**Taleemabad Evaluation by CERP Analytics**

Table of contents

1. Background
2. Taleemabad’s Evaluation  
   2.1. Instruments 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. Analyses by CERP  
   3.1. Data Management  
   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  
   3.4. Results using Truncated Sample  
   3.4.1. ASER Grades 1 - 3  
   3.4.2. ASER Grades 4 - 5  
   3.4.3. MELQO
4. Limitations of the design
5. Conclusion

**1. Background**

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.

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.

**2. Taleemabad’s Evaluation**

The primary objective of the evaluation conducted by Taleemabad was to discern the effectiveness of its offerings. The evaluation was conducted through third-party standardized assessments of student learning outcomes and class observations on both Taleemabad schools and comparable “control” schools.

**2.1. Instruments and Data Collection**

Taleemabad administered the assessments using globally renowned 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. Lastly, motor skills are observed through pupils’ ability to replicate shapes presented on paper, highlighting their levels of skill development.

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.

**2.2. Sampling and Measurement**

Taleemabad employed a two-step sampling strategy. First, using a random sampling method, approximately 30 schools were chosen at random for the experimental group from Taleemabad’s extensive network of 140 schools. This was then followed by a selection of nearby non-Taleemabad schools as controls. The control schools are deliberately chosen based on their close geographical proximity to the treatment schools that are subscribed to the Taleemabad program.

Some of the experimental schools that were randomly chosen at baseline existed the Taleemabad program. Hence, these schools were replaced by comparable schools 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:**

Overall, 50.7% of all Taleemabad students met the criteria for all 3 subjects compared to 29.4% of control students at endline. This indicated a difference of 0.46 standard deviations compared to the control group (1.58 LAYS).

During baseline, the difference between treatment and control schools is statistically insignificant from zero for English and Mathematics. For Urdu, however, the difference was -13.6% between treatment and control schools. At the endline, the difference between treatment 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.

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

Overall, 7.2% of all Taleemabad students met the criteria for all 3 subjects compared to 6.2% of control students at endline. For 2 subjects, this stood at 49.3% for Taleemabad and 46.8% for control. This indicated a difference of 0.05 standard deviations compared to the control group (0.4 LAYS).

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.

**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 the 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.

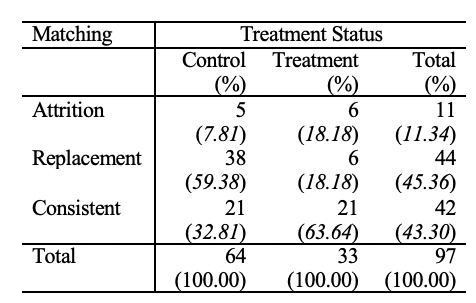
**3. Analyses by CERP:**

**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 to bring 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 provided by Taleemabad, schools were assigned unique ids to help create a master school-level dataset for ASER grades 1 - 3, ASER grades 4 - 5, and MELQO that would consolidate information from each respective baseline and endline datasets.
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. Since there are no similar thresholds for MELQO, the percentage scores obtained by students for pre-numeracy, pre-literacy and motor skills are calculated.
4. IV. IV. 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 schools 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 the school level. The baseline and endline school level 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.

Since not all schools during the baseline assessment were included in the endline, and new schools were introduced as replacements at the endline, 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*

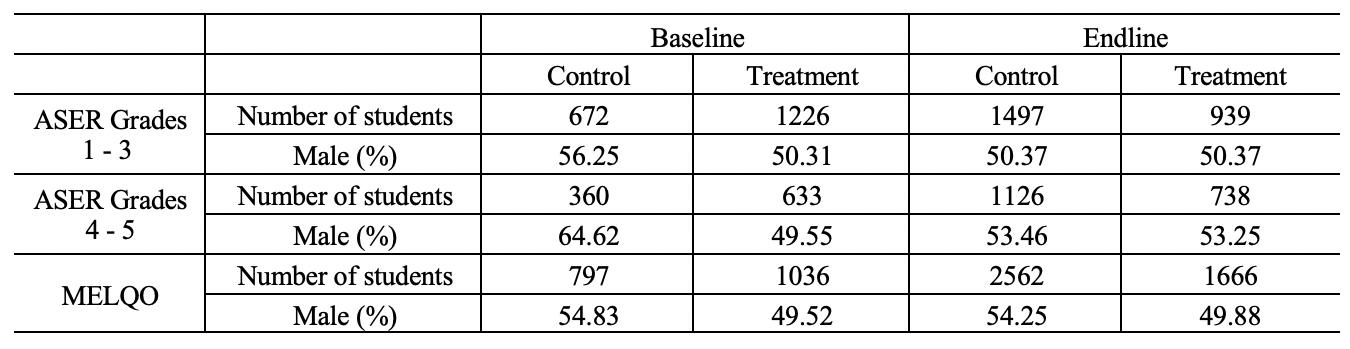
Table 2 provided above demonstrates the distribution of schools categorized as Consistent, Attrition, and Replacement, across treatment and control groups. The first row is the number of schools whereas the respective column percentage is shown in parenthesis in the second row. There are 97 unique schools, out of which 11 schools dropped after baseline while 44 schools were added as replacements at the endline. Out of 44 replacement schools, there are only six schools that were added to replace the schools dropped after baseline.

