Evaluation of Organizational Skills Training (OST)
Program for Upper Elementary Students

NCT03443323

June 5, 2022

Data Management and Analysis

Analyses followed the intention-to-treat (ITT, as randomized) principle to examine the effect of OST-T2 at post-treatment and follow-up. Time zero was established as the date of consent. All measures were scored according to published instructions, and mean item scores were used in analyses. If an item was missing, mean item scores for each scale were computed without the missing item. An examination of pre-specified baseline student-level covariates, indicated in Table 1, noted some imbalances between groups. For this reason, we rebalanced the data using inverse probability of treatment weights derived from a logistic regression with treatment assignment as the outcome and all covariates as predictors (Loux & Huang, 2023). Once weighted, the baseline covariates collectively produced no association with treatment assignment (Model $p = 0.99$). This balancing procedure was then used in all outcome models to adjust for student-level covariates.

The pattern of medication use to treat ADHD symptoms was similar across conditions. Caregiver-reported rates of student medication use at baseline were 18% for the OST-T2 condition and 23% for TAU/WL; rates were 14% for OST-T2 vs. 18% for TAU/WL at post-treatment, and 16% vs. 21% at 12-month follow-up. Overall, only 9% of students in each condition had a change in medication status at any point in the trial. As such, medication status was not controlled in the analyses.

Prior to analyzing outcome data, we found that actual administration dates of assessments (a) departed frequently from pre-specified times within each assessment period, (b) varied by the informant completing the assessments (caregivers, teachers), (c) varied between treatment arm, and (d) overlapped so that post-treatment data for some students were sometimes collected later than the 5-month follow-up data for others. Discrepancies in data collection from pre-specified times generally were due to variations among informants in response to data requests. For that reason, we preplanned to model time as a continuous variable and allow for non-linear effects of treatment assignment by means of restricted cubic splines with knots at pre-specified locations (Fitzmaurice et al., 2011).

Linear mixed effects models were used to estimate the effects of OST-T2 on outcomes. Models had three levels: assessment time nested within student and students nested within school. All models included time (as splines), treatment, and time-by-treatment interactions. Mixed effects models included random intercepts for school (to account for clustering of students within schools), random slopes (over time) to allow for different trajectories across schools, and a random intercept and random slope for student to allow for variation in the effect of treatment assignment across students over time. Modeling for the clustering effect of schools accounted for more than one student being rated by multiple teachers; teachers rated a mean of 1.6 students and approximately 85% of teachers rated three or fewer students at each time point. Statistical significance was evaluated using a threshold p value of .0083, accounting for the estimation of three contrasts of time-by-treatment differences per outcome and two measures per domain (OTMP skills [COSS-P, COSS-T], homework performance [HPC, HPQ-T], and academic performance [ACES, APS]).

Mixed effects models facilitated both the graphic display, and contrasts, of curves depicting the expected values of outcomes over time from time zero until 12-month follow-up for the student with the last data collected. All confidence bounds reflected variances based on the multi-level linear model. To evaluate differences between conditions, we estimated model-based, expected values for outcomes at four preplanned times, 0 (baseline), 14 (post-treatment), 22 (5-month follow-up) and 55 (12-month follow-up) weeks, except for academic grades, which

were estimated for beginning and end of school year. Those actual times were translated into their equivalent values in the models for restricted cubic splines. Expected values were then the product of these times and estimated coefficients from the mixed effects models. The primary overall estimate of intervention effect for each outcome was the between-group difference of the change in expected values (means) from time zero to the prespecified outcome assessment times. The mixed effects models applied in this study assume that missing scores were missing at random, a reasonable assumption given that we also adjusted for student baseline covariates that might explain missed assessments. Analyses used all available scores, consistent with an ITT approach. Sensitivity analyses were conducted to determine whether data collected after the school shutdown due to the COVID-19 pandemic (March 16, 2020) altered study findings.

Mixed effects models were also used to estimate student-level outcomes at each assessment time. We estimated random effects for students and schools to determine model-based estimates, which reflect expected values (average score of the entire group) plus individual random effects (departures of the individuals from average). Statistical theory suggests that model-based predictions provide better estimates of true values than the actual data. Model-based predictions, called "empirical Bayes" estimates, combine individual measures with group-level expected values in a manner that permits individual estimates to derive strength from the group effects (Efron & Morris, 1973). Using these estimates for each student at each assessment point, we categorized all students based on thresholds derived from normative data sets or common-sense scale values on measures. For the COSS-P, COSS-T, HPC, and HPQ-T, thresholds were set at one-half standard deviation below/above the mean of the normative sample to indicate performance within the normal range (for normative information, see masked for review; Power et al., 2006; Power et al., 2007), as this corresponds with a medium effect size. For the ACES and APS, threshold values were set at a value reflecting performance at the standard for grade level (3.0) or above. For academic grades, the threshold was set at a value of 3.0 or above.

To estimate the magnitude of the intervention effect, we computed Cohen's *d*. This statistic was calculated as the ratio of between-group differences in changes in expected values over time divided by the standard deviation of the combined, unadjusted baseline scores for each outcome measure.