

Supplementary Material S3

Details Regarding the Coding Process

All four researchers engaged in independent coding of the focus group and personal interview data using an in vivo approach [49] to create a codebook of all major emergent descriptive labels from the data. The researchers then met to collaboratively discuss each other's codebooks and establish consensus over the final list of themes and codes. The researchers also included a brief description of each label in the codebook, to ensure that they were all defining the code in a consistent manner and would be able to conduct further analyses of the data objectively [49, 162]. Emergent codes with commonalities were organized under the three broad themes identified at the outset. For example, the codes of *daily dosage*, *backward design*, and *systematic ongoing assessment* all relate to some aspect of curriculum and hence were categorized under the theme of *curriculum*. If an emergent code did not correspond with any of the three a priori themes from the research literature, it was organized under a fourth theme titled 'Other Variables'. The codes identified under this theme were *supportive leadership*, *parent engagement*, and *change management*. The codebook was then used by Researcher 1 and Researcher 2 to independently analyze all focus group discussions and personal interview video recordings.

The two researchers who conducted the second cycle of coding were chosen because they were both actively engaged with helping ABC Schools introduce and scale up the phonics program at the time of the intervention, and hence possessed deeper context of the changes that the school had experienced over the years. Both researchers were asked to independently review all transcripts and assign text segments from the data to codes as per the codebook's guidelines, using NVivo software. A text segment was considered to be an entire statement or series of statements from an interview or focus group transcript that corresponded with one or more of the codes in the codebook. The researchers assigned every

single statement in the interview and focus group transcripts to a code, to ensure that they were both considering the entire pool of available data in their analysis.

Every time both researchers assigned a specific text segment to the same code, intercoder agreement was considered to be achieved. A mean inter-coder agreement of 97% and Kappa reliability value of 0.56 was found between Researcher 1 and Researcher 2, demonstrating less than satisfactory reliability [47, 52, 163]. Two major issues were identified as potential causes of low agreement between the researchers - *discriminant capability* and *unitization*. Discriminant capability refers to the extent to which coders can easily and objectively categorize excerpts of text from the interview transcripts under specific codes, using the codebook [164].

Given the high number of codes that existed in the first version of the codebook, and the inherent overlaps between similar codes, the likelihood of researchers disagreeing on which code a specific text segment fit best under was relatively high. In order to address this, researchers 1 and 2 met to reduce the number of codes and eliminate any potential overlaps between them. Similar codes were combined with each other to reduce the probability of misassignment. For example, the codes of *structured* and *goal-directed* under the theme of 'Curriculum' were combined given that they both attempted to address the process of designing the written curriculum in a cascading manner beginning with goals and moving on to assessments, instructional strategies, and resources. Further, the descriptive labels for each code were also revised in order to eliminate any overlaps with other codes. For example, the researchers added the statement "DOES NOT include demos, roleplays, or other sensemaking activities of curricular materials that are CONFIRMED already, NOT any matter related to reflection on student data etc." to the label for the code *collaborative decision making* in order to effectively differentiate it from the code for *collective sensemaking and reflection*.

Additionally, the issue of unitization refers to lack of clarity on what specific segments of text - or "*units of analysis*" [164] - are to be coded thereby leading to them choosing different portions of text during the coding process. For example, one coder may choose an entire paragraph to assign under a specific code while the other might only choose two sentences from it. This leads to lower levels of agreement even if both coders agree on which code to assign the text segment to. In order to address this issue, we followed the methodology suggested in recent literature [163,164] where the principal investigator demarcated specific segments of text to be separately coded before the coding process. This ensured that the units of analysis for both coders to follow were standardized, and the possibility of them choosing different portions of the transcript for code assignment was eliminated.

Once the problems of discriminant capability and unitization were addressed in the manner described above, the coders embarked on a second round of coding on NVivo using the revised codebook and transcripts with predetermined units of analysis. There was a significant increase in the inter-coder agreement, with the researchers achieving a mean Kappa score of 0.73 which can be regarded as substantial [52]. Despite the measures taken to improve reliability, there were still a few codes with less than satisfactory agreement. In order to address this, the researchers used a "negotiated agreement" approach [164] where specific instances of disagreement were discussed until the coders reached consensus about which code the said text segment would fit best under [165]. In instances where the coders initially disagreed but then mutually agreed with each other's rationale after discussion, they decided to assign the said text segment to both codes in question as the same statement/s would support inferences about more than one code/ theme.