Based on the guidelines provided by Taleemabad, the number of replacement control schools were observed to be higher due to two main reasons. Firstly, when a treatment school dropped after baseline then it is likely that more than one nearby control school was introduced as a replacement for that 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.

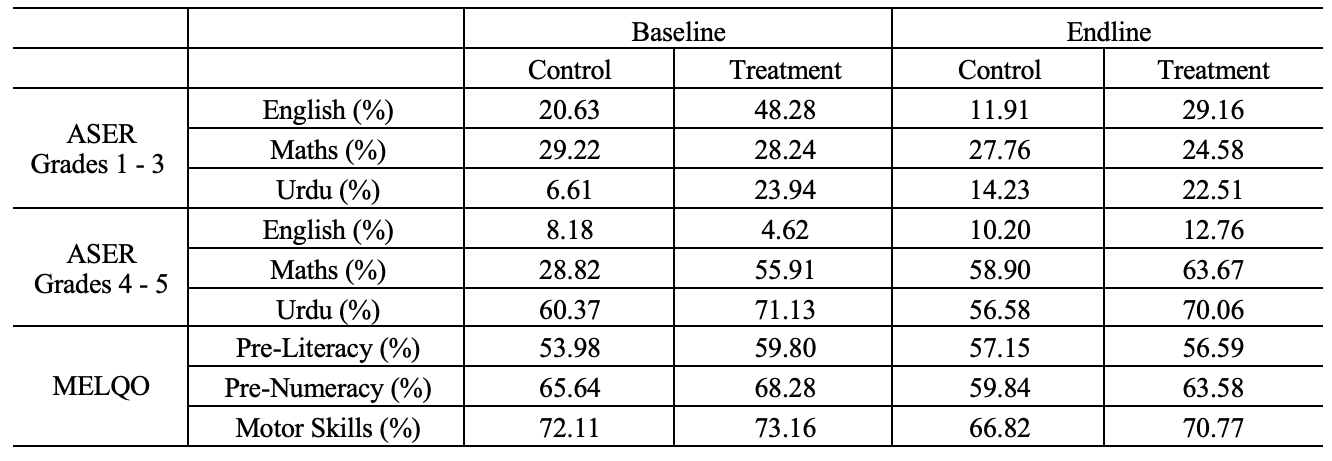
Furthermore, three schools initially enrolled in the treatment group withdrew from the Taleemabad program after baseline but agreed to undergo the endline assessment. Hence, during analysis by the CERP team, their treatment status was preserved as “control” for both baseline and endline.

In the full sample, there are 97 unique schools, out of which 33 are treatment schools and 64 are control schools. The truncated sample, which consists of only Consistent schools (i.e. schools present at both baseline and endline), contains 21 treatment schools and 21 control schools.

