AMERICAN PSYCHOLOGICAL ASSOCIATION

Journal of Consulting and Clinical Psychology

ISSN: 0022-006X

2024, Vol. 92, No. 10, 674–691
https://doi.org/10.1037/ccp0000909

School-Based Organizational Skills Training for Students in Grades 3–5: A Cluster Randomized Trial

Jenelle Nissley-Tsiopinis[1, 2], Thomas J. Power[1, 3], Phylicia F. Fleming[1], Katie L. Tremont[1], Bridget Poznanski[1], Shannon Ryan[1], Jaclyn Cacia[1], Theresa Egan[1], Cristin Montalbano[1], Alex Holdaway[1], Ami Patel[1], Richard Gallagher[4], Howard Abikoff[4], A. Russell Localio[2], and Jennifer A. Mautone[1, 2]

[1] Department of Child and Adolescent Psychiatry and Behavioral Sciences, Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, United States

[2] Department of Psychiatry, Perelman School of Medicine at University of Pennsylvania

[3] Department of Pediatrics, Psychiatry, and Education, Perelman School of Medicine at University of Pennsylvania

[4] Department of Child and Adolescent Psychiatry, New York University Grossman School of Medicine

Objective: Research has demonstrated the effectiveness of interventions to reduce organizational skills deficits and homework problems, including the clinic-based Organizational Skills Training (OST-C) program (Abikoff et al., 2013). In this study, OST-C was adapted for schools as a small-group (Tier 2) intervention delivered by school partners (OST-T2). ***Method:*** The study was conducted in 22 schools serving students from diverse backgrounds. Students (n = 186; 122 male) in Grades 3–5, ages 8–12 (M = 9.7 years; SD = 0.88) with organizational skills deficits referred by teachers were enrolled. Schools were randomly assigned to OST-T2 or treatment as usual with waitlist. OST-T2 consisted of sixteen 35-min child sessions, two caregivers, and two teacher consultations. Outcomes were evaluated with longitudinal mixed effects modeling at posttreatment, 5-month and 12-month follow-up using caregiver and teacher reports of organizational skills, homework, and academic performance. ***Results:*** OST-T2 resulted in reductions in organizational skills deficits on caregiver and teacher report ($p < .001$) at posttreatment and 5-month follow-up (effect sizes [ES], Cohen's d = 0.96, 1.20). Findings also revealed a reduction in caregiver-reported homework problems at posttreatment and 5-month follow-up ($p < .001$, ES = 0.60, 0.72), and an improvement in teacher-rated homework at posttreatment ($p = .007$, ES = 0.64). Effects were attenuated at 12-month follow-up. The effects of OST-T2 on academic measures were not significant. ***Conclusions:*** Findings provide evidence for the immediate and short-term effectiveness of OST-T2 delivered by school professionals.

What is the public health significance of this article?
This study demonstrated the effectiveness of organizational skills training provided by school professionals for students ages 8–11, with teacher-identified organization, time management, and planning difficulties.

Keywords: organizational skills, child skills training, school mental health, intervention, executive functioning

Supplemental materials: https://doi.org/10.1037/ccp0000909.supp

Bunmi Olatunji served as action editor.

Jenelle Nissley-Tsiopinis https://orcid.org/0000-0003-2680-3605
Thomas J. Power https://orcid.org/0000-0001-6651-4002
Bridget Poznanski https://orcid.org/0000-0001-5558-0928
Shannon Ryan https://orcid.org/0009-0007-4661-885X
Jaclyn Cacia https://orcid.org/0000-0001-7363-0802
Alex Holdaway https://orcid.org/0000-0002-8921-7307
Ami Patel https://orcid.org/0000-0001-5697-7257
Richard Gallagher https://orcid.org/0000-0002-6612-3463
Howard Abikoff https://orcid.org/0000-0002-6360-3605
A. Russell Localio https://orcid.org/0000-0002-0506-851X
Jennifer A. Mautone https://orcid.org/0000-0002-2712-5255

This study was approved by the internal review board of the hospital and preregistered at https://clinicaltrials.gov (NCT03443323). All data analytic code, research materials and a detailed statistical analysis plan are available from the authors upon request. The Organizational Skills Training–Tier 2 (OST-T2) manual is an adaptation of the clinic-based treatment manual that is copyright protected by Guilford Press (Gallagher et al., 2014). Research teams that have an interest in the OST-T2 intervention manual can contact Richard Gallagher at richard.gallagher@nyulangone.org or Jenelle Nissley-Tsiopinis at nissleyj@chop.edu.

Multiple uses of data collected from the same sample: This article has not been published previously and is not currently under consideration by another journal. The authors do not have any previously published or currently in-press works stemming from the data set used in the submitted article. The data reported here have been previously presented in six different article presentations at local, national, and international academic conferences, with only brief abstracts published for each presentation.The main numerical results and analyses for all of the variables reported in this article were not included in any of these published abstracts.

Howard Abikoff and Richard Gallagher and the New York University Grossman School of Medicine receive royalties from sales of the Children's Organizational Skills Scale and the manual for the clinic version of

continued

Executive functioning (EF) is a set of cognitive processes that allow individuals to control thoughts, emotions, and actions in new or complex situations to plan and execute goal-directed behavior (Best et al., 2009). Impairments in EF occur in several clinical populations. An estimated 90% of children with attention-deficit/hyperactivity disorder (ADHD) show deficits in EF (Kofler et al., 2019). EF deficits are also common among children with learning disabilities (Crisci et al., 2021), autism spectrum disorders (O'Hearn et al., 2008), neurological disorders (Slick et al., 2006), and children who have had adverse psychosocial experiences (Lund et al., 2020). Demands for EF increase as students advance from preschool, and many children, including those without a clinical diagnosis, manifest impaired EF in elementary school. Given the clear associations between EF and academic performance, and the fact that EF in childhood predicts academic achievement both among all students (Best et al., 2011) and those with clinical diagnoses (Rinsky & Hinshaw, 2011), the consequences of EF difficulties are substantial. Furthermore, EF deficits also impact home behavior and contribute to family conflict (Abikoff et al., 2013).

Recognition of the impact of EF on school success, behavioral adjustment, and emotional functioning led to the development of interventions to improve EF skills. Effective interventions target behavioral manifestations of EF that are directly related to real-world functioning, such as organization, time management, and planning (OTMP) skills, as opposed to cognitive processes (Chacko et al., 2014). For example, in a sample of middle school students with ADHD, Langberg et al. (2013) found that organization of materials and task planning were associated with homework performance and school grades over and above associations with the symptoms of ADHD. Similarly, organizational problems such as failing to record homework assignments and complete assignments were associated with teacher-reported learning problems (Power et al., 2006). Further, teachers reported lower achievement in children who misplaced assignments or struggled with task readiness (Langberg et al., 2011). Improvements in OTMP skills following organizational skills treatment are associated with improvements in academic productivity and performance, reductions in family conflict, and improvements in children's attitudes toward teachers and school (Abikoff et al., 2013). These studies highlight the importance of OTMP skills for academic success, as well as improved family interactions, suggesting that these behavioral manifestations of EF are important intervention targets.

Numerous interventions have been developed to reduce OTMP deficits in children, and most of the research has focused on children with ADHD. Some interventions are embedded in comprehensive programs with multiple components (e.g., Challenging Horizons, Evans et al., 2016; Collaborative Life Skills, Pfiffner et al., 2016); others have a specific focus on OTMP deficits (e.g., Homework, Organization, and Planning Skills program, Langberg et al., 2018; Supporting Teens' Autonomy Daily, Sibley et al., 2016). Most of the existing research has targeted students at the middle school level (e.g., Evans et al., 2016; Langberg et al., 2018; Sibley et al., 2016). Overall, research supports the efficacy of these interventions in reducing OTMP deficits and homework problems (Evans et al., 2018).

The clinic-based Organizational Skills Training (OST-C) program was designed to help students in Grades 3–5 with ADHD and OTMP deficits to organize themselves to meet task demands at school and home. Although these students receive considerable organizational supports, there are increasing demands for developing organizational skills as they prepare for middle school (Best et al., 2011). OST-C is

Organizational Skills Treatment. Thomas J. Power and Jennifer A. Mautone have a financial interest in the Homework Performance Questionnaire, which was published as an appendix in a book they coauthored, *Family–School Success for Children with Attention-Deficit/Hyperactivity Disorder: A Guide for Intervention* (Guilford Press, 2024). The other authors have no financial interests.

The research was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A170052-20 to Thomas J. Power, Jenelle Nissley-Tsiopinis, and Jennifer A. Mautone at the Children's Hospital of Philadelphia. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education. The authors thank the students, caregivers, and school professionals who participated in, and supported, the study, as well as the research team members who worked on this project.

Jenelle Nissley-Tsiopinis played a supporting role in validation and an equal role in conceptualization, funding acquisition, investigation, methodology, project administration, resources, supervision, visualization, writing–original draft, and writing–review and editing. Thomas J. Power played an equal role in conceptualization, funding acquisition, investigation, methodology, project administration, resources, supervision, validation, visualization, writing–original draft, and writing–review and editing. Phylicia F. Fleming played a supporting role in investigation, methodology, supervision, validation, and writing–review and editing. Katie L. Tremont played a supporting role in data curation, investigation, project administration, validation, visualization, and writing–review and editing. Bridget Poznanski played a supporting role in data curation, formal analysis, investigation, validation, and writing–review and editing. Shannon Ryan played a supporting role in data curation, investigation, and writing–review and editing. Jaclyn Cacia played a supporting role in data curation, investigation, methodology, project administration, supervision, validation, and writing–review and editing. Theresa Egan played a supporting role in investigation and writing–review and editing. Cristin Montalbano played a supporting role in investigation and writing–review and editing. Alex Holdaway played a supporting role in investigation and writing–review and editing. Ami Patel played a supporting role in data curation, investigation, and writing–review and editing. Richard Gallagher played a supporting role in conceptualization, funding acquisition, methodology, resources, supervision, and writing–review and editing. Howard Abikoff played a supporting role in conceptualization, funding acquisition, methodology, resources, supervision, and writing–review and editing. A. Russell Localio played a lead role in formal analysis and software, a supporting role in funding acquisition and writing–review and editing, and an equal role in conceptualization, data curation, methodology, resources, validation, and visualization. Jennifer A. Mautone played a lead role in data curation, a supporting role in validation, and an equal role in funding acquisition, investigation, methodology, project administration, resources, supervision, visualization, writing–original draft, and writing–review and editing.

