysis across key variables such as textbooks, training, and coaching. These groups include:

- 17 control schools where no interventions were implemented.
- 60 treatment schools divided across four intervention arms.

The program prioritizes student learning outcomes as the ultimate measure of success, ensuring that all activities and interventions are aligned with this goal.

The table below outlines the specific interventions included or excluded in each variation of the program, providing a clear comparison of the components being tested:

Intervention	Variation 1	Variation 2	Variation 3	Variation 4	Control Group
Teacher Training	TRUE	TRUE	FALSE	TRUE	FALSE
Lesson Plans	TRUE	TRUE	TRUE	TRUE	FALSE
Textbooks	TRUE	FALSE	FALSE	FALSE	FALSE
Coaching	TRUE	TRUE	TRUE	FALSE	FALSE
Assessments	TRUE	TRUE	TRUE	TRUE	FALSE
At Home Learning	TRUE	TRUE	TRUE	TRUE	FALSE

Note: Due to the Balochistan Government's denial to publish the textbooks, intervention arms 1 and 2 became exactly the same, effectively functioning as a single intervention arm for this report.

Hypotheses

The effectiveness of the proposed interventions is evaluated through a set of key hypotheses, designed to isolate the impact of individual program components and assess the overall model's efficacy. The following table outlines these hypotheses and the primary comparisons intended to address each:

Hypothesis	Insights Sought
How effective is our current model when deployed in Balochistan?	To assess the overall efficacy of the blended CPD model in the provincial context.

How much impact does teacher training add in the current model?	To identify the specific contribution of teacher training to student learning gains.
How much impact do we lose/gain if we remove coaches?	To understand the role of on-site coaching in instructional consistency and outcomes.
How much do children learn when not exposed to any one of our intervention arms?	To establish baseline learning levels in the absence of program interventions.

Study Population and Sampling Strategy

To ensure the feasibility of data collection while maintaining statistical rigor, the study adopted a fixed-sample-size approach based on available resources and operational considerations. A strategic trade-off was made between the number of students per class and statistical power. By reducing the number of participating classes and increasing the number of students sampled within each class, we aimed to enhance the precision of estimates while working within logistical constraints.

To select students for the sample, a stratified random sampling approach was employed, ensuring representation across schools and accounting for key variables such as grade level and gender. This method ensured that the sample accurately reflected the diversity of the student population within the selected districts, thereby increasing the generalizability of the findings. The random selection within strata minimized selection bias and enhanced the statistical validity of the baseline assessment.

Given these constraints, data were successfully collected from 3,576 students across all intervention and control groups. A fixed sampling approach was used at the school level, wherein data were collected from 47 students per school across Grades 2 to 4 (approximately 16 students per grade). Recognizing the impact of sample distribution on statistical power, adjustments were made to optimize study reliability: the number of grades included in the sample was reduced while the fixed number of students per grade increased. The revised sampling approach focused on collecting data from approximately 16 students each from Grades 2, 3, and 4, significantly improving the study’s statistical power while maintaining a manageable scope for data collection and analysis.

Analytical Approach

For this baseline report, the analysis will primarily involve two key approaches:

- **Descriptive Analysis:** This included an overview of the basic statistics like

frequencies and averages.

- **Comparative Analysis of Control and Treatment Groups:** We conducted a comparative analysis between the control group and the various treatment groups. This initial comparison highlighted any pre-existing differences in student learning outcomes between groups before the interventions were fully implemented. While this was a baseline, understanding these initial differences was crucial for future impact evaluations.

For both assessments (ASER Grades 1 – 3 and 4 – 5), benchmarks for each subject and grade were set as described below.

Targets for English:	Targets for Urdu:	Targets for Maths:
1: 100% at Words 2: 100% at G2 Sentence 3: 100% at G2 Sentence 4: 100% at G5 Sentence Meaning 5: 100% at Story	1: 100% at Words 2: 100% at Sentence 3: 100% at G2 Story Fluency 4: 100% at G5 Story Fluency 5: 100% at G5 Story Fluency	1: 100% at 3-digit recognition 2: 100% at Subtraction 3: 100% at G2 Division 4: 100% at G5 Division 5: 100% at G5 Division

Summary of Results:

Initial descriptive statistics for the sampled student population across the different intervention and control groups are presented below. This table provides an overview of the demographic characteristics at the baseline, including mean age and gender distribution within each group.