**3.2. Balance Test**



*Table 3.1: Balance Test using student demographics*



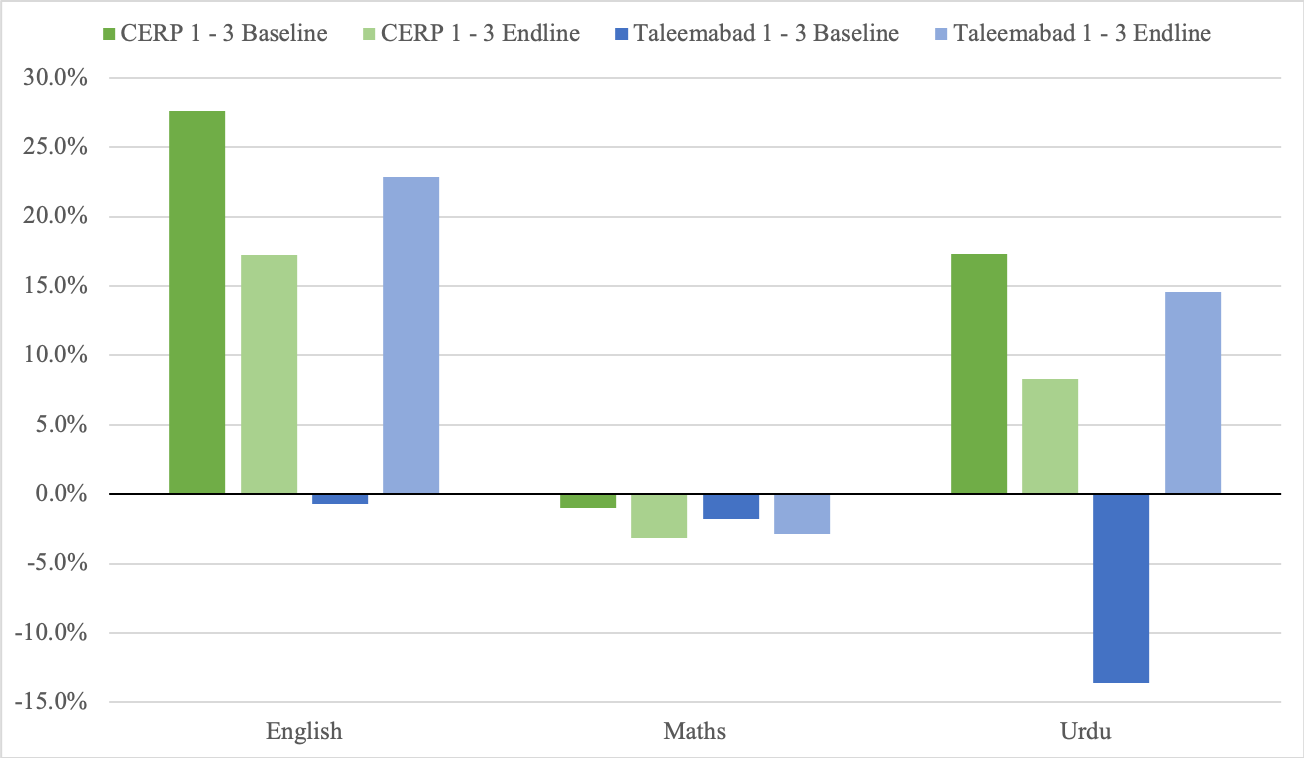
*Table 3.2: Balance Test using outcome variables*

Table 3.1 and Table 3.2 demonstrate a lack of balance between treatment and control groups at baseline. Notably, differences in the number of students, proportion of male students, and outcome variables are evident. This imbalance can be attributed to the non-random selection of control schools currently done based on proximity to treatment schools.

**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. The figures plot results by Taleemabad and CERP for comparison.

**3.3.1. ASER Grades 1 - 3**



*Figure 1: % Difference in Grade Threshold Clearing Between Treatment and Control Schools at Baseline and Endline using ASER Grades 1 - 3 instrument*