Correspondence concerning this article should be addressed to Jenelle Nissley-Tsiopinis, Department of Child and Adolescent Psychiatry and Behavioral Sciences, Children's Hospital of Philadelphia, 2716 South Street, Philadelphia, PA 19146, United States. Email: nissleyj@chop.edu

an individualized, child skills training intervention that includes substantial caregiver involvement and some teacher involvement to promote skills generalization. In a large-scale randomized controlled trial, OST-C was compared to a caregiver and teacher training intervention (Parents and Teachers Helping Kids Organize) and a waitlist control. Both interventions reduced OTMP deficits and homework problems and improved academic performance relative to the control group (Abikoff et al., 2013). Gains in OTMP functioning persisted into the next school year as reported by teachers who were unaware of the intervention. Although both interventions produced similar effects, OST-C resulted in significantly greater reductions in caregiver-reported OTMP deficits, reductions in family conflict, and long-term improvements in children's attitudes toward school and teachers (Abikoff et al., 2013).

Despite the success of OST-C, questions have been raised about its accessibility, feasibility, generalizability, and sustainability. The intervention was offered in a clinic by highly trained clinicians exclusively to individual children with OTMP deficits diagnosed with ADHD. To extend its reach, Di Bartolo et al. (2021) adapted OST as a small-group intervention implemented in schools by school professionals for students with OTMP deficits, regardless of ADHD diagnostic status and with limited caregiver and teacher involvement. The involvement of school professionals as interventionists was designed to expand the workforce of individuals in the community to provide OST, thereby increasing accessibility and sustainability. A pilot study of the school version (OST-S) in an open trial provided evidence of acceptability and feasibility and preliminary support for intervention effectiveness (Di Bartolo et al., 2021).

Based on the success of OST-C and the pilot of OST-S, the present study was a large-scale, cluster-randomized trial of OST-S across schools serving students from diverse backgrounds. To promote accessibility and feasibility, further adaptations were made to OST-S. Specifically, the length of the intervention was reduced from 19 sessions to 16, the amount of content per session was reduced to fit within a 35-min period, and group size was increased from three to four. As such, the revised program is consistent with a Tier 2 (small-group, targeted) intervention within a multitier system of support (Sugai & Horner, 2006) and called OST-T2. The OST-T2 intervention was provided by staff employed by participating schools (called school partners) who received training and consultation from the study team. To promote generalizability, the intervention targeted students with OTMP difficulties regardless of diagnostic and special education classification status. Supplemental Figure S1 depicts the theory of change. OST-T2 includes (a) cognitive framing with students to promote a sense of self-efficacy, (b) in-session skills training and practice with students, (c) between-session practice using skills to promote generalization, (d) consultation with caregivers and teachers to support practice between sessions, and (e) consultation with school partners to promote fidelity and student engagement. OST-T2 is designed to improve OTMP skills, which in turn can lead to improved homework and academic performance. In this study, schools were assigned to OST-T2 or treatment as usual with waitlist (TAU/WL) to identify the incremental benefit of OST-T2 compared to organizational supports typically offered to students. See Supplemental Table S1 for a comparison among OST-C, OST-S, and OST-T2.

The main study aim was to evaluate the effectiveness of OST-T2 on participant outcomes, including OTMP skills deficits, homework performance, academic performance, and student self-efficacy at posttreatment and follow-up 5 months and 12 months from baseline. We hypothesized that compared to TAU/WL, OST-T2 would be superior on all study outcomes at posttreatment and follow-up. An additional aim examined the acceptability, usability, and feasibility of OST-T2.

Method

Study Design

The study employed a cluster randomized controlled design with a 1:1 random assignment of schools to conditions: OST-T2 versus TAU/WL. Randomization was stratified to achieve group balance on school subsidized lunch rates. Outcome evaluation was planned for baseline, posttreatment, 5 months following baseline (2 months after treatment), and 12 months following baseline (9 months after treatment). In 18 of 22 schools, there were two study cohorts (assigned to OST-T2 or TAU/WL) per school year (i.e., fall and spring). In addition, 17 of 22 schools participated for one school year; the remaining five were involved for two school years.

Study Sites

Study enrollment occurred between 2017 and 2022 at a tertiary-care children's hospital in a large metropolitan area in the northeast United States. Schools were recruited across urban and suburban counties. Schools were strategically invited to participate to reflect socioeconomic diversity, based on subsidized lunch rates.

Participants and Recruitment Procedures

Fifty-two (52) schools were invited to participate in the study, and 22 agreed. Free and reduced-price lunch rates were similar between participating (M = 53.4%, SD = 36.1%) and nonparticipating schools (M = 59.0%, SD = 38.0%). Criteria for school participation included (a) principal approval, (b) superintendent and/or school board approval, and (c) consent by one or more school professionals to receive training and deliver the intervention (i.e., school partners). Across the 12 schools in OST-T2, there were 13 school partners (11 non-Hispanic White; 1 Hispanic, White; 1 non-Hispanic, Black) employed by participating schools: Five (38.5%) were in instructional roles (i.e., one general education teacher, four special education teachers) and eight (61.5%) were in student support roles (i.e., seven counselors, one school psychologist). Additionally, one postdoctoral psychology fellow from the research team provided the program for one group when a school staff member was not available.

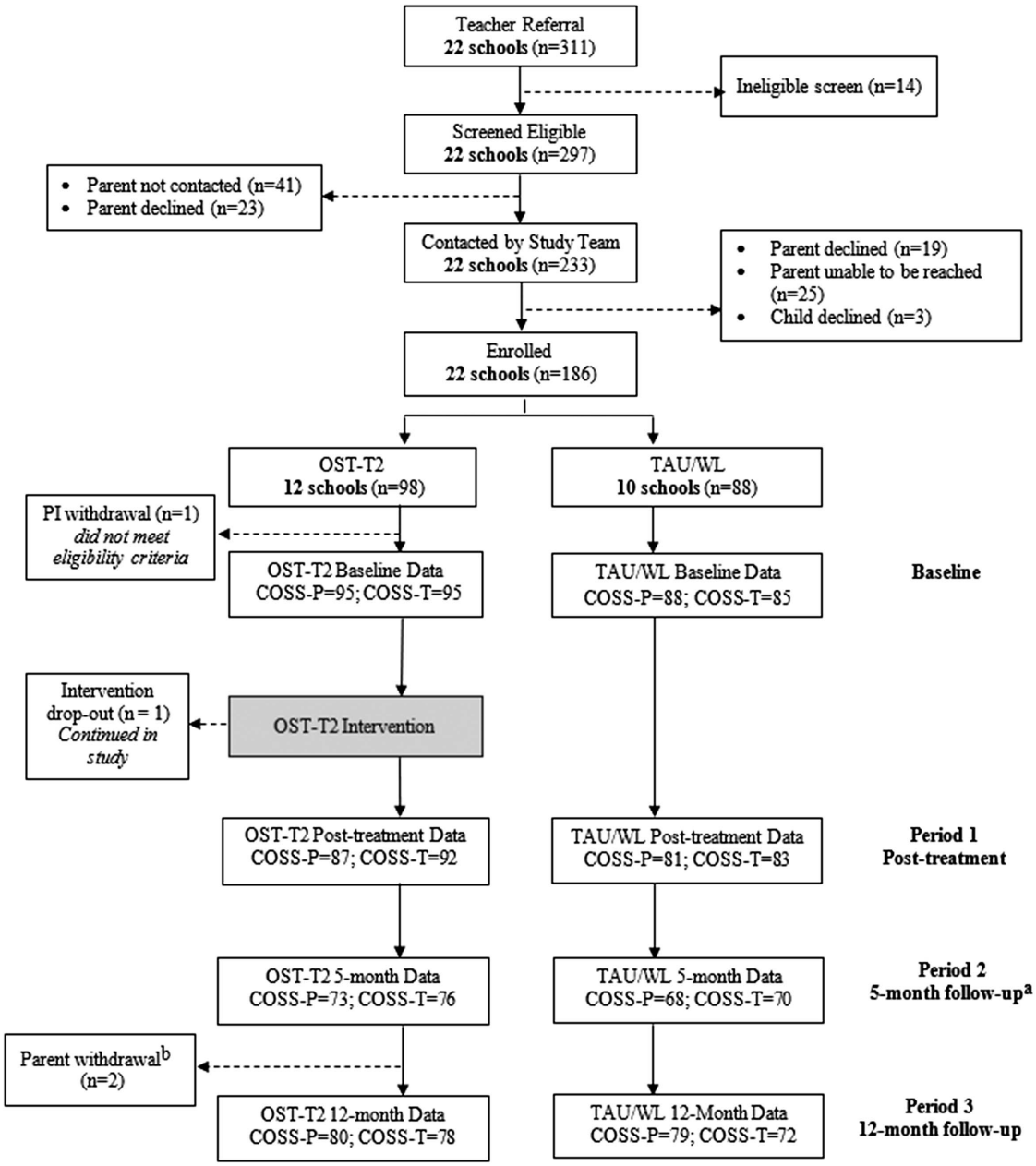
Students in Grades 3–5 were enrolled; one grade level was targeted per school, restricted partially by highest grade level in the building. This allowed one academic year for the intervention period and the following academic year for follow-up data collection and delivery of the intervention to those in the control condition. For example, if the highest grade level in the school was Grade 5, we targeted students in Grades 3 or 4. The study statistician, who was masked to the identity of the schools, determined the specific grade level targeted within each school to achieve balance in grade levels across conditions.

General education teachers nominated 1–4 students, depending on the number of classrooms at the targeted grade level. Teachers nominated students in order of concern about OTMP deficits, with

the most concerning students indicated first. Eligibility criteria included (a) students with OTMP deficits contributing to academic difficulties, with a teacher rating of "pretty much" or "very much" on at least one of the four interference items from the Children's Organizational Skills Scale (COSS; Abikoff & Gallagher, 2009); (b) students placed in general education for at least 50% of the school day; and (c) students with at least one English-speaking caregiver. Students with co-occurring learning, behavioral, and emotional problems were eligible if OTMP deficits were the primary concern. Students were excluded if they received support from a one-to-one instructional aide. The CONSORT diagram indicates the number of students referred, enrolled, and randomized to each condition (see Figure 1). A total of 186 students were enrolled and randomized at the school level.

