

Supplementary Material S4

Reliability and Validity of the Data

Reliability of data analysis was facilitated by video recording all focus group discussions and personal interviews [47]. To further support reliability in data analysis, intercoder agreement was reported in terms of percentage agreement and Kappa reliability [47,48]. In order to maximize the validity of a qualitative study's findings, Creswell [47] presents eight strategies of which he prescribes any two to be sufficient for use. In our study, we employed four of three strategies - triangulation, member checking, and clarifying researcher bias. *Triangulation* entailed the use of diverse types of data from a variety of sources to support one's findings [47,50,167]. In our study, data from Focus groups discussions, personal interviews, parent emails, school timetables, the master curriculum and daily planning documents, the school's PD workshop content and website and classroom photographs were analyzed to extract findings related to the research question. *Member checking* required the researchers to take their analysis and interpretations back to the participants of the study to assess the credibility and accuracy of the findings [47,48,167]. We shared the final draft of the *findings* section with all focus groups and interview participants along with a Google Form that asked them two questions: (1) Do the findings of the researchers accurately match your experience? and (2) If you selected no above, please suggest which statements were inaccurate and suggest what they should be changed to. Participants were given 8 days to respond. Out of the total 12 participants, 8 did not respond, while 4 expressed that the findings section accurately reflected their experience and they did not suggest any edits.

Finally, with regard to *clarifying researcher bias*, this entailed that the researchers must explicitly declare their associations with the study site along with their personal biases, goals, orientations, positions on matters that might influence the study [47,48]. Two of the

researchers were leaders in the organization and were responsible for spearheading the administration of an international standardized curriculum based measure - DIBELSNext, to assess the foundational literacy skills of students across the school. They were surprised to note that a majority of the students showed a reading performance below grade level. These results initiated the introduction of a synthetic phonics curricular reform. The DIBELSNext battery of assessments were administered every year, and a steady and significant improvement was noted over the previous four years. This data caught the attention of the third researcher of our study. She was an Indian-origin professor at a university in the United States of America, and she specialized in the field of literacy and special education. Having already carried out prior research on the problem of foundational literacy skills in the context of Indian private schooling, she was interested in the significant improvement that students at ABC English Medium School had made on their foundational reading scores. She had been acquainted with the school leaders at ABC English Medium School previously due to another research project they had collaborated on. Finally, the fourth author was serving as a research executive on the teams of the two leaders at the school under study.