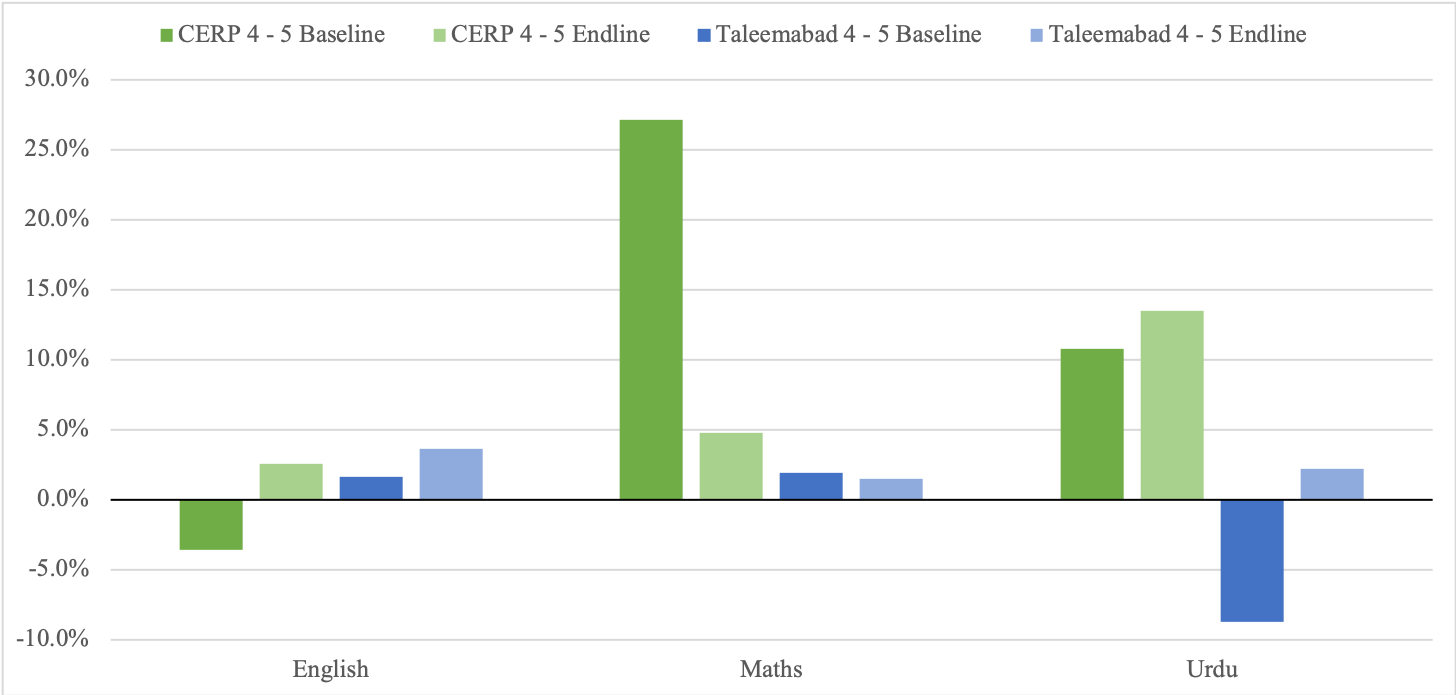
At endline, 2.54% of treatment schools met the criteria for all 3 subjects compared to 1.06% of control schools. This represents a difference of .452 standard deviations compared to the control group, which is equivalent to 3.48 LAYS.

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**



*Figure 2: % Difference in Grade Threshold Clearing Between Treatment and Control Schools at Baseline and Endline using ASER Grades 4 - 5 instrument*

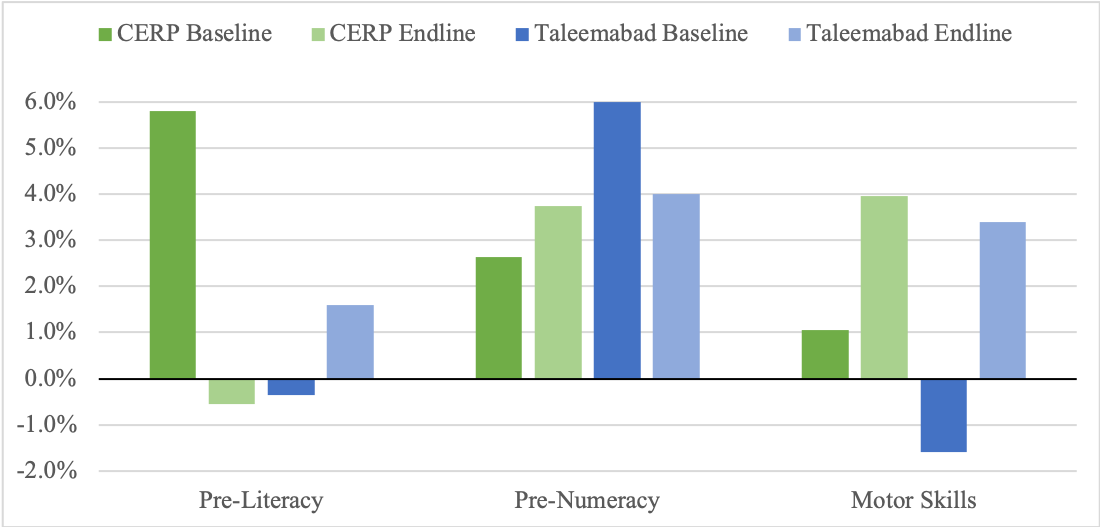
At endline, 6.77% of treatment schools met the criteria for all 3 subjects compared to 4.41% of control schools. This represents a difference of .234 standard deviations compared to the control group, which is equivalent to 1.8 LAYS.