Figure 1

CONSORT Diagram Illustrating Participant Flow From Screening to 12-Month Follow-Up



Note. CONSORT = consolidated standards for reporting trials; OST-T2 = Organizational Skills Training–Tier 2; TAU/WL = treatment as usual with wait list condition; COSS-P = Child Organizational Skills Scale–Parent versions; COSS-T = Child Organizational Skills Scale–Teacher versions; PI = principal investigator.

[a] Data collection for 5-month follow-up during Spring 2020 was not completed due to school closures during the COVID-19 pandemic. [b] In one of these two cases, 12-month follow-up data had been collected from the teacher at the time of caregiver withdrawal.

Intervention Investigated: OST-T2

OST-T2 includes 16 sessions with an introductory session and four modules designed to teach the main skills required to complete work effectively in school and during homework: tracking assignments (two sessions), managing materials (five sessions), time management (four sessions), and planning (four sessions). Each session was about 35-min and administered to a small group of three to four students. The intervention was adapted from OST-C and OST-S (Di Bartolo et al., 2021; Gallagher et al., 2014). Sessions follow a consistent format: (a) review homework with a focus on between-session use of skills learned in prior sessions, (b) teach new skills via instruction and demonstration, (c) practice skills in session, and (d) prepare for between-session skills practice. Students are reinforced for participation and practice using a token reward system. OST-T2 also involves caregivers and teachers to promote skill generalization. Like the clinic-based program, OST-T2 is based on social–cognitive theory of motivation, behavioral psychology, and social learning theory. The following are the main components of OST-T2. See Supplemental Table S2 for a description of program sessions and content.

Cognitive Framing to Prepare Students for Change

OST-T2 is grounded in attribution theory, which posits that individuals are more persistent in facing challenges when they see problems as malleable, under their control, and able to be changed with effort (Dweck & Elliott-Moskwa, 2010). OST-T2 providers foster a belief that students can become more organized by externalizing OTMP problems as manifestations of mental "glitches" that can be controlled by building the skills of the "Mastermind" through program participation.

In-Session Skills Training and Practice With Students

A major purpose of each OST-T2 session is skills training and practice. Skills training included verbal instruction, modeling, shaping, guided practice, positive reinforcement, and strategic use of verbal correction. A key component of each session is repeated practice of the new skill (at least three rounds of practice).

Strategies to Promote Generalization

For a skills training intervention to be effective, it is critically important to include strategies to support generalization (Abikoff, 2009).

Student-Focused Strategies. OST-T2 prioritizes in-session and between-session opportunities for students to practice OST-T2 skills. During sessions, students practice applying skills in a range of tasks in which skills will be used. In addition, students practice new skills between sessions during completion of their daily home and school tasks, including homework. At the following session, students show evidence of their between-session practice and receive reinforcement (i.e., points and prizes).

Caregiver and Teacher Consultation. During two caregiver and two teacher 20-min consultation sessions provided at the onset and midway through the program, caregivers and teachers are informed about session content and are guided to prompt children to practice using skills between sessions and to praise skill use. Caregiver and teacher meetings include development of a plan for verification and reinforcement of student between-session skills practice and discussion about how to support skills use. Caregivers and teachers receive a manual at the start of their participation and information sheets and a worksheet (i.e., Mastermind sheet) for documenting student skills use after each session, which they are instructed to complete and return to the school partner at the next student session.

Training and Consultation to Support Implementation by School Partners

School partners who agreed to deliver OST-T2 received 6–8 hr of in-person training prior to intervention, provided by a postdoctoral psychology fellow on the study team. This training included an overview of the theoretical model of OST-T2 and evidence to support the intervention, a description of key processes for delivering the program (see process fidelity dimensions outlined below), and instruction in the curriculum for each session. School partners were involved actively during training; approximately 67% of training time was dedicated to modeling OST-T2 strategies and practicing with feedback. School partners were given an OST-T2 implementation guide and all materials needed to conduct sessions.

School partners received individual, in-person, 20-min consultation and coaching sessions after each child group session (i.e., twice weekly) from a postdoctoral fellow or psychologist on the study team. Consultation and coaching followed adapted procedures that were successfully employed in a similar school-based trial (Eiraldi et al., 2020). Adaptations were made according to published recommendations to address feasibility concerns raised by school staff (Beidas & Kendall, 2010). Before each consultation session, the consultant reviewed videos of child group sessions. Consultation sessions included discussion of at least one aspect of the prior session that was implemented effectively and at least one aspect of the session for which there was an opportunity for improvement. In addition, consultation included preparation for the next OST-T2 session. Consultants often advised school partners about how to implement session content in a manner that would be feasible and efficient for them.

Control Condition

Students in the control condition received treatment as usual (TAU) during the study (baseline through the follow-up periods), which included OTMP supports that were already provided to students at a school, classroom, or individual level. At baseline, teachers reported that 11% of students in TAU received extra support for OTMP difficulties through special education, 504 plans, or other programs, which was comparable to the percentage of students (4%) in the OST-T2 condition who received this support. After 12-month follow-up assessment, students in TAU received OST-T2 from a school partner or member of the study team.

School Partner Fidelity in Implementing OST-T2

The study team attempted to video record all OST-T2 sessions. School partners audio recorded each session as a backup. To evaluate whether the intervention was delivered as intended, sessions were randomly selected and coded for content and process fidelity. Random selection of sessions for both initial and reliability coding was stratified by intervention module (tracking assignments,

managing materials, time management, planning) to ensure coding represented all modules of the program.

Content Fidelity

Content fidelity was examined to determine whether school partners implemented key program content. Each content fidelity coding form consisted of 16 items in three sections: (a) reviewing skill use since the last session, (b) teaching and practicing a new skill, and (c) preparing students for between-session practice. Members of the research team observed session video recordings or audio recordings if a complete video recording was not available, and completed content fidelity coding forms. Each item was rated 0 = not implemented or 1 = implemented. Between five and eight randomly selected sessions for each OST-T2 group were coded for content fidelity (n = 157 of the 378 sessions [42%] were coded). Of the sessions coded for content fidelity, 20% (n = 32) were randomly selected for reliability coding by an independent coder. The reliability of content fidelity coding across all sessions was excellent (unweighted κ = .81), and across the four modules, reliability was good to excellent (unweighted κ range = .71–.94).

Process Fidelity

Process fidelity was examined using an adaptation of the process coding form developed for the Early Alliance Project (Dumas et al., 2001). Five process dimensions identified as being key elements of OST-T2 implementation were coded: (a) effectiveness in tailoring OST-T2 sessions to each student; (b) effectiveness in using engaging learning strategies with students; (c) effectiveness in using positive reinforcement for skills use, practice, and active listening during sessions; (d) effectiveness in using a positive coping reframe (e.g., referring to glitches) when students encountered challenges; and (e) clarity of presenting OST-T2 session content. Accurate coding of process fidelity required use of video recordings. Trained members of the study team observed video recordings of sessions and completed process coding forms. Coders were trained to differentiate whether each process item occurred *almost all of the time* (rating = 4), *most of the time* (rating = 3), *sometimes* (rating = 2), or *rarely* (rating = 1). Between five and eight randomly selected sessions from each OST-T2 group were coded for process fidelity (n = 154 of 378 sessions [41%] were coded). Of the coded sessions, 42 (27%) were randomly selected for double coding by an independent coder. The reliability of all process fidelity dimensions was excellent (intraclass correlation coefficients range = .85–.96).

Participant Involvement in OST-T2

Student degree of involvement in intervention was determined by the number of OST-T2 sessions attended and the rate of attendance (number of sessions attended divided by number of sessions scheduled). Degree of caregiver and teacher involvement in intervention was also determined by the number of sessions attended and the rate of caregiver and teacher attendance.

Measures

Demographic and Clinical Characteristics of the Sample

Caregivers completed a questionnaire to assess demographic characteristics of the child and family, including child assigned sex, ethnicity, race, and medication status, and caregiver level of education. Caregivers and teachers completed the *Behavior Assessment System for Children* (3rd ed.; Reynolds & Kamphaus, 2015) and caregivers completed the Autism Spectrum Rating Scale Short Form (Goldstein & Naglieri, 2009) to characterize the sample. A geo-mapping tool (Area Deprivation Index; Kind & Buckingham, 2018) estimated family level of disadvantage based on multiple indicators of education, employment, housing quality, and poverty. National percentile rank was used in analyses.

Student Outcomes: OTMP Skills

COSS–Parent Version. The COSS–Parent Version (COSS-P) quantifies caregiver perceptions of their child's OTMP functioning (Abikoff & Gallagher, 2009). Caregivers rated the frequency of each OTMP behavior on a 4-point scale, ranging from 1 = *hardly ever or never* to 4 = *just about all the time*. The scale includes items related to organized actions, task planning, and memory and materials management. A factor analysis of the COSS-P supports the presence of a robust 26-item general factor that has an acceptable level of internal consistency (Ryan et al., 2024). Higher scores indicate more OTMP skills deficits. In the current sample, the α coefficient for the total score was .85.

COSS–Teacher Version. The COSS–Teacher Version (COSS-T) assesses teacher perceptions of student OTMP deficits (i.e., organized actions, task planning, memory, materials management; Abikoff & Gallagher, 2009), using the same response anchors used by parents. A factor analysis identified a robust general factor that forms a 28-item total score (Ryan et al., 2024). Higher scores indicate more OTMP skills deficits. In the current sample, the α coefficient for the total score was .87. The COSS-T includes four additional items assessing the extent to which OTMP deficits interfere with academic achievement, which were used for eligibility screening.

Student Outcomes: Homework Performance

Homework Problem Checklist. The Homework Problem Checklist (HPC) is a 20-item caregiver-report measure that assesses homework performance (Anesko et al., 1987). Caregivers indicate the frequency of homework behaviors on a 4-point scale (0 = *never* to 3 = *very often*). Psychometric analysis confirms factorial and convergent validity (Power et al., 2006) and treatment sensitivity (e.g., Abikoff et al., 2013; Langberg et al., 2018). Higher scores indicate more homework problems. In the current sample, the α coefficient for the total score was .93.