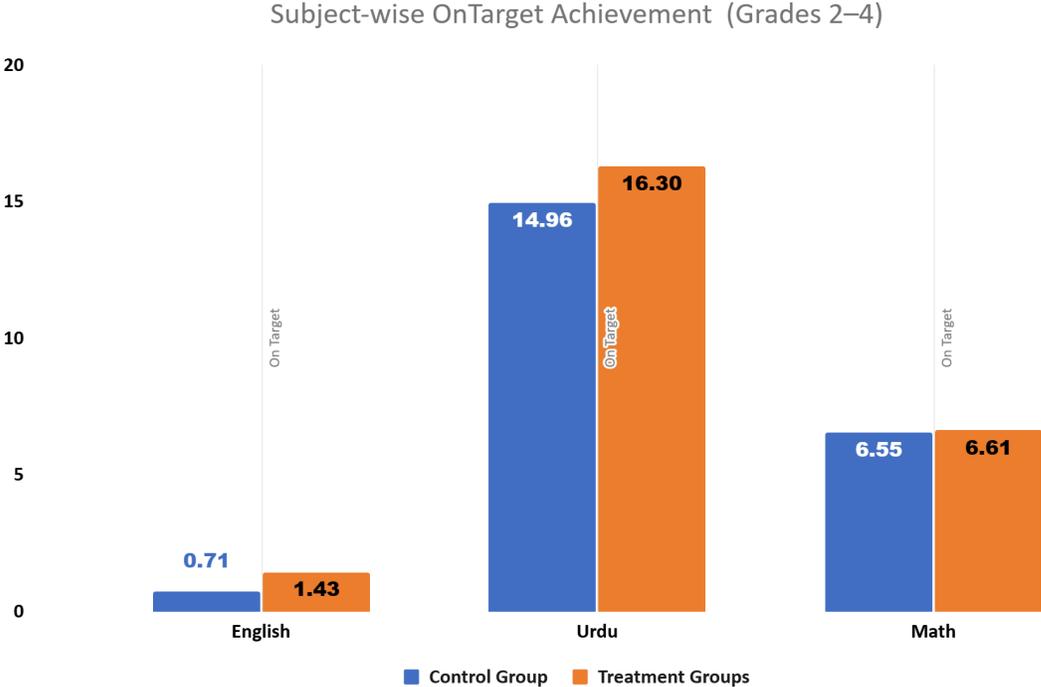
Group	Mean Age	Male	Female	Total Students
T1	10.6	371	355	726
T2	10.6	335	434	769
T3	10.8	233	461	694
T4	10.3	355	330	685
CG	10.5	261	441	702
Total	--	1555	2021	3576

The mean age across all groups is highly consistent, ranging from 10.3 to 10.8 years, indicating a uniform age distribution across the intervention and control arms. This consistency in mean age suggests that the randomization process was effective in creating groups that are comparable in terms of student age at the outset of the study.

In terms of gender distribution, the overall sample shows a higher representation of female students (2,021) compared to male students (1,555). This pattern is generally consistent across individual treatment groups (T1, T2, T3) and the Control Group (CG), with a slight deviation in T4 where male students marginally outnumber females. The total number of students per group is also relatively balanced, ranging from 685 to 769, ensuring sufficient sample sizes for comparative analyses. These baseline demographics confirm the suitability of the groups for robust comparative analysis, as they exhibit similar characteristics before the interventions are fully implemented.

Comparative Analysis

This section presents the baseline performance data for student learning outcomes, comparing the control group with the various treatment arms across key subjects. These comparisons highlight the initial state of student proficiency prior to the differentiated interventions.



At baseline, student performance against grade-level targets is generally low across all subjects: in English, barely 1 % of learners in either cohort meet the benchmark (0.7 %

in control schools and 1.4 % in schools slated for treatment), highlighting a critical gap in foundational language skills. Urdu shows the strongest starting point, yet still only about one in five students in treatment schools (16.3 %) and one in seven in control schools (15 %) achieve the expected level, indicating room for significant improvement even in the best-performing subject. Math sits in the middle but is uniform across groups, with just 6.6 % of students on target in both. These figures establish a clear intervention baseline: large learning deficits exist overall, treatment schools begin with a modest advantage in Urdu, and English emerges as the area most urgently in need of support.