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 3: % 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.

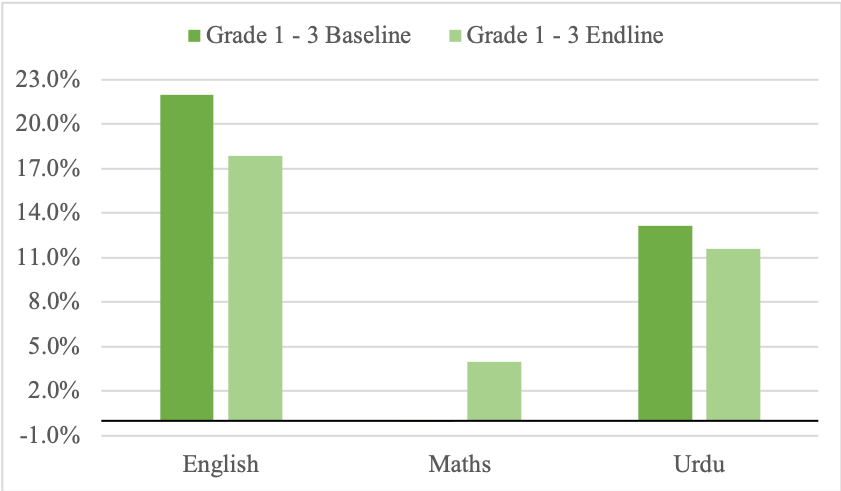
At baseline, the percentage difference between treatment and control schools for pre-numeracy is 2.64% which is statistically insignificant from zero. At endline, the percentage difference between treatment and control schools for pre-numeracy is 3.74% which is statistically insignificant from zero. 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 motor skills at endline.

**3.4. Results using Truncated Sample**

The following results pertain to the truncated sample of 42 consistent schools, with 21 schools part of the treatment group and an equivalent number part of the control group.

**3.4.1. ASER Grades 1 - 3**



*Figure 1: % Difference in Grade Threshold Clearing Between Treatment and Control Schools at Baseline and Endline using ASER Grades 1 - 3 instrument*

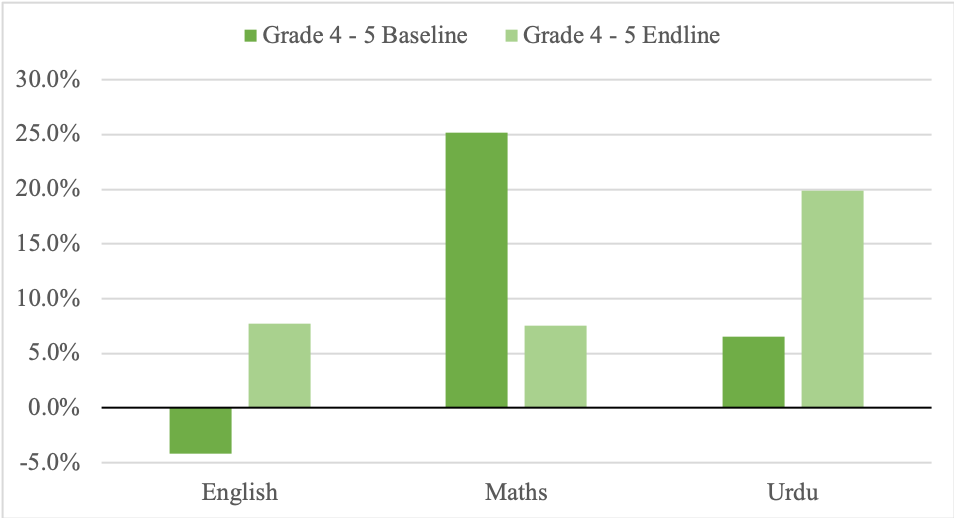
At endline, 3.02% of treatment schools met the criteria for all 3 subjects compared to 1.15% of control schools. This represents a difference of .582 standard deviations compared to the control group.

At baseline, the percentage difference between treatment and control schools for English is 22% which is statistically significant at the 5% level. At endline, the percentage difference between treatment and control schools for English is 17.85% which is statistically significant at 5%. Expressed in standard deviations, treatment schools are 1.375 standard deviations above control schools for English at endline.

