Supplemental Material S1. Detailed description of curriculum-based measures (CBMs) analyses.

For each CBM the multivariate distribution of the six repeated measures was examined using chi-square QQ plots and Mardia’s test for multivariate normality within each level of treatment. Chi-square QQ plots and Boxplots were examined to assess the severity of the departure from multivariate normality. The plots revealed that responses to one of the CBMs, Upper Case Letter Naming, had a pronounced ceiling effect (i.e., right censored), particularly for the measurements taken at later time points. Because the measures were counts of number correct, the square root transformation was considered, but the transformation had little success in producing a distribution closer to normal. So, for this variable, two models were fit: (1) a linear mixed effects model with a normal distribution for the response variable, and (2) a linear mixed effects model with a right censored normal distribution for the response variable. Both models are described below.

Bartlett’s test was used to assess homogeneity of covariance matrices across the two treatment groups. The null hypothesis of homogeneity of variance was rejected for each of the six CBMs, due primarily to the ceiling effect mentioned previously. Models for heterogeneity across groups were considered.

A linear mixed model using a normal distribution for the response variable was employed with fixed effects factors for treatment and time, random effects for school, and random coefficients at the child level. Child is crossed with time, but children are nested within classrooms and classrooms are nested within treatment level. Although some teachers taught an am and pm class, only one class was selected for inclusion in the research. Although the research included multiple schools, there were too few schools to consider an effect for classroom nested within school. The design is essentially a split plot with an added random factor for schools. The treatment factor had two levels, TELL vs BAU. The mixed model specified random coefficients for intercept and linear slope, so each child has his or her own growth trajectory, and the random intercept and slope terms also accommodate heterogeneous variance and covariance across time. The mixed model can be expressed in the following form:

$$y_{\mathrm{ijkl}} = \beta_{0i} + \beta_{1i} t_{\mathrm{ij}} + \beta_2 t_{\mathrm{ij}}^2 + u_{k(l)} + \beta_3 (x_{1\mathrm{ikl}} - \bar{x}_1) + \beta_4 x_{2\mathrm{ikl}} + \beta_5 t_{\mathrm{ij}} x_{2\mathrm{ikl}} + \beta_6 t_{\mathrm{ij}}^2 x_{2\mathrm{ikl}} + \varepsilon_{\mathrm{ijkl}}$$

where y_{ijkl} is an outcome measure for child *i* in classroom *k* at time *j*, $\beta_{0i} = \beta_0 + u_{0i}$, $\beta_{1i} = \beta_1 + u_{1i}$, $x_{1\mathrm{ikl}}$ is a fixed effect covariate such as mother’s education, x_{2ikl} is a dummy variable representing level of the treatment factor, $u_{k(l)}$ is a random effect for classroom *k* nested within treatment level *l*, with $u_{k(l)} N\left(0, \sigma_{u_k}^2\right)$, u_{0i} and u_{1i} as random effects for intercept and slope with $(u_{0i}, u_{1i})' N_2(0, G)$, and ε_{ijkl} as a random error term with $\varepsilon_{ijkl} N(0, \sigma^2)$ and independent of $(u_{0i}, u_{1i})'$ and $u_{k(l)}$.

Since the random intercept and slope terms were used to accommodate heterogeneity of variance across time, the independence structure, $\sigma_\varepsilon^2 I$, was selected for R_i, the covariance matrix of ε_i, the vector of error terms for the six measurements on child *i*. Variance components were estimated by the method of restricted maximum likelihood. The Kenward-Rogers (KR) method was used to calculate approximate degrees of freedom for model fixed effects. The likelihood ratio test and the Bayes Information Criterion (BIC) were used to compare models with heterogeneity of variance across treatment groups. The model fixed effects were estimated by the method of generalized least squares using SAS PROC MIXED. Measurements on PELI subtests were also analyzed by this method.

As mentioned above, due to the ceiling effect for Upper Case Letter Naming, a linear mixed effects model with a right-censored normal distribution for the response variable was employed (Vock, Davidian, & Tsiatis, 2011) for this variable. This approach is sometimes known as a tobit model. The mixed model for Y_{ijkl}, which would have been observed if there had been no censoring, is similar to the model specified earlier. However, due to censoring, we observe Q_{ijkl}, which takes on the value Y_{ijkl} for $Y_{ijkl} < l_{ij}$ and takes on the value l_{ij}, the known upper limit, otherwise. For each CBM, the known upper limit l_{ij} was the same for all time points and treatment levels. Variance components and fixed effects parameters were estimated by the method of maximum likelihood. The likelihood equation for the right-censored model is given in the following expression:

$$L(\theta, Q) = \prod_{i=1}^{N} \int \prod_{j=1}^{n_i} \left[\left\{1 - \Phi_1\left(m_{ijkl}\right)\right\}^{I\left(Q_{ijkl}=l_{ij}\right)} \left\{\frac{1}{\sigma} \phi_1\left(m_{ijkl}\right)\right\}^{I\left(Q_{ijkl}<l_{ij}\right)} \right] \varphi_q(u_i) du_i$$

where θ is a parameter vector containing the variance component parameters and the fixed effect parameters, Q is a vector containing all observed responses, N is the total number of children in the study, n_i is the number of observations on child i nested in kl, $m_{ijkl} = \left\{Q_{ijkl} - x'_{ijkl}\beta - z'_{ijkl}u_i\right\}/\sigma$, x' is a row of the fixed effects model matrix, z' is a row of the random effects model matrix, u_i is a vector of random effects, $I(\cdot)$ is the indicator function, $\Phi_1(\cdot)$ represents the normal distribution function, and $\varphi_q(\cdot)$ represents the q-dimensional normal density.

Parameter estimates from the linear mixed effects model without right censoring were used as start values for maximum likelihood estimation of the right censored model. The right-censored model was estimated with SAS PROC NLMIXED, which presents some limitations compared to SAS PROC MIXED. The limitations, namely maximum two random components, necessitated a choice between a random effect at the child level and random effects. Estimation of the right-censored model with random effect at the child level encountered substantial difficulties with convergence and the right-censored model with both random intercept and random slope almost never converged. Therefore, most of the right-censored model results presented below were estimated with random teacher effect. Under this random effect specification, all measurements on students under the same teacher are correlated, both within student and across student, which captures the main source of correlation among the measurements in the same classroom Since the structure of the covariance matrix for the six measurements across time may have been misspecified in this approach, the robust “sandwich” estimator (Liang & Zeger, 1986; White, 1980) was used for estimated standard errors and tests of fixed-effect parameters. Comparing results from the SAS PROC MIXED model and the right censored-model, the same model fixed effects are found significant in both approaches. Usually the TELL effect is larger in magnitude under the right-censored model, which is intuitively reasonable since TELL children tend to reach the maximum possible score at an earlier time point than their peers in BAU classes. Hypothesis tests on fixed effects parameters were conducted using Wald statistics. Whereas the Kenward-Rogers denominator degrees of freedom from SAS PROC MIXED were over 900, the degrees of freedom for t-tests (square root of Wald statistics) were obtained from the number of teachers and were usually equal to 60. But the effect on *p*-values for tests of the fixed effect parameters were minimal since degrees of freedom equal to 60 is large enough so that the distribution of the test statistics is quite close to normal.