Subject Wise Performance Percentage by Intervention Arms

		Variation 1 & 2	Variation 3	Variation 4	Control Group
English	Off Target	98.0	99.4	99.0	99.3
	On Target	2.0	0.6	1.0	0.7
Urdu	Off Target	82.0	84.6	86.3	85.0
	On Target	18.0	15.4	13.7	15.0
Math	Off Target	93.6	93.7	92.7	93.5
	On Target	6.4	6.3	7.3	6.6

At a glance, the baseline picture is stark: almost every child is below grade-level expectations in all three subjects. In English, off-target rates hover near total—98 % in Variations 1 & 2, 99.4 % in Variation 3, 99 % in Variation 4, and 99.3 % in the control group—leaving, at best, just 2 % of students on target. Urdu fares a little better but still keeps the majority behind: on-target performance peaks at 18 % (Variations 1 & 2), slips to 15.4 % (Variation 3) and 13.7 % (Variation 4), and lands at 15 % in control schools; roughly four out of five learners remain off target. Math sits squarely in the middle, with on-target rates tightly bunched—6.4 % (Variations 1 & 2), 6.3 % (Variation 3), 7.3 % (Variation 4), and 6.6 % in the control group—meaning over nine in ten students still miss the benchmark. Taken together, these numbers confirm sizable learning deficits before any programme work begins: English demands urgent, system-wide attention; Urdu shows the strongest foothold yet still leaves most pupils behind; and Math, while slightly better than English, remains a steep climb for nearly every child.

Imagine a classroom of 100 students. When we check how many already meet the grade-level bar before any programme starts, the numbers are eye-opening. In English, only one or two children out of every hundred hit the mark—whether they’re in our future treatment schools (about 1–2 % on target) or in the control schools (less than 1 %). Urdu looks a little brighter: roughly 15–18 pupils in treatment schools and 15 in control schools are where they should be, but that still leaves more than four-fifths of the class needing help. Math sits in between: about 6–7 students in every hundred reach the target everywhere, meaning the other 90-plus are behind. In short, English demands urgent attention, Urdu offers the strongest—though still modest—starting point, and Math shows a consistent, low baseline across all groups. These figures set a clear, shared starting line for everyone involved: plenty of ground to cover and a big opportunity for the forthcoming intervention to make a visible difference.

Baseline Performance by Grade, Subject, and Group – % of Students On-Target vs Off-Target:

Grades		Grade 2		Grade 3		Grade 4	
		Control	Treatment	Control	Treatment	Control	Treatment
English	Off Target	100%	99.4%	100%	98.1%	97.6%	98.3%
	On Target	0%	0.6%	0%	1.9%	2.4%	1.7%
Urdu	Off Target	91.8%	92.0%	79.0%	83.5%	84.0%	75.7%
	On Target	8.2%	8.0%	21.0%	16.5%	16.0%	24.3%
Math	Off Target	92.1%	89.1%	94.2%	96.0%	94.1%	95.1%
	On Target	7.9%	10.9%	5.8%	4.0%	5.9%	4.9%

The table shows, for every subject, how many children are already hitting the grade-level bar (on-target) and how many still need a lift (off-target).

What jumps out

- English: almost everyone is struggling. In Grade 2 not a single control-group child and only six in every thousand treatment-group children meet the target. Grades 3 and 4 inch up to roughly 1–2 % on-target, but the story stays the same: 97–100 kids out of 100 need serious help with English across the board.

- Urdu: small islands of strength. Grade 2 starts modestly (about 8 % on-target in both groups). Grade 3 flips the script—one in five control-group children makes the cut, while treatment slips to one in six. By Grade 4, treatment catches a gust and climbs to one in four, the best number anywhere in the dataset.
- Math: low and uneven foundations. Grade 2 shows the best Math footing (about 11 % of treatment and 8 % of control students on-target). Grades 3 and 4 sink back to roughly 4–6 % across both groups, leaving nine-tenths or more of every class still below par.

Across the three core subjects, Grade 2 has the lowest share of children already on-target: practically none in English ($\approx 0\%$), only about 8 % in Urdu, and roughly 9 % in Math. Grades 3 and 4 inch ahead—averaging around 8–9 % on-target—but they still face large gaps. A nuance worth noting: Grade 3 is the weakest specifically in Math ($\approx 5\%$ on-target), while Grade 4 edges slightly ahead in both Urdu and English. Taken as a whole, though, Grade 2 learners are starting from the greatest disadvantage and will need the most intensive early support.