Appendix C: Multilevel Analysis: Combining Standardized Bias-Corrected Effect Sizes

A regression-based effect size capturing the effectiveness of peer tutoring was estimated using the equation suggested by Van den Noortgate and Onghena (2003a, 2003b):

$$y_{ijk} = \beta_{0jk} + \beta_{1jk}D_{ijk} + e_{ijk} \text{ with } e_{ijk} \sim N(0, \sigma^2_{e_{ijk}}) \quad (1)$$

where y_{ijk} refers to the outcome score at measurement occasion *i*, for subject *j*, nested within study *k*. D_{ijk} is a dummy variable, equaling 0 during the baseline and 1 during the treatment. As a consequence, β_{0jk} refers to the intercept, reflecting the outcome level during the baseline. β_{1jk} reflects the difference between the treatment level and the baseline level, referred to as the average immediate treatment effect. In addition to the immediate treatment effect, we wanted to capture the effectiveness of peer tutoring across time. Therefore, an extension of the piecewise regression equation suggested by Center, Skiba and Casey (1885-1886) was used:

$$y_{ijk} = \beta_{0jk} + \beta_{1jk}D_{ijk} + \beta_{3jk}D_{ijk}Time_{ijk} + e_{ijk} \text{ with } e_{ijk} \sim N(0, \sigma^2_{e_{ijk}}) \quad (2)$$

In Equation 2 a time coded variable was added, $Time_{ijk}$ beginning at 0 and continuing to the end of the experiment. For a more detailed discussion of the interpretation of the regression coefficients, we refer the reader to Moeyaert et al. (2014).

In this study, Equations 1 and 2 need to be extended as we have different types of outcomes (i.e., social-behavioral outcomes and academic outcomes). Therefore, we added a dummy variable in the model indicating the type of outcome, O_{ijk}, representing whether measurement occasion, *i*, nested within case *j*, from study *k* was obtained for social outcomes (O_{ijk} = 0) or academic outcomes (O_{ijk} = 1). The mathematical model without trends is displayed in Equation 3:

$$y_{ijk} = \beta_{00jk} + \beta_{01jk}O_{ijk} + \beta_{20jk}D_{ijk} + \beta_{21jk}D_{ijk}O_{ijk} + e_{ijk}$$

with $e_{ijk} \sim N(0, \sigma^2_{e_{ijk}})$ (3)

β_{00jk} represents the baseline level for social outcomes; therefore, $\beta_{00jk} + \beta_{01jk}O_{ijk}$ is the baseline level for academic outcomes, and $\beta_{20jk}D_{ijk}$ indicates the immediate treatment effect for social outcomes (O_{ijk} = 0), whereas $\beta_{21jk}D_{ijk}O_{ijk}$ indicates the difference in immediate treatment effect between social outcomes and academic outcomes. Given our research questions, the effect sizes of interest are: β_{20jk}, β_{21jk}, β_{30jk}, and β_{31jk}. The effect of the treatment effect on the time trend can also be modelled using the following equation:

$$y_{ijk} = \beta_{00jk} + \beta_{01jk}O_{ijk} + \beta_{20jk}D_{ijk} + \beta_{21jk}D_{ijk}O_{ijk} + \beta_{30jk}D_{ijk}Time_{ijk} + \beta_{31jk}D_{ijk}Time_{ijk}O_{ijk} + e_{ijk} \text{ with } e_{ijk} \sim N(0, \sigma^2_{e_{ijk}}) \quad (4)$$

By using this extended regression equation, two extra regression coefficients are obtained, namely β_{30jk} and β_{31jk} respectively, indicating the change of the effectiveness of the treatment across time for social outcomes ($O_{ijk} = 0$) and the difference in effectiveness across time between social outcomes and academic outcomes.

Prior to combining the effect sizes, the raw data were standardized as the included studies included a variety of outcome measures. For instance in Chiang, Thorpe, and Darch's (1980) study of the effect of peer tutoring on word recognition, the outcome scale was the percentage of words that were correctly recognized. However, Neddenriep, Skinner, Wallace, and McCallum (2009) studied the effect of peer tutoring on oral reading fluency and the outcome scale was words read correct per minute. The standardization procedure suggested by Van den Noortgate and Onghena (2008) and validated by Ugille, Moeyaert, Beretvas, Ferron, and Van den Noortgate (2012) was applied prior to combining the effect sizes. For instance, the standardized immediate treatment effect for social outcomes equals:

$$b'_{20jk} = \frac{b_{20jk}}{\hat{\sigma}_e} \quad (5)$$

Because SCEDs result in small data sets, Ugille, Moeyaert, Beretvas, Ferron, and Van den Noortgate (2013) suggested correcting the standardized effect sizes (i.e., b'_{20jk} in Equation 5) for small-sample bias by multiplying the effect size by Hedges' bias correction factor (Hedges, 1981), which is approximately equal to 1-[3/(4*m*-1)], with *m* indicating the degrees of freedom. In the models discussed above, *m* equals the number of measurement occasions (*I*) minus the number of predictors (*p*) in the regression model minus 1 (i.e., $m = I - p - 1$). For instance, the corrected standardized immediate treatment effect equals:

$$b'^{C}_{20jk} = b'_{20jk}\left(1 - \frac{3}{4\,(I - p - 1) - 1}\right) \quad (6)$$

As a final step, the standardized bias-corrected effect sizes are combined across cases and across studies using the multivariate multilevel meta-analytic model as illustrated by Moeyaert et al. (2014). These results are shown in Model 1 in the current study. Potential moderators of the effect sizes were added in Model 2.