**Taleemabad Evaluation by CERP Analytics**

Table of contents

1. Introduction
2. Taleemabad’s Evaluation  
   2.1. Instrument and Data Collection  
   2.2. Sampling and Measurement  
   2.3. Summary of Results  
   2.3.1. ASER Grades 1 - 3  
   2.3.2. ASER Grades 4 - 5  
   2.3.3. MELQO
3. Analytics by CERP  
   3.1. School level Master Dataset and Variables  
   3.2. Balance Test  
   3.3. Results  
   3.3.1. ASER Grades 1 - 3  
   3.3.2. ASER Grades 4 - 5  
   3.3.3. MELQO
4. Conclusion

**1. Introduction**

Taleemabad, an EdTech learning platform, is dedicated to creating educational content aligned with national standards to enhance student outcomes. The platform caters to both students and teachers through a proprietary learning platform that integrates content delivery, student assessment, and engagement with parents and school administration. The Learning Management System (LMS) is offered for licensing to small-scale entrepreneurs who facilitate the enhancement of existing school operations with effective pedagogical methods and administrative tools. The platform’s content is accessible through a standalone Android app and is also broadcasted on TV and radio.

Taleemabad has collaborated with CERP for an external validation of their internal assessments. For this purpose, CERP has carried out an audit using the existing data that Taleemabad has collected. The following sections delve deeper into the specifics of both evaluations, highlighting their methodologies, analyses, and key differences. This comprehensive analysis provides valuable insights into the strengths and limitations of the approaches used and ultimately contributes to a more robust assessment of Taleemabad’s impact on educational outcomes.

**2. Taleemabad’s Evaluation**

The primary objective of Taleemabad evaluation was to evaluate the effectiveness of its offerings. The evaluation was conducted through third-party standardized assessments of student learning outcomes and class observations. The assessment and observation processes extended to both Taleemabad schools and comparable “control” schools. Currently encompassing 140 schools, Taleemabad employs a random sampling method to randomly select schools for these assessments and observations. In contrast, the control schools are deliberately chosen based on their close geographical proximity to the treatment schools.

**2.1. Instrument and Data Collection**

Taleemabad utilized established assessment instruments to evaluate learning outcomes across different grade levels. For grades 1-3 and 4-5, the ASER (“Annual Status of Education Report”) instrument was employed, which rigorously tests foundational skills in English, local languages (Urdu), and Mathematics. These instruments are widely recognized and validated for their application in low and middle-income countries.

In the pre-primary grades, Taleemabad employed MELQO, a UNESCO-led global initiative aimed at measuring pupils’ development and learning at the commencement of primary school, along with assessing the quality of pre-primary learning environments. MELQO assesses pre-numeracy, pre-literacy (English), and motor skills. Pre-numeracy evaluates a child’s understanding of basic mathematical concepts such as counting and number recognition. Pre-literacy (English) assesses foundational reading and writing skills, including letter recognition and sounds. Additionally, fine motor skills are observed through pupils’ ability to replicate shapes presented on paper, indicating promising levels of skill development.

**2.2. Sampling and Measurement**

The evaluation involved a comprehensive data collection process conducted at baseline during October-November 2022 and at endline during April-May 2023. This timeframe corresponds with the commencement of the fall term for baseline and the conclusion of the academic year for endline. Taleemabad employed a two-step sampling strategy, initially selecting schools randomly in the experimental group and choosing nearby non-Taleemabad subscribing schools as controls.

During baseline, approximately 30 experimental schools were randomly chosen, but not all of them participated in the endline data collection due to exits from the Taleemabad program. These schools were replaced by comparable ones for the endline assessment. Consequently, the evaluation focused on differences between the treatment and control groups at both baseline and endline, presenting these differences in terms of standard deviations of the control group, and later utilizing the metric of Learning Adjusted Years of Schooling (LAYS). The assessments, specifically ASER administered to grades 1-3 and 4-5, utilized a consistent metric across waves as detailed below:



Table 1: Metric used for analysing ASER Grades 1 - 3 and ASER Grades 4 - 5

**2.3. Summary of Results**

**2.3.1. ASER Grades 1-3:**

During baseline, the difference between Taleemabad and control students is statistically insignificant from zero for English and Mathematics. For Urdu, however, the difference was -13.6% between Taleemabad and control. At the endline, the difference between Taleemabad and control schools rises to 22.9% for English and 14.6% for Urdu (both significant at 5% level). For Mathematics, however, the difference is -2.9% (i.e. in favor of control schools, significant at 5% level). Expressed in standard deviations, Taleemabad students are 0.46 of a standard deviation above control schools for English and 0.29 standard deviations for Urdu, whereas the control schools are 0.07 of a standard deviation above Taleemabad schools at the endline.

For reference, between 2019 and 2021, there was a decline in learning outcomes nationally equivalent to 0.06 standard deviations averaged across subjects for grades 1-3 (~0.03 sd/year). Pre-pandemic (2015 vs 2018) this stood at 0.17 (~0.06 sd/year) respectively although we were starting at a much lower baseline in terms of learning levels.

**2.3.2. ASER Grades 4-5:**

During baseline, the difference between Taleemabad and control students is statistically insignificant from zero for English and Mathematics. For Urdu, however, the difference was -8.7% between Taleemabad and control, meaning that control schools were at a higher level than intervention schools. At the endline, the difference between Taleemabad and control schools rises to 3.6% for English, 1.5% for Mathematics and 2.2% for Urdu (all significant at 5% level). Expressed in standard deviations, Taleemabad students are 0.11 of a standard deviation above control schools for English, 0.03 for Mathematics and 0.05 standard deviations for Urdu.

For reference, between 2019 and 2021, there was a decline in learning outcomes nationally equivalent to 0.07 standard deviations averaged across subjects for grades 4-5 (~0.035 sd/year). Pre-pandemic (2015 vs 2018) this stood as a gain of 0.18 (~0.06 sd/year) although we were starting at a much lower baseline in terms of learning levels.

**2.3.3. MELQO:**

For MELQO, the difference between average scores obtained by treatment and control schools for pre-numeracy, pre-literacy and motor skills is calculated. A Threshold such as one for ASER is not imposed due to a lack of available guidelines. At baseline, Taleemabad schools outperformed control schools by ~6 percent of overall score in pre-numeracy but lagged behind control schools in pre-literacy and motor skills by 0.35% and 1.58% respectively. At the endline, the difference between Taleemabad and control schools reduced to ~4% in pre-numeracy whereas Taleemabad schools performed better than control schools in pre-literacy and motor skills by 1.6% and 3.4% respectively. These results are statistically significant at the 5% level. Expressed in standard deviations, the difference between Taleemabad and control schools is 0.16, 0.11 and 0.17 for pre-numeracy, pre-literacy and motor skills respectively

**3. Analytics by CERP:**

**3.1. Data Management**

Taleemabad provided six student-level raw datasets which included data collected at baseline and endline for ASER grades 1 - 3, ASER grades 4 - 5, and MELQO. The data cleaning process of these datasets involved a series of systematic steps to enhance data quality and analytical reliability.

1. The first crucial step involved a thorough revision of variable names for bringing clarity and consistency across all six datasets. This meticulous effort significantly enhanced the usability of the data for subsequent analysis.
2. Additionally, the school names needed a rigorous cleaning to standardize them across all six datasets. The correction files provided to Taleemabad mapped various spellings and discrepancies in the school names. Once consistent correct names were created, schools were assigned a unique id which was used to create a master school-level dataset that incorporated information from all baseline and endline datasets for ASER grades 1 - 3, ASER grades 4 - 5, and MELQO.
3. Furthermore, the raw student-level datasets included student marks for each question in the ASER instrument. To facilitate analysis, categorical variables were generated using the ASER ranking methodology for the respective grade levels. These variables categorized test scores into predefined groups, representing various levels of proficiency in English, Urdu, and Maths. Subsequently, dummy variables were created for each proficiency category, indicating whether a student belonged to a specific proficiency level. Due to the lack of guidelines regarding thresholds for MELQO, the percentage scores obtained by students for pre-numeracy, pre-literacy and motor skills are calculated.

