**Supplemental Method: Measurement of Teacher Tone and Interactions**

In addition to direct observation of teachers’ procedural fidelity, we used a 5-point, Likert-type scale to rate the overall quality of teacher tone and interactions (1 = highly negative/punitive, 3 = neutral, 5 = highly supportive/encouraging) during sessions. Data collectors used examples of teacher behaviors to determine each rating (e.g., frequency and tone of praise, facial expressions). We hypothesized that differences in teacher behavior could correlate with student behavior (e.g., a student displaying higher engagement in a condition with higher quality tone and interactions). Jalen’s mean quality was 3.50 for paper sessions and 3.40 for CAI sessions, Koby’s mean quality was 3.80 for paper and 3.75 for CAI, and Berto’s mean quality was 5 for paper sessions and 4.50 for CAI sessions. Because these ratings were comparable, we concluded it was unlikely that differences in student behavior were related to the quality of teachers’ tone and interactions.