Homework Performance Questionnaire–Teacher Version. The Homework Performance Questionnaire–Teacher Version (HPQ-T) assesses teacher perceptions of student homework behavior (Power et al., 2007). Teachers rate the percentage of time each homework behavior occurs and the percentage of work completed or performed accurately on a 5-point scale (0 = 0%–39% to 4 = 90%–100%). The seven-item factor assessing homework productivity and efficiency was used in this study. Higher scores indicate better performance. This factor has been supported through factor analysis and has high

internal consistency (Power et al., 2007). In this study, the α coefficient for this factor was .91.

Student Outcomes: Academic Performance

Academic Competence Evaluation Scales. The Academic Competence Evaluation Scales (ACES) is a teacher-report scale that assesses student academic competence (DiPerna & Elliott, 2000). The Reading/Language Arts (11 items) and Math (eight items) subscales were used. Teachers rated student skills in relation to grade level expectations in the school (1 = far below; 3 = grade level; 5 = far above). The α coefficients and test–retest correlations for these subscales have been shown to be above .90. Scores on these subscales were combined to form a total academic competence score. In the current sample, the α coefficient for this total score was .93.

Academic Proficiency Scale. The Academic Proficiency Scale (APS) is a teacher-report measure assessing child proficiency in six academic subjects (reading, spelling, language arts, social studies, science, math) that is sensitive to OST treatment effects (Abikoff et al., 2013). Teachers rate students on a 5-point scale (1 = well below standard expected at this time of year; 3 = at standard; 5 = well above standard). In the present study, the α coefficient was .91.

Student Outcomes: Student Self-Efficacy

Patterns of Adaptive Learning Scales. The five-item Academic Efficacy subscale of the Patterns of Adaptive Learning Scales (Midgley et al., 1998) was used to explore students' perceptions of their ability to overcome challenges to perform their class work (e.g., "I can do even the hardest work in this class if I try."). Items are rated on a 5-point scale from 1 = *not at all true* to 5 = *very true*. Higher scores are more favorable. In this study, the α coefficient for this scale was .72.

Student Outcomes: Student Report Card Grades

We obtained report card grades from schools for the academic year in which the student was enrolled and recorded academic grades for the first and last marking period for all school years except for students enrolled in the 2019–2020 academic year due to the onset of COVID-19. Final grades for the 2019–2020 academic year were typically issued around March 2020, so these grades were used for that year. Standard procedures were developed to harmonize report card grades across school districts with varying grading scales (available from authors) to a uniform grading scale (1 = D/F, 2 = C, 3 = B, 4 = A). A composite academic grade score was computed for each student based on their grades in Math and Reading/Language Arts.

Intervention Acceptability, Feasibility, and Usability

The *Treatment Evaluation Inventory–Short Form*, a nine-item measure, was used to assess caregivers' views of intervention acceptability (Kelley et al., 1989). The *Children's Intervention Rating Profile* (CIRP), a seven-item measure, was used to assess children's perceptions of intervention acceptability (Witt & Elliott, 1985). Items on the Treatment Evaluation Inventory–Short Form and Children's Intervention Rating Profile are rated on a 5-point scale ranging from *strongly disagree* to *strongly agree*. Participating teachers and school partners completed the *Usage Rating Profile-Intervention–Revised* (Briesch et al., 2013), a 29-item scale. Each item is rated on a scale ranging from 1 = *strongly disagree* to 6 = *strongly agree*. For all these measures, higher scores reflect more favorable views of the intervention. For this study, three subscales were examined: *Acceptability, Feasibility*, and *School Climate* (perception of the extent to which the intervention is a good fit with the school).

Data Management and Analysis

Analyses followed the intention-to-treat (as randomized) principle to examine the effect of OST-T2 at posttreatment 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 prespecified 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 $p = .99$). 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 versus 18% for TAU/WL at posttreatment, and 16% versus 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 completion dates of assessments (a) departed frequently from prespecified 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 posttreatment data for some students were sometimes collected later than the 5-month follow-up data for others. Discrepancies between prespecified and actual time of data collection generally were due to variations in response time among informants to data requests. For that reason, we preplanned to model time as a continuous variable and allow for nonlinear effects of treatment assignment by means of restricted cubic splines with knots at prespecified 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 × 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. Teachers rated a mean of 2.4 students (range = 1–10); approximately 85% of teachers rated three or fewer students. Because schools typically had one or only a few teachers for the participating grade, and to limit the number of modeled random

Table 1
Demographic and Clinical Characteristics of Sample

Demographic or clinical characteristic	OST-T2 (n = 97)		TAU/WL (n = 88)		Group comparison
	n	%	n	%	
Child sex assigned at birth[a]					$\chi^2 = 0.40$, n.s.
Male	66	68.0	56	63.64	
Child ethnicity					$\chi^2 = 4.03$, n.s.
Hispanic or Latino	10	10.8	13	14.8	
Child race					$\chi^2 = 5.93$, n.s.
Asian	5	5.2	1	1.1	
Black or African American	25	26.0	23	26.1	
American Indian or Alaska Native	0	0.0	0	0.0	
Native Hawaiian or Other Pacific Islander	1	1.0	0	0.0	
White	50	52.1	55	62.5	
More than one race	11	11.5	8	9.1	
Unknown or other	5	4.2	1	1.1	
Child grade[a]					$\chi^2 = 4.23$, n.s.
3	27	27.8	37	42.0	
4	46	47.4	32	36.4	
5	24	24.8	19	21.6	
Caregiver mean level of education[a]					$\chi^2 = 8.12$, $p < .05$
High school graduate or less	13	13.8	22	25.0	
Partial college/specialized training	30	31.9	36	40.9	
Standard college or graduate degree	51	54.3	30	34.1	
Children on ADHD medication (baseline)	12	12.5	15	17.0	$\chi^2 = 0.76$, n.s.
Likely ADHD diagnosis[b]	46	47.9	46	52.3	$\chi^2 = 0.35$, n.s.
Demographic or clinical characteristic	*M*	*SE*	*M*	*SE*	
Free and reduced-price lunch percentage[a, c]	52.8	10.8	53.9	11.0	$t = -0.70$, n.s.
Area deprivation index national percentile[a]	39.3	2.5	52.8	2.6	$t = -3.77$, $p < .001$
BASC-T attention problems T-score[a]	61.2	0.9	65.1	0.8	$t = -3.32$, $p < .001$
BASC-P anxiety T-score[a]	54.1	1.2	53.7	1.0	$t = 0.26$, n.s.
ASRS T-score[a]	50.9	1.2	54.1	0.9	$t = -2.59$, $p < .05$

Note. *SE* = standard error; OST-T2 = Organizational Skills Training–Tier 2; TAU/WL = treatment as usual with wait-list; n.s. = not significant; ADHD = attention-deficit/hyperactivity disorder; BASC-T = Behavior Assessment Scale for Children, Teacher Rating Scales; BASC-P = Behavior Assessment Scale for Children, Parent Rating Scales; ASRS = Autism Spectrum Rating Scales.
[a] Variable was selected prior to analyses as a covariate to be included in analytic models. [b] Likely ADHD diagnosis was determined based on caregiver and teacher BASC-3 T-scores. Children were considered likely to meet criteria for an ADHD diagnosis if caregiver *and* teacher ratings were 60 or above on Attention Problems and/or Hyperactivity, *and* at least one reporter had a score of 65 or above on one of these factors. All children in the sample met impairment criteria based on study inclusion criteria. [c] School-level variable.

effects, we accounted for within-school correlations of students and teachers using random intercepts and random slopes at the school level. 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]). Report card grades were evaluated using a p of .05 as it was unique as a naturalistic measure of outcome and there was a comparison only at 5-month follow-up.

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 multilevel linear model. To evaluate differences between conditions, we estimated model-based, expected values for outcomes at four preplanned times, 0 (baseline), 14 (posttreatment), 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 intention-to-treat 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 0.5 *SD* below/above the mean of the normative sample to indicate performance well within the normal range (for normative information, see Power et al., 2006, 2007; Ryan et al., 2024). This approach is consistent with the one used by Swanson et al. (2001), who defined intervention success as the absence of the presenting problem, which was operationalized in this study as scoring well within the normal range and therefore not likely to need intervention. 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 3.0 or above.

To estimate magnitude of intervention effect, we computed Cohen's *d*, 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.

Transparency and Openness

The study was approved by the internal review board of the hospital and preregistered with https://clinicaltrials.gov (Registration Number NCT03443323). Study data were collected and managed using Research Electronic Data Capture electronic data capture tools hosted at the Children's Hospital of Philadelphia (Harris et al., 2009). We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study (See Supplemental Section 1 for additional details). All models were fit with the program "mixed" in Stata software Versions 17 and 18. All data, data analytic code, research materials, and a detailed statistical analysis plan are available from the authors upon request. The OST-T2 manual is an adaptation of the clinic-based treatment manual that is copyright protected by Guilford Press (Gallagher et al., 2014). Research teams that have an interest in the OST-T2 intervention manual can contact Dr. Richard Gallagher or Dr. Jenelle Nissley-Tsiopinis.

Results

Characteristics of Sample

As indicated in Table 1, 34.2% of the sample was female (65.8% male); 12.5% was Hispanic; 35.3% was Black/African American; 10.3% was multiracial; 66.8% was White. Overall, 44% of caregivers reported a college education or higher, which is slightly higher than the national average (38%; U.S. Census Bureau, 2023). The mean age of participating students was 9.7 years ($SD = 0.88$; range 8.1–12.3 years). The application of multi-informant criteria based on Behavior Assessment Scale for Children, Parent Rating Scales and Behavior Assessment Scale for Children, Teacher Rating Scales scores (see table note) indicated that 49.7% of the sample displayed elevated symptoms of ADHD and likely met criteria for the diagnosis. Caregivers reported that 14.6% of children were on medication for ADHD at baseline.

The OST-T2 and TAU/WL groups were generally highly similar, which was expected based on the cluster-randomized design. However, the TAU/WL group had a higher level of deprivation than the OST-T2 group on the Area Deprivation Index, lower percentage of caregivers with a college education, higher level of inattention assessed by the Behavior Assessment Scale for Children, Teacher Rating Scales, and greater severity of autism spectrum disorder symptoms assessed by the Autism Spectrum Rating Scale. The mixed effects models examining intervention outcomes adjusted for these variables and other covariates, including child assigned sex, child grade level, presence of a likely diagnosis of ADHD, school subsidized lunch rate, and ratings of anxiety on the Behavior Assessment Scale for Children, Parent Rating Scales.