Following the cleaning of endline and baseline data for ASER grades 1 - 3, ASER grades 4 - 5 and MELQO, the individual datasets were collapsed on school to create school level datasets. Consequently, dummy variables that were previously created for each proficiency category now indicate the proportion of students in each school belonging to specific proficiency levels. Similarly, pre-numeracy, pre-literacy and motor skills indicators now represented average percentage scores at school level. The baseline and endline datasets for ASER grades 1 - 3, ASER grades 4 - 5 and MELQO were merged on the basis of schools to create a school level master dataset.

Sine not all schools during the baseline assessment were included in the endline, and new schools were introduced as replacements, a matching variable was created to categorize whether a school was present in the baseline, the endline, or both. Schools present in both baseline and endline assessments were labelled as “Consistent”. Those exclusively present in the baseline were labeled as “Attrition,” signifying their absence in the endline. Conversely, schools introduced in the endline were categorized as “Replacement.”

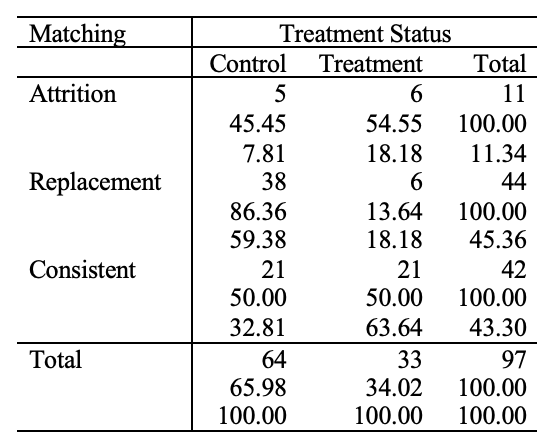


Table 2: Categories of schools by Treatment status

The table provided above demonstrates the distribution of schools categorized as Consistent, Attrition, and Replacement within both the treatment and control groups. The first row has frequencies; second row has row percentages and third row has column percentages. There are 97 unique schools in the master dataset, out of which 11 schools dropped after baseline and 44 schools were added as replacements at the endline stage. Out of 44 replacement schools, six schools were added to replace that the schools that were dropped after baseline. However, the number of replacement control schools is much higher due to two reasons. Firstly, when a treatment school dropped after baseline then nearby control school(s) were also introduced as replacements in addition to a treatment school. Secondly, ambiguous school names in the endline datasets which could not be corrected were consequently renamed as unknown and added to the control group. Both of these decisions were taken by Taleemabad.

Furthermore, three schools were initially enrolled in the Taleemabad program at baseline but withdrew from the program but agreed to undergo the endline assessment. Hence, their treatment status changed from treatment in baseline to control in endline. Specifically, these schools have been regarded as control schools in both baseline and endline.

In the full sample, there are 97 unique school out of which 33 are treatment schools and 64 are control schools. The truncated sample, which consists of Consistent schools only, contains 21 treatment schools and 21 control schools.

**3.2. Balance Test**

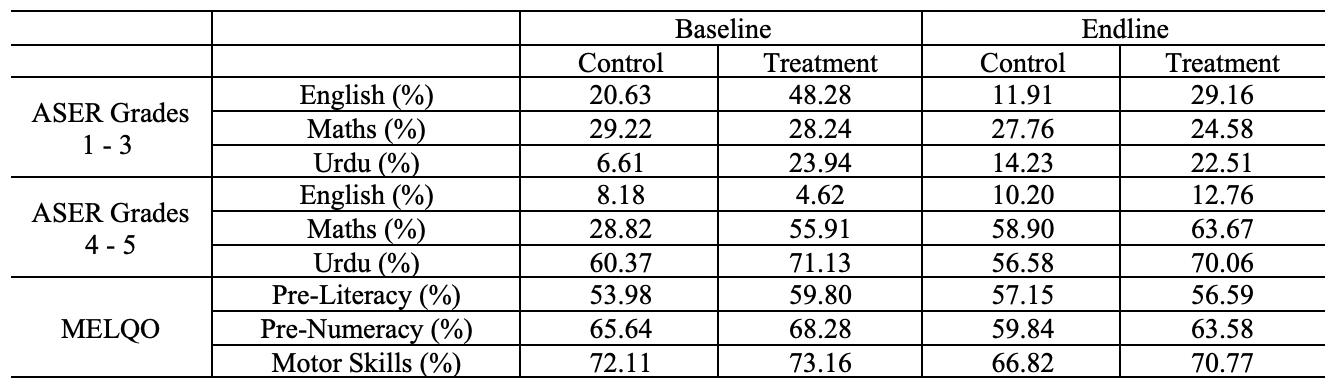


Table 3: Balance Test

The above table depicts whether the treatment and control groups are balanced at baseline and endline in terms of the outcome variables.

**3.3. Results**

This study evaluates the impact of the Taleemabad program on the student learning outcomes by comparing the mean differences in scores, in percentages, between the treatment and control groups at baseline and endline, and testing whether these differences are statistically significant. The following results pertain to the full sample of 97 schools.

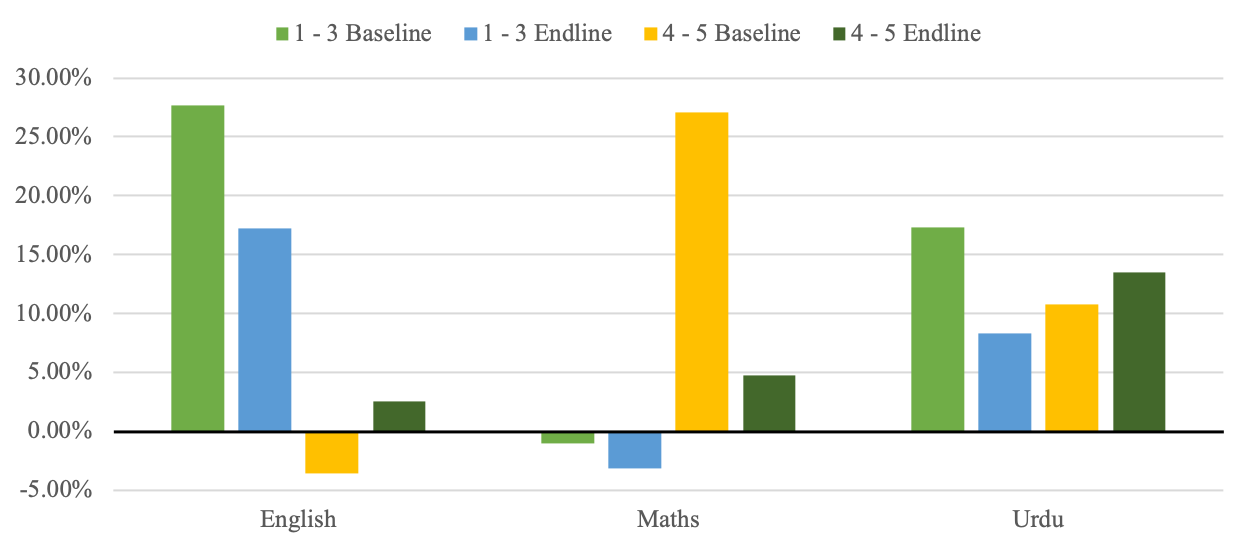


Figure 1: Percentage Difference in Grade Threshold Clearing Between Treatment and Control Schools at Baseline and Endline using ASER Instruments

**3.3.1. ASER Grades 1 - 3**

At baseline, the percentage difference between treatment and control schools for English is 27.65% which is statistically significant at the 5% level. At endline, the percentage difference between treatment and control schools for English is 17.25% which is statistically significant at 5%. Expressed in standard deviations, treatment schools are 1.073 standard deviations above control schools for English at endline.

At baseline, the percentage difference between treatment and control schools for Maths is -.98% which is statistically insignificant from zero. At endline, the percentage difference between treatment and control schools for Maths is -3.17% which is statistically insignificant from zero. Expressed in standard deviations, treatment schools are .161 standard deviations below control schools for Maths at endline.

At baseline, the percentage difference between treatment and control schools for Urdu is 17.33% which is statistically significant at the 5% level. At endline, the percentage difference between treatment and control schools for Urdu is 8.285% which is statistically insignificant from zero. Expressed in standard deviations, treatment schools are .366 standard deviations above control schools for Urdu at endline.

**3.3.2. ASER Grades 4 - 5**

At baseline, the percentage difference between treatment and control schools for English is -3.56% which is statistically insignificant from zero. At endline, the percentage difference between treatment and control schools for English is 2.55% which is statistically insignificant from zero. Expressed in standard deviations, treatment schools are .156 standard deviations above control schools for English at endline.

At baseline, the percentage difference between treatment and control schools for Maths is 27.09% which is statistically significant at the 5% level. At endline, the percentage difference between treatment and control schools for Maths is 4.77% which is statistically insignificant from zero. Expressed in standard deviations, treatment schools are .18 standard deviations above control schools for Maths at endline.

At baseline, the percentage difference between treatment and control schools for Urdu is 10.76% which is statistically insignificant from zero. At endline, the percentage difference between treatment and control schools for Urdu is 13.48% which is statistically insignificant from zero. Expressed in standard deviations, treatment schools are .376 standard deviations above control schools for Urdu at endline.

**3.3.3. MELQO**

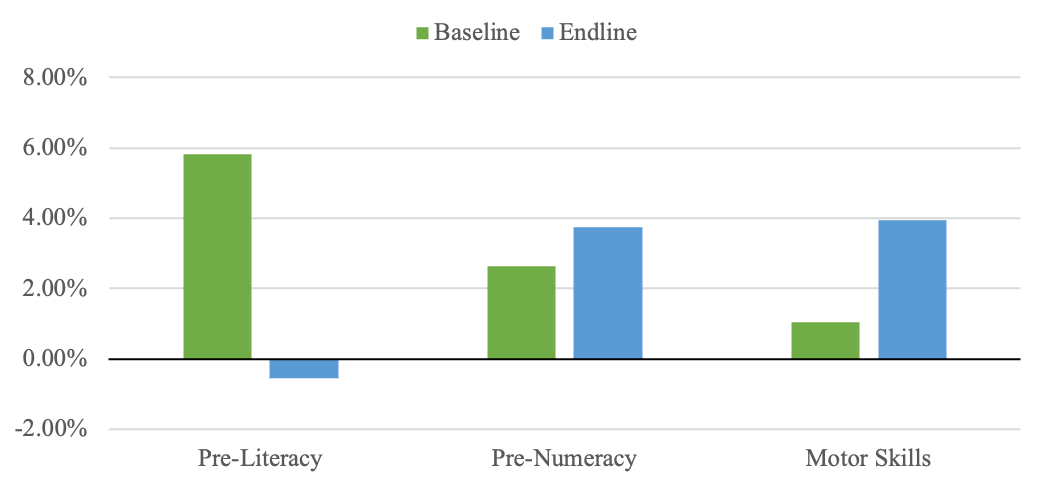


Figure 2: Percentage Difference in Scores between Treatment and Control Schools at Baseline and Endline using MELQO

At baseline, the percentage difference between treatment and control schools for pre-literacy is 5.81% which is statistically insignificant from zero. At endline, the percentage difference between treatment and control schools for pre-literacy is -.55% which is statistically insignificant from zero. Expressed in standard deviations, treatment schools are .038 standard deviations below control schools for pre-literacy at endline.

At baseline, the percentage difference between treatment and control schools for pre-numeracy is 2.64% which is statistically insignificant from zero. At endline, the percentage difference between treatment and control schools for pre-numeracy is 3.74% which is statistically insignificant from zero. Expressed in standard deviations, treatment schools are .203 standard deviations above control schools for pre-numeracy at endline.

At baseline, the percentage difference between treatment and control schools for motor skills is 1.05% which is statistically insignificant from zero. At endline, the percentage difference between treatment and control schools for motor skills is 3.95% which is statistically insignificant from zero. Expressed in standard deviations, treatment schools are .288 standard deviations above control schools for pre-numeracy at endline.

**4. Conclusion**