At baseline, the percentage difference between treatment and control schools for Maths is -.06% which is statistically insignificant from zero. At endline, the percentage difference between treatment and control schools for Maths is 3.99% which is statistically insignificant from zero. Expressed in standard deviations, treatment schools are .277 standard deviations above control schools for Maths at endline.

At baseline, the percentage difference between treatment and control schools for Urdu is 13.18% which is statistically significant at the 5% level. At endline, the percentage difference between treatment and control schools for Urdu is 11.57% which is statistically insignificant from zero. Expressed in standard deviations, treatment schools are .572 standard deviations above control schools for Urdu at endline.

**3.4.2. ASER Grades 4 - 5**



*Figure 2: % Difference in Grade Threshold Clearing Between Treatment and Control Schools at Baseline and Endline using ASER Grades 4 - 5 instrument*

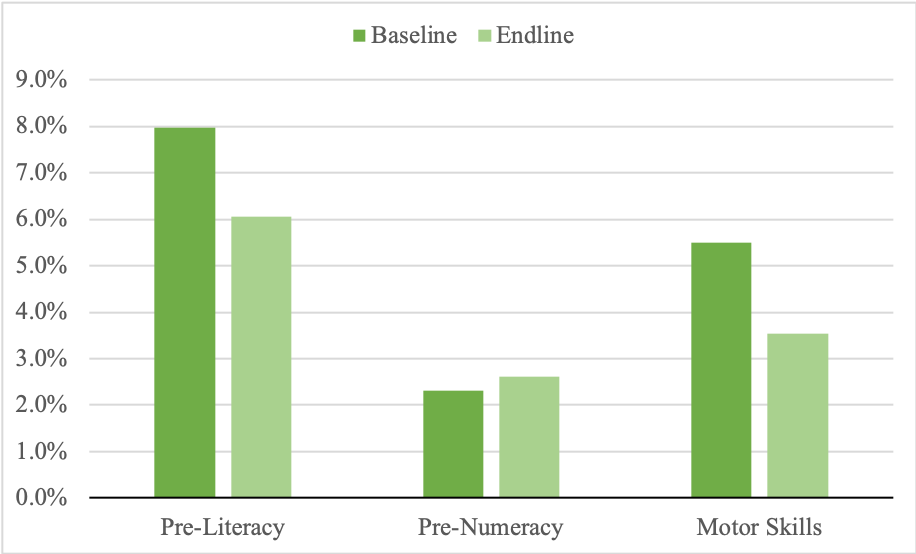
At endline, 7.66% of treatment schools met the criteria for all 3 subjects compared to 1.67% of control schools. This represents a difference of 1.614 standard deviations compared to the control group.

At baseline, the percentage difference between treatment and control schools for English is -4.18% which is statistically insignificant from zero. At endline, the percentage difference between treatment and control schools for English is 7.71% which is statistically insignificant from zero. Expressed in standard deviations, treatment schools are .905 standard deviations above control schools for English at endline.

At baseline, the percentage difference between treatment and control schools for Maths is 25.19% which is statistically significant at the 5% level. At endline, the percentage difference between treatment and control schools for Maths is 7.5% which is statistically insignificant from zero. Expressed in standard deviations, treatment schools are .271 standard deviations above control schools for Maths at endline.

At baseline, the percentage difference between treatment and control schools for Urdu is 6.55% which is statistically insignificant from zero. At endline, the percentage difference between treatment and control schools for Urdu is 19.85% which is statistically insignificant from zero. Expressed in standard deviations, treatment schools are .636 standard deviations above control schools for Urdu at endline.

**3.4.3. MELQO**



*Figure 6: % 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 7.96% which is statistically insignificant from zero. At endline, the percentage difference between treatment and control schools for pre-literacy is 6.06% which is statistically insignificant from zero. Expressed in standard deviations, treatment schools are .383 standard deviations above control schools for pre-literacy at endline.

At baseline, the percentage difference between treatment and control schools for pre-numeracy is 2.31% which is statistically insignificant from zero. At endline, the percentage difference between treatment and control schools for pre-numeracy is 2.62% which is statistically insignificant from zero. Expressed in standard deviations, treatment schools are .17 standard deviations above control schools for pre-numeracy at endline.

At baseline, the percentage difference between treatment and control schools for motor skills is 5.5% which is statistically insignificant from zero. At endline, the percentage difference between treatment and control schools for motor skills is 3.53% which is statistically insignificant from zero. Expressed in standard deviations, treatment schools are .26 standard deviations above control schools for motor skills at endline.