School Partner Intervention Fidelity

The school partners implemented the content of OST-T2 with a high level of fidelity. Across the 25 groups receiving OST-T2, 87.7% of the intervention content ($SD = 6.5\%$; range = 69%–99%) was implemented by school partners. In general, the modules of intervention were delivered with a similar level of fidelity: introductory session ($M = 91.3\%$, $SD = 10.5\%$); two tracking assignment sessions ($M = 91.6\%$, $SD = 8.8\%$); five managing materials sessions ($M = 89.9\%$, $SD = 8.6\%$); four time management sessions ($M = 85.5\%$, $SD = 9.1\%$), but the four planning sessions had a somewhat lower level of content fidelity ($M = 79.4\%$, $SD = 12.3\%$). The process with which OST-T2 was implemented by school partners was generally competent (favorable score >2.5 with a maximum score of 4.0). Mean process fidelity scores across OST-T2 groups were 3.5 ($SD = 0.3$) for individualized tailoring of session; 3.8 ($SD = 0.2$) for use of engaging learning strategies; 3.4 ($SD = 0.5$) for positive reinforcement of student effort; 3.3 ($SD = 0.3$) for highlighting student coping in response to challenges; and 3.6 ($SD = 0.3$) for clarity of presenting OST-T2 content.

Student, Caregiver, and Teacher Involvement in OST-T2

Student, caregiver, and teacher attendance was high. Students attended a mean of 14.0 sessions ($SD = 3.4$); 75% of students attended at least 13 sessions. Student attendance rate was high ($M = 92.4\%$, $SD = 9.8\%$); 75% of students had an attendance rate of 88% or higher. One student dropped out of intervention after session 13; this student remained in the study and agreed to provide outcome data. Caregivers attended a mean of 1.5 sessions ($SD = 0.7$) of a maximum of 2; the mean attendance rate for caregivers was 83.9% ($SD = 31.4\%$). Teachers attended a mean of 1.6 sessions ($SD = 0.6$) of a maximum of 2; the mean attendance rate for teachers was 91.4% ($SD = 25.1\%$). Caregivers and teachers had the opportunity to complete Mastermind sheets after Sessions 1–15. Caregivers completed a mean of 5.4 sheets per student ($SD = 4.4$), and teachers completed a mean of 5.2 sheets per student ($SD = 4.1$). At least one of these sheets was completed and returned by 83.3% of caregivers and 75.0% of teachers.

Intervention Outcomes

Model-adjusted means and standard deviations at the four assessment periods are presented in Table 2. Unadjusted means and standard deviations at each assessment period were similar to the adjusted values (Table 3). The number and percentage of students whose adjusted scores were clearly within the normal range and at or above the standard for academic performance are indicated in Table 4.

Table 2
Expected Outcome Scores by Time Point Based on Model Predictions

Outcome/treatment group	Baseline		Post (14 weeks)		5-month follow-up (22 weeks)		12-month follow-up (55 weeks)	
	EV	95% CI	EV	95% CI	EV	95% CI	EV	95% CI
COSS-P[a]								
OST-T2	2.65	[2.59, 2.72]	2.20	[2.09, 2.31]	2.25	[2.14, 2.36]	2.33	[2.17, 2.49]
TAU/WL	2.59	[2.53, 2.64]	2.51	[2.40, 2.61]	2.53	[2.44, 2.62]	2.35	[2.18, 2.51]
COSS-T[a]								
OST-T2	2.89	[2.80, 2.99]	2.27	[2.12, 2.41]	2.32	[2.19, 2.45]	2.40	[2.22, 2.59]
TAU/WL	2.86	[2.73, 2.98]	2.71	[2.61, 2.82]	2.71	[2.62, 2.80]	2.57	[2.45, 2.70]
HPC[a]								
OST-T2	1.49	[1.35, 1.63]	0.94	[0.75, 1.14]	0.97	[0.81, 1.12]	1.13	[0.95, 1.32]
TAU/WL	1.35	[1.23, 1.47]	1.20	[1.06, 1.35]	1.26	[1.11, 1.41]	1.14	[0.95, 1.34]
HPQ[b]								
OST-T2	1.76	[1.38, 2.14]	2.68	[2.23, 3.13]	2.55	[2.17, 2.93]	1.97	[1.35, 2.59]
TAU/WL	2.07	[1.64, 2.51]	2.32	[2.12, 2.53]	2.44	[2.22, 2.66]	2.14	[1.69, 2.60]
ACES[b]								
OST-T2	2.46	[2.31, 2.62]	2.71	[2.55, 2.88]	2.72	[2.57, 2.87]	2.63	[2.42, 2.84]
TAU/WL	2.31	[2.11, 2.51]	2.35	[2.17, 2.53]	2.46	[2.30, 2.61]	2.42	[2.20, 2.63]
APS[b]								
OST-T2	2.59	[2.44, 2.74]	2.77	[2.60, 2.94]	2.80	[2.65, 2.95]	2.67	[2.46, 2.88]
TAU/WL	2.57	[2.30, 2.84]	2.54	[2.40, 2.67]	2.53	[2.44, 2.63]	2.62	[2.34, 2.90]
PALS[b]								
OST-T2	3.90	[3.69, 4.10]	4.09	[3.89, 4.29]	4.18	[4.03, 4.34]	4.12	[3.75, 4.48]
TAU/WL	3.88	[3.78, 3.97]	3.98	[3.87, 4.10]	4.08	[3.84, 4.31]	4.22	[3.87, 4.57]
GRADES[b]								
OST-T2	2.50	[2.29, 2.69]			2.72	[2.49, 2.94]		
TAU/WL	2.52	[2.32, 2.72]			2.71	[2.52, 2.91]		

Note. All participants have fitted values over time, even if they do not have actual values at a given period. The OST-T2 condition had a total of 97 participants, and the TAU/WL condition had a total of 88 participants. Academic grades assigned in the marking period corresponding to baseline and 5-month follow-up assessment are reported in this table. EVs are adjusted based on mixed effects longitudinal models. EV = model-based expected values; CI = confidence interval; COSS-P = Child Organizational Skills Scale–Parent report; OST-T2 = Organizational Skills Training–Tier 2; TAU/WL = treatment as usual with wait-list; COSS-T = Child Organizational Skills Scale–Teacher report; HPC = Homework Problem Checklist; HPQ = Homework Performance Questionnaire; ACES = Academic Competence Evaluation Scales; APS = Academic Proficiency Scale; PALS = Patterns of Adaptive Learning Scales.
[a] Higher scores on these measures are less favorable. [b] Higher scores on these measures are more favorable.

Outcomes for OTMP Skills

Caregiver Report. Children in the OST-T2 intervention demonstrated a significant reduction in caregiver ratings of OTMP skill deficits assessed by the total score on the COSS-P at posttreatment in comparison to TAU/WL (adjusted mean difference = −0.37, $p < .001$, 95% CI [−0.22, −0.53]); see Figure 2. The effect of intervention remained significant at the 5-month follow-up or 8 weeks after posttreatment (adjusted mean difference = −0.35, $p < .001$, 95% CI [−0.22, −0.48]). However, there was an attenuation of intervention effect at 12-month follow-up, and group differences were not significant at this time point. Effect sizes (ES) based on Cohen's *d* at posttreatment, 5-month follow-up, and 12-month follow-up were 1.01, 0.95, and 0.34, respectively. The percentage of students scoring in the normal range at posttreatment increased by 54.5% for the OST-T2 group versus 10.7% for the controls, compared to baseline levels; the percentage normalized at 5-month follow-up increased by 50.3% for OST-T2 versus 11.4% for controls (compared to baseline), and the percentage normalized at 12-month follow-up increased by 44.5% for OST-T2 versus 22.9% for controls (compared to baseline).

Teacher Report. OST-T2 also resulted in a significant reduction in teacher ratings of OTMP skill deficits assessed by the total score on the COSS-T at posttreatment in comparison to TAU/WL (adjusted mean difference = −0.48, $p < .001$, 95% CI [−0.28, −0.68]); see Figure 3. The effect of intervention remained significant at 5-month follow-up (adjusted mean difference = −0.43, $p < .001$, 95% CI [−0.24, −0.61]). However, there was an attenuation of intervention effect at 12-month follow-up. Cohen's *d* at posttreatment, 5-month follow-up, and 12-month follow-up was 1.21, 1.08, and 0.42, respectively. The percentage of students scoring in the normal range at posttreatment increased by 30.6% for the OST-T2 group versus 1.3% for controls (compared to baseline); the percentage normalized at 5-month follow-up increased by 33.3% for OST-T2 versus 0.8% for controls (compared to baseline), and the percentage normalized at 12-month follow-up increased by 19.1% for OST-T2 versus 10.4% for controls (compared to baseline).

Outcomes for Homework Performance

Caregiver Report. Students in OST-T2 responded more favorably compared to those in TAU/WL based on caregiver ratings on the total score of the HPC (adjusted mean difference = −0.40, $p < .001$, 95% CI [−0.17, −0.63]); see Supplemental Figure S2. The effect of intervention remained significant at 5-month follow-up (adjusted mean difference = −0.44, $p < .001$, 95% CI [−0.27, −0.61]). However, there was an attenuation of intervention effect at 12-month follow-up; Cohen's *d* at posttreatment, 5-month follow-up, and 12-month follow-up was 0.60, 0.72, and 0.27, respectively. The percentage of students scoring in the normal range at posttreatment increased by 35.3% for the OST-T2 group versus 8.0% for controls

Table 3
Mean Outcome Scores by Time Point Based on Actual, Unadjusted Data

Outcome/treatment group	Baseline	Post (14 weeks)	5-month follow-up (22 weeks)	12-month follow-up (55 weeks)
	M (*SD*)	*M* (*SD*)	*M* (*SD*)	*M* (*SD*)
COSS-P[a]				
OST-T2	2.64 (0.38)	2.23 (0.44)	2.23 (0.47)	2.29 (0.55)
TAU/WL	2.59 (0.35)	2.51 (0.38)	2.52 (0.34)	2.45 (0.39)
COSS-T[a]				
OST-T2	2.78 (0.39)	2.31 (0.49)	2.33 (0.54)	2.40 (0.42)
TAU/WL	2.80 (0.41)	2.74 (0.40)	2.75 (0.39)	2.51 (0.50)
HPC[a]				
OST-T2	1.48 (0.63)	0.98 (0.58)	0.97 (0.57)	1.10 (0.69)
TAU/WL	1.33 (0.58)	1.18 (0.53)	1.25 (0.57)	1.14 (0.58)
HPQ[b]				
OST-T2	2.03 (1.11)	2.75 (1.06)	2.59 (1.12)	2.31 (1.22)
TAU/WL	2.22 (0.96)	2.38 (0.96)	2.50 (0.91)	2.37 (1.20)
ACES[b]				
OST-T2	2.51 (0.68)	2.69 (0.76)	2.72 (0.70)	2.63 (0.79)
TAU/WL	2.35 (0.62)	2.36 (0.56)	2.37 (0.56)	2.50 (0.63)
APS[b]				
OST-T2	2.64 (0.69)	2.74 (0.75)	2.79 (0.76)	2.68 (0.80)
TAU/WL	2.55 (0.66)	2.52 (0.60)	2.50 (0.63)	2.57 (0.70)
PALS[b]				
OST-T2	3.92 (0.72)	4.04 (0.61)	4.19 (0.61)	4.10 (0.59)
TAU/WL	3.82 (0.71)	3.96 (0.72)	3.99 (0.79)	4.07 (0.67)
GRADES[b]				
OST-T2	2.49 (0.59)		2.75 (0.55)	
TAU/WL	2.56 (0.52)		2.72 (0.68)	

Note. The OST-T2 condition had a total of 97 participants, and the TAU/WL condition had a total of 88 participants. Academic grades assigned in the marking period corresponding to baseline and 5-month follow-up assessment are reported in this table. Standard procedures for harmonizing grades across schools resulted in scores of 1 = D/F, 2 = C, 3 = B, and 4 = A. COSS-P = Child Organizational Skills Scale–Parent report; OST-T2 = Organizational Skills Training–Tier 2; TAU/WL = treatment as usual with wait-list; COSS-T = Child Organizational Skills Scale–Teacher report; HPC = Homework Problem Checklist; HPQ = Homework Performance Questionnaire; ACES = Academic Competence Evaluation Scales; APS = Academic Proficiency Scale; PALS = Patterns of Adaptive Learning Scales.
[a] Higher scores on these measures are less favorable. [b] Higher scores on these measures are more favorable.

(compared to baseline); the percentage normalized at 5-month follow-up increased by 36.0% for OST-T2 versus 7.4% for controls (compared to baseline), and the percentage normalized at 12-month follow-up increased by 26.5% for OST-T2 versus 15.8% for controls (compared to baseline).

Teacher Report. Students in OST-T2 responded more favorably to OST-T2, compared to those in TAU/WL based on teacher ratings on the homework productivity and efficiency factor for the HPQ-T (adjusted mean difference = 0.67, $p = .007$, 95% CI [0.18, 1.16]); see Supplemental Figure S3. However, there was an attenuation of intervention effect at 5-month and 12-month follow-up. Cohen's *d* at posttreatment, 5-month follow-up, and 12-month follow-up was 0.64, 0.41, and 0.08, respectively. The percentage of students scoring in the normal range at posttreatment increased by 17.4% for the OST-T2 group versus 1.2% for controls (compared to baseline); the percentage normalized at 5-month follow-up increased by 10.5% for OST-T2 versus 0.3% for controls (compared to baseline); and the percentage normalized at 12-month follow-up increased by 8.9% for OST-T2 versus 1.6% for controls (compared to baseline).

Outcomes for Academic Performance

Teacher Report of Academic Skills. OST-T2 did not result in a significant improvement in teacher ratings of language arts and math skills assessed by the ACES at posttreatment, 5-month follow-up, and 12-month follow-up, although there was evidence of a potentially small effect at posttreatment. Cohen's *d* at posttreatment, 5-month follow-up, and 12-month follow-up was 0.31, 0.16, and 0.03, respectively. The percentage of students scoring in the normal range at posttreatment increased by 11.1% for OST-T2 versus a decline of 1.0% for controls (compared to baseline); the percentage normalized at 5-month follow-up increased by 23.5% for OST-T2 versus 3.0% for controls (compared to baseline); and the percentage normalized at 12-month follow-up increased by 12.1% for OST-T2 versus 6.4% for controls (compared to baseline).

Teacher Report of Overall Academic Proficiency. Although OST-T2 did not result in a significant improvement in teacher ratings on the APS at the adjusted α level of .0083 at any time point (posttreatment $p < .06$; 5-month follow-up $p < .03$), there was evidence of a potentially small effect at both time points. Cohen's *d* at posttreatment, 5-month follow-up, and 12-month follow-up was 0.31, 0.36, and 0.06, respectively. The percentage of students scoring in the normal range at posttreatment increased by 7.1% for OST-T2 versus 0.4% for controls (compared to baseline); the percentage normalized at 5-month follow-up increased by 14.0% for OST-T2 versus −0.8% for controls (compared to baseline); and the percentage normalized at 12-month follow-up increased by 14.1% for OST-T2 versus 2.9% for controls (compared to baseline).

Table 4
Threshold Analyses: Participants Scoring in the Clearly Normative Range by Time Point Based on Model-Adjusted Outcome Data

	Baseline *N* (%)		Post (14 weeks) *N* (%)		5-month follow-up (22 weeks) *N* (%)		12-month follow-up (55 weeks) *N* (%)	
Outcome measure	OST-T2	TAU/WL	OST-T2	TAU/WL	OST-T2	TAU/WL	OST-T2	TAU/WL
COSS-P	16 (16.8)	21 (23.9)	62 (71.3)	28 (34.6)	49 (67.1)	24 (35.3)	49 (61.3)	37 (46.8)
COSS-T	5 (5.3)	3 (3.5)	33 (35.9)	4 (4.8)	28 (36.8)	3 (4.3)	19 (24.4)	10 (13.9)
HPC	14 (14.7)	18 (20.5)	44 (50.0)	23 (28.4)	37 (50.7)	19 (27.9)	33 (41.3)	29 (36.3)
HPQ	0 (0.0)	1 (1.2)	16 (17.4)	2 (2.4)	8 (10.5)	1 (1.5)	7 (8.9)	2 (2.8)
ACES	15 (17.1)	10 (12.4)	24 (28.2)	9 (11.4)	28 (40.6)	10 (15.4)	21 (29.2)	13 (18.8)
APS	33 (34.7)	14 (16.5)	38 (41.8)	14 (16.9)	37 (48.7)	11 (15.7)	39 (48.8)	14 (19.4)

Note. The thresholds were determined based on the means and standard deviations of each measure in a normative sample as a function of child sex. The threshold for performing in the clearly normative range on the COSS-P, COSS-T, HPC, and HPQ was defined as participants whose scores were greater than/less than or equal to the normative mean score plus or minus 0.5 *SD*. The directionality (greater than vs. less than, plus vs. minus) was determined based on whether higher or lower scores represent more favorable outcomes. The following threshold values were used for each of these measures: COSS-P (Boys = 2.48; Girls = 2.21), COSS-T (boys = 2.36; girls = 1.94), HPC (boys = 1.03; girls = 0.72), and HPQ (boys = 3.49; girls = 3.64). For the academic performance measures, the threshold was based on scale anchors, whereby 3 indicates "at grade level" for ACES and "at the standard expected at this time of year" for APS. The percentages at each assessment point reflect the number of students scoring within the clearly normal range divided by the number of completed assessments at that time point. See Figure 1 for sample sizes for parent- and teacher-report measures at each time point. OST-T2 = Organizational Skills Treatment–Tier 2; TAU/WL = Treatment as usual with wait list; COSS-P = Child Organizational Skills Scale–Parent versions; COSS-T = Child Organizational Skills Scale–Teacher versions; HPC = Homework Problem Checklist; HPQ = Homework Performance Questionnaire; ACES = Academic Competence Evaluation Scales; APS = Academic Proficiency Scale.

Student Report Card Grades. Students in OST-T2 demonstrated minimal, if any, improvement in report card grades at the end of the school year (generally corresponding to 5-month follow-up) compared to those in the control condition. Cohen's *d* reflecting changes from the first to last marking period of the academic year of enrollment was 0.05 for combined math and reading/language arts grades (0.15 for math and 0.00 for reading/language arts).

Outcomes for Student Self-Efficacy

Student Self-Report of Academic Self-Efficacy. Students in OST-T2 did not report a significant improvement in their sense of self-efficacy on the Academic Efficacy subscale of the Patterns of Adaptive Learning Scales at any time point. Cohen's *d* at posttreatment, 5-month follow-up, and 12-month follow-up was 0.12, 0.13, and 0.11, respectively. Scores at baseline were high (>3.8), thereby limiting opportunities for improvement.

Sensitivity Analyses Accounting for Effects of COVID-19

Sensitivity analyses excluded all data collected after the shutdown of schools due to COVID-19. Doing so omitted posttreatment data for 16 participants in the OST-T2 group and 16 in the TAU/WL group, follow-up data for 28 participants in the OST-T2 group and 24 participants in TAU/WL, and all data for 11 participants (7 OST-T2, 4 TAU/WL) enrolled during the 2021–2022 school year. Excluding these participants resulted in the mean number of sessions attended being 15.2 (*SD* = 1.6). The direction and magnitude of effects on each outcome measure were similar to those identified using the intention-to-treat analyses, which included all data.

Acceptability, Feasibility, and Usability of OST-T2

Participant ratings provided evidence of the acceptability, feasibility, and usability of OST-T2. Caregiver ratings of the acceptability of OST-T2 at posttreatment, as assessed by the Treatment Evaluation Inventory–Short Form, were relatively high (*M* item score = 4.1, *SD* = 0.6; maximum possible score of 5.0), Similarly, child ratings of the acceptability of OST-T2 after treatment, as assessed by the Children's Intervention Rating Profile, were high (*M* item score = 4.3, *SD* = 0.6; maximum possible score of 5.0).

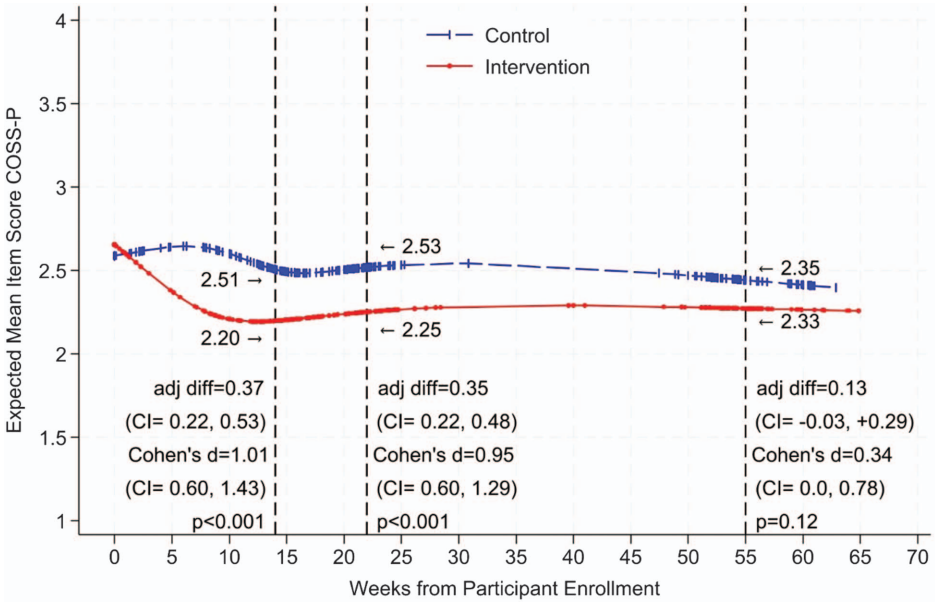
Teacher ratings ($n = 40$) on the Usage Rating Profile-Intervention–Revised were relatively high (maximum possible score of 6.0): Acceptability (*M* = 4.8, *SD* = 0.6), Feasibility (*M* = 4.4, *SD* = 0.8), and School Climate (*M* = 4.8, *SD* = 0.5). School partner ratings ($n = 13$) on the Usage Rating Profile-Intervention–Revised were also relatively high (maximum possible score of 6.0): Acceptability (*M* = 4.9, *SD* = 0.6), Feasibility (*M* = 4.3, *SD* = 0.8), and School Climate (*M* = 4.6, *SD* = 0.7).

Discussion

This study was designed to evaluate the effectiveness of school-based implementation of the OST-T2 intervention by school professionals for students with OTMP skills deficits in Grades 3–5 (ages 8–11) referred by teachers. The findings replicated the results of a clinic-based trial of OST-C (Abikoff et al., 2013) by demonstrating the effectiveness of this intervention in improving OTMP skills and homework performance. Most skills training interventions designed to improve student organizational skills have targeted students in middle and high school, who are ages 11–17 (DuPaul et al., 2021; Evans et al., 2016; Langberg et al., 2018), so this study and the study by Abikoff and colleagues are noteworthy in demonstrating that these programs can be effective with students as young as Grade 3. OTMP skills training programs are effective when delivered by highly trained clinicians (Abikoff et al., 2013) or school mental health professionals who are members of research teams (Langberg et al., 2018). This study corroborates the findings of an earlier study (Langberg et al., 2012) that demonstrated OTMP skills training interventions can be effective when delivered by school staff in school settings.

Figure 2

Model-Based Adjusted Mean COSS-P Item Scores at Each Assessment Period



Note. Lower scores reflect lower organizational skills deficits. Measurement completion varied somewhat across children leading to these outcome assessment estimations: 14 weeks from participant enrollment for posttreatment; 22 weeks from enrollment for 5-month follow-up; and 55 weeks from enrollment for 12-month follow-up. Markers on curves indicate mixed-effects model-based COSS-P mean scores at actual individual measurement times and curves join these markers. Figure notations indicate between-group differences of estimated change from baseline at the designated outcome assessment weeks, the associated confidence intervals, the corresponding Cohen's *d*, 95% confidence interval for Cohen's *d*, and *p* values. To account for multiple comparisons, differences in changes from baseline to posttreatment and follow-up were deemed statistically significant if $p < .083$. Between-group differences in changes over time are adjusted for baseline scores and do not necessarily equal within-group differences at each reported week of estimation. COSS-P = Children's Organizational Skills Scale–Parent version; CI = confidence interval; adj diff = difference in model-adjusted scores.

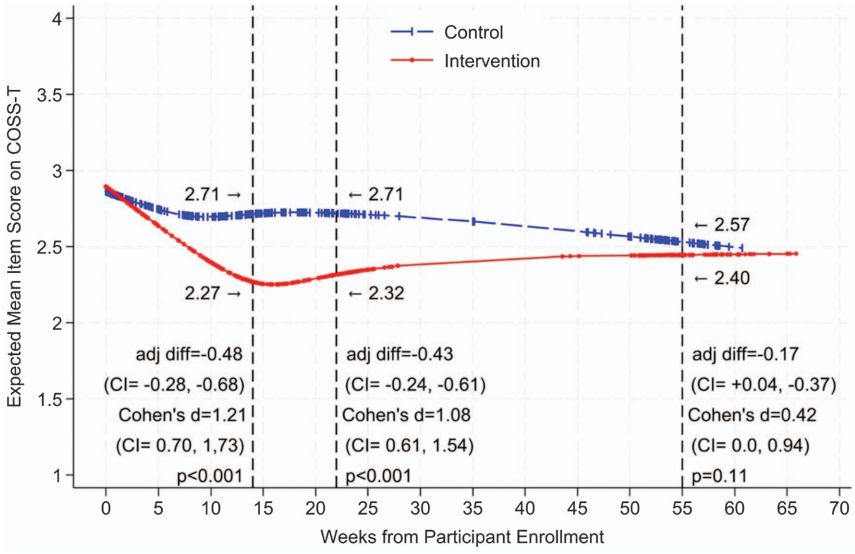
The findings indicate that OST-T2 can be delivered with high fidelity by school professionals in various roles across schools serving students from diverse socioeconomic and racial backgrounds. Further, most studies examining the effectiveness of skills training programs for students with OTMP skill deficits have focused on individual intervention and targeted children with ADHD. This study provides evidence that OST can be highly effective as a Tier 2, small-group intervention for students referred for OTMP deficits impacting academic performance who may not meet criteria for a diagnosis of ADHD or classification in special education. Indeed, teacher ratings of OTMP skills deficits at baseline for students in this study were comparable to ratings in a prior study of OST in elementary-age students with ADHD (Abikoff et al., 2013), although caregivers in the present study reported less OTMP skill impairment at baseline than the previous study. As such, the intervention tested in this study (OST-T2) can be effective for a broad range of students in Grades 3–5 with teacher-referred OTMP skills deficits without considering diagnostic status.

Immediate and Follow-Up Effects of OST-T2

OST-T2 succeeded in improving OTMP skills, which were the direct targets of intervention. Large effect sizes were demonstrated at posttreatment and 2 months after intervention, based on both caregiver and teacher report on the COSS. The degree of improvement in OTMP skills in this study compared to the clinic-based, individually delivered, more intensive OST-C intervention (Abikoff et al., 2013) was similar in magnitude of effect based on teacher report (ES = 1.08 this study; ES = 1.18 Abikoff et al.,'s 2013 study) and lower based on caregiver report (ES = 0.95 this study; ES = 2.77 Abikoff et al.,'s 2013 study). The larger effect on caregiver ratings of OTMP skills in the OST-C study may have been due to multiple factors, including the higher level of caregiver involvement in OST-C, likely resulting in more practice at home, as well as the more impaired sample in the prior study reflected in higher caregiver ratings of OTMP skills deficits at baseline compared to this study. In addition, OST-T2 resulted in a substantial improvement in homework performance; medium to large effects were demonstrated on the caregiver-rated HPC at posttreatment and 8 weeks afterward, and medium effects were shown on the teacher-rated HPQ-T (although short-term follow-up effects were not significant). The degree of improvement in homework performance in this study, compared to OTMP intervention for middle school students (Langberg et al., 2018), was higher in magnitude of effect based on teacher ratings and lower based on caregiver ratings.

Figure 3

Model-Based Adjusted Mean COSS-T Item Scores at Each Assessment Period



Note. The note for Figure 2 applies to Figure 3. COSS-T = Children's Organizational Skills Scale–Teacher version; CI = confidence interval; adj diff = adjusted difference or the difference adjusted for model corrections.

The impact of OST-T2 on academic performance was limited. The effect on overall teacher ratings of academic proficiency (APS) was small (ES = 0.31 at posttreatment; 0.36 at 5-month follow-up), and comparable to the OST-C study at posttreatment (Abikoff et al., 2013), although it did not reach statistical significance. The APS likely reflects overall teacher impressions of student performance related to both academic skills and enablers (i.e., efficiency, productivity). In contrast, there appeared to be little, if any, effect of OST-T2 on teacher-reported academic skills in math and reading/language arts as assessed by the ACES. Further, OST-T2 did not demonstrate an effect on overall report card grades, which stands in contrast to an improvement in grades, albeit modest, for older students in response to OTMP interventions (Evans et al., 2016; Langberg et al., 2018). Although the difference in findings across studies may indicate that OST-T2 is less effective in improving academic grades than OTMP interventions for older students, differences in grading systems and how these are applied in middle and high school versus elementary school may have contributed to differential effects. Schools in this study differed substantially in how grades were assigned, the scales used in grading, and frequency of grading. In addition, specific academic components were evaluated in some marking periods and not others. As such, it is likely the APS was a more accurate indicator of overall academic performance than grades. To summarize, the findings of this study suggest that OST-T2 may result in modest, short-term gains in teacher perceptions of overall student academic performance (assessed by APS), but not acquisition of academic skills and performance on report cards.

A relatively high percentage of students receiving OST-T2 were functioning well within normal limits on OTMP skills rated by caregivers at posttreatment and 5-month follow-up (71%, 67%, respectively). A lower percentage were rated as functioning within normal limits by teachers at these time points (36%, 37%). The percentage of students receiving OST-T2 who were normalized on homework performance at posttreatment and 5-month follow-up showed a similar pattern (50%, 51% by caregiver report on HPC; 17%, 11% by teacher report on HPQ, respectively). It was noted that teachers generally viewed students as more impaired on OTMP skills at baseline than caregivers. Thus, although caregivers and teachers generally perceived large improvements in OTMP skills and medium improvement in homework performance after intervention, teachers perceived that a high percentage of students continued to have organizational and homework problems. Many students failing to respond sufficiently to OST-T2 may need more intensive, Tier 3 intervention.

There was a reduction in the effect of OST-T2 at 12-month follow-up. Although effects were not statistically significant at 12-month follow-up at conventional thresholds, effect sizes (Cohen's *d*) on both versions of the COSS and the HPC were in the small to medium range (0.27–0.42). In addition, rates of normalization on these measures at 12-month follow-up were higher for those in the OST-T2 than the TAU/WL condition, although the difference between groups was relatively small (5%–15%). It is worth noting that there was an effect size of 0.42 on the COSS-T, which was completed during the school year following the intervention by teachers who did not participate in OST-T2 and were likely unaware of condition assignment.

There was a greater attenuation of effects at 12-month follow-up in this study than in the OST-C study. This may be due in part to the higher dosage of OST-C compared to OST-T2 (20 vs. 11 hr; individual vs. group format), which offered students more opportunities to practice skills, as well as greater caregiver involvement. In addition, the attenuation of effects at 12-month follow-up in this study suggests the need for additional sessions after the focused intervention period has ended, including opportunities for child skills practice and

consultation with teachers and caregivers to encourage and praise skill use. Indeed, there is evidence that more and continual opportunities to practice skills are associated with better treatment outcomes and maintenance of effects (Evans et al., 2021; Schultz et al., 2017). Research is needed to test the long-term benefits of increasing the amount of practice during and after OST-T2. Given the evidence of modest benefits of OST-T2 at 12-month follow-up and the chronicity of OTMP difficulties for many students, especially those with ADHD (Molina et al., 2009), providing students with continual opportunities to practice OTMP skills throughout their schooling would likely be useful.

The small effects of OST-T2 on academic skills are not surprising given that OST-T2 does not target the acquisition of these skills. Many students with academic skills deficits who receive OST-T2 will likely need additional academic support. However, given evidence that OTMP skills are associated with future homework and academic achievement (Langberg et al., 2013), which reflects academic skills and enablers, and that the association between homework performance and academic achievement increases in later school years (Cooper et al., 2006), there may be long-term academic benefits to receiving OST at the elementary level. The long-term academic benefits of early organizational skills training require additional research.

Implementation of OST-T2 by School Professionals

OST-T2 generally was implemented with a high level of fidelity by school professionals, including counselors and general and special education teachers. Overall, high fidelity was achieved after 6–8 hr of initial training and twice weekly consultation offered to school partners by postdoctoral level research team members. During consultation sessions, delivery of the OST-T2 curriculum to students was frequently adjusted to accommodate school partner preferences and time constraints. Although fidelity was generally high, it was somewhat lower for the Planning module, suggesting that students received less of the Planning content. Planning is covered at the end of intervention, when students often needed more time to practice skills for modules introduced earlier. In addition, there was a fair amount of variability across school partners and modules of OST-T2. Understanding factors contributing to the variability in school partner fidelity is an important area for future research.

Participants in OST-T2 generally had favorable views of the program. School partners expressed high interest in providing OST-T2, and indicated the program was generally feasible and was a good fit with the values and priorities of the school. Teachers' views about the acceptability, feasibility, and usability of OST-T2 were similar to those of school partners. Importantly, students and their caregivers also viewed OST-T2 favorably. These findings suggest that school-based organizational skills training is valued and viewed as appropriate to be offered by school professionals. Nonetheless, school partners expressed some concerns about the feasibility of consultation supports and, to a lesser extent, the intervention itself and offered suggestions to enhance feasibility and sustainability. Consultation and coaching meetings were burdensome given school partners' limited prep time, and managing the large number of materials needed to implement OST-T2 was challenging. Their feedback suggested that with streamlining of the consultation process and implementation materials, follow-up practice sessions might be feasible and effective.

Generalizability of Findings, Limitations, and Future Directions

Participating schools were from a large metropolitan area in the northeastern United States. All school professionals, including principals, school partners, and teachers, participated voluntarily. To recruit a school, we first obtained the approval of the principal. Although many principals expressed strong interest in participating, some indicated that doing so was not feasible. For schools where the principal offered approval, informed consent was obtained from at least one school professional who could deliver OST-T2 and serve as a liaison with the study team. Several schools were not able to identify anyone to assume this role. Teachers were given the choice of whether to nominate students for participation in the study and whether to complete assessments. As such, the findings of this study are generalizable to the subset of schools in a large metro area for which OST-T2 was viewed as aligned with their school mission, feasible to deliver, and could be implemented by a school professional.

Students participating in this study were diverse with respect to race/ethnicity and socioeconomic status, addressing to some extent the need for intervention research that includes a substantial portion of children and families from marginalized backgrounds (Evans et al., 2018). Nonetheless, children identified by their caregivers as Black or White were overrepresented (35% and 66%, respectively), and those identified as Hispanic/Latino were underrepresented (12%) compared to recent national census data. In general, schools serving students across the full socioeconomic spectrum were included in the study, although schools serving a high proportion of children living in poverty or deep poverty were underrepresented.

Although prestudy power calculations supported the adequacy of the study design to detect meaningful levels of improvement across the outcome measures, the onset of the COVID-19 pandemic resulted in school and student participation challenges that limited the final sample size. Another limitation is that the results may have been subject to rater bias. It is possible that caregivers and teachers, who were aware of intervention condition, were biased to report more favorable outcomes for OST-T2, leading to inflated posttreatment effect sizes (Sonuga-Barke et al., 2013). However, it is questionable whether the large effects found at posttreatment and 5-month follow-up (approximately 1.0) could be accounted for entirely or largely due to bias. Ideally, outcome assessments would have included observations of student behavior in school, although such assessments are highly challenging to develop and validate and can be time consuming and costly to implement. Further, the magnitude of the effect of intervention on OTMP skills at 12-month follow-up, as rated by teachers who were likely unaware of treatment assignment, was near the medium range. Nonetheless, research to identify objective, ecologically valid measures of outcome that can be administered in a relatively efficient manner is needed (Abikoff et al., 2013; Evans et al., 2018). Additionally, the timing of student report card grades did not align closely with assessment periods. As such, examining changes in report card grades from the first to the last marking period was limited as an approach to assessing intervention effects. It should also be noted that the threshold for normality used in this study was

functioning well within normal limits. The use of a higher threshold (e.g., near clinical) would have resulted in differences in the proportion of students identified in the normal range.

There are numerous directions for future research. First, it is important to identify whether the effects of OST-T2 vary for subgroups defined by demographic (e.g., age, socioeconomic status) and clinical characteristics (e.g., severity of ADHD and anxiety symptoms). It is also important to explore whether and how changes in OTMP skills impact students' general adjustment, the level of family conflict they experience due to baseline OTMP deficits, and other related behaviors and emotions. Second, additional research should explore the mechanisms underlying the effect of OST-T2 on homework and academic outcomes. Third, future research is needed to understand patterns that may explain variations in fidelity and the degree to which school partner fidelity, as well as student, caregiver, and teacher engagement in OST-T2 affect outcomes. Fourth, additional research is needed to determine whether any additional adaptations to OST-T2 are needed to be responsive to current demands placed on students to use electronic systems to manage and complete assigned tasks. Finally, it is important to identify specific implementation factors to improve acceptability and feasibility of the training and consultation approach and the intervention. This will result in improved understanding of the potential for sustaining OST-T2 with limited support from consultants.

Conclusions

This study replicates many of the findings of a previous study (Abikoff et al., 2013) demonstrating the immediate and short-term effectiveness of OST in improving OTMP skills and homework performance among students ages 8–11. This study extends previous research by demonstrating that OST is effective when delivered in a school setting by well-trained and supported school professionals as a small-group, Tier 2 intervention. OST-T2 demonstrated minimal effect on academic skill acquisition, suggesting that students with OTMP deficits and academic skills problems need targeted academic skills training in addition to a program like OST-T2. The effects of OST-T2 were attenuated, although still perhaps meaningful 9 months after intervention when students advanced to the next grade level, suggesting the need for follow-up sessions providing additional opportunities to practice skills. Additional research is needed to investigate strategies to improve long-term effects and examine whether OST-T2 can be effective under routine school conditions when school partners receive training and support aligned with authentic practice.

References

Abikoff, H. (2009). ADHD psychosocial treatments: Generalization reconsidered. *Journal of Attention Disorders*, *13*(3), 207–210. https://doi.org/10.1177/1087054709333385

Abikoff, H., & Gallagher, R. (2009). *Children's organizational skills scale: Manual*. Multihealth Systems.

Abikoff, H., Gallagher, R., Wells, K. C., Murray, D. W., Huang, L., Lu, F., & Petkova, E. (2013). Remediating organizational functioning in children with ADHD: Immediate and long-term effects from a randomized controlled trial. *Journal of Consulting and Clinical Psychology*, *81*(1), 113–128. https://doi.org/10.1037/a0029648

Anesko, K. M., Schoiock, G., Ramirez, R., & Levine, F. M. (1987). The Homework Problem Checklist: Assessing children's homework difficulties. *Behavioral Assessment*, *9*, 179–185.

Beidas, R. S., & Kendall, P. C. (2010). Training therapists in evidence-based practice: A critical review of studies from a systems-contextual perspective. *Clinical Psychology: Science and Practice*, *17*(1), 1–30. https://doi.org/10.1111/j.1468-2850.2009.01187.x

Best, J. R., Miller, P. H., & Jones, L. L. (2009). Executive functions after age 5: Changes and correlates. *Developmental Review*, *29*(3), 180–200. https://doi.org/10.1016/j.dr.2009.05.002

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences*, *21*(4), 327–336. https://doi.org/10.1016/j.lindif.2011.01.007

Briesch, A. M., Chafouleas, S. M., Neugebauer, S. R., & Riley-Tillman, T. C. (2013). Assessing influences on intervention implementation: Revision of the usage rating profile-intervention. *Journal of School Psychology*, *51*(1), 81–96. https://doi.org/10.1016/j.jsp.2012.08.006

Chacko, A., Bedard, A. C., Marks, D. J., Feirsen, N., Uderman, J. Z., Chimiklis, A., Rajwan, E., Cornwell, M., Anderson, L., Zwilling, A., & Ramon, M. (2014). A randomized clinical trial of Cogmed working memory training in school-age children with ADHD: A replication in a diverse sample using a control condition. *Journal of Child Psychology and Psychiatry*, *55*(3), 247–255. https://doi.org/10.1111/jcpp.12146

Cooper, H., Robinson, J. C., & Patall, E. A. (2006). Does homework improve academic achievement? A synthesis of research, 1987–2003. *Review of Educational Research*, *76*(1), 1–62. https://doi.org/10.3102/00346543076001001

Crisci, G., Caviola, S., Cardillo, R., & Mammarella