**4. Limitations of the design**

1. Inability to Track Test Scores for the Same Student at Baseline and Endline:

One of the significant limitations of the evaluation design is the inability to consistently track test scores for the same student at both baseline and endline assessments. This constraint hinders our capacity to conduct comprehensive child-level analyses, which could provide valuable insights into individual progress over the course of the intervention.

Additionally, due to the unavailability of test scores for the same student at baseline and endline, we are compelled to aggregate test scores at the school level for analysis. However, collapsing on the school level introduces a potential distortion in our findings as the average test scores at the school level may not accurately represent how individual children within the same school are actually performing. Variability in student performance is obscured, and the aggregated school-level analysis may provide a skewed representation of Taleemabad program's impact on educational outcomes.

1. Lack of Sufficient Data on External Factors:

The evaluation design faces a challenge due to the insufficient availability of data on external factors. The absence of comprehensive information on variables, such as socio-economic status, family support, or prior educational experiences, poses a substantial obstacle to effectively isolating the treatment effect from other potential influences.

For instance, without detailed data on socio-economic status, we may overlook how financial resources within households could independently impact a student's academic performance. If a student's improved outcomes align with an increase in parental involvement, discerning the isolated impact of the program becomes intricate. Household characteristics, including parental education levels or the availability of educational resources at home, could significantly influence a student's performance, but our current design lacks the granularity to account for these nuances.

Consequently, the observed differences in means cannot be conclusively attributed to the treatment alone. This limitation not only undermines the internal validity of our study but also restricts the extent to which causal inferences can be drawn. The complexity introduced by unaccounted household characteristics highlights the need for more comprehensive data collection strategies to strengthen the robustness of our analysis and enhance the validity of our conclusions.

1. Inherent Difficulty in Establishing Causation:

The design's limitations extend to the broader issue of establishing causation. The lack of a controlled environment and comprehensive data on confounding variables, such as teacher quality or school resources, introduces complexities in confidently attributing changes in outcomes solely to the implemented treatment. For instance, improved educational outcomes may be linked not only to the Taleemabad program but also to unaccounted variations in teaching methods or additional resources allocated to specific schools.

1. Potential for Selection Bias in Selection of Control Schools:

A notable limitation in our study arises from the process of selecting control schools based on proximity to treatment schools. This method, though practical, introduces the potential for selection bias. The proximity-based selection of control schools may inadvertently lead to differences in contextual factors between treatment and control groups. Local variations in community characteristics and educational resources may exist impacting the baseline comparability of the two groups. Variations in characteristics between treatment and control groups may confound our ability to attribute observed changes solely to the Taleemabad program.

**5. Conclusion**

There is lack of balance between treatment and control groups at baseline in terms of the number of students, proportion of male students, and outcome variables evaluated. This imbalance can be attributed to the non-random selection of control schools that is carried out based on proximity to treatment schools in this evaluation design.

By evaluating using the full sample of 97 schools, at endline, we observe better results in grades 1 - 3 and grades 4 -5 of the magnitude 0.452 standard deviations and 0.234 standard deviations respectively compared to the control group. This corresponds to approximately 3.48 years of Learning Adjusted Years of Schooling for grades 1 - 3 and 1.8 years of Learning Adjusted Years of Schooling for grades 4 - 5. Evaluating using the full sample, at endline, we observe better results in pre-primary for pre-numeracy and motor skills for treatment schools compared to control schools to the order 0.203 standard deviations and 0.288 standard deviations. In contrast, treatment schools are 0.038 standard deviations below control schools for pre-literacy at endline. Similar results are observed in the truncated sample of 42 schools. Out of the 21 consistent treatment schools, SRM School, DOA 12 and My School most often appeared in the top 20% best performing schools for each outcome variable.

The design of the evaluation study could be further strengthened by gathering data on relevant external factors (e.g., family support, socio-economic status, school-level resources) that would allow future analyses to test baseline comparability between treatment and control groups. Moreover, implementing unique student identifiers and consistent data recording procedures would not only minimize tracking errors but enable a deeper understanding of individual learning trajectories. Furthermore, if the schools in Taleemabad’s cohort can be included, if not already, as part of ASER or other evaluation studies that gather household, child-level, school-level or village-level information, it would enable more nuanced analysis of the program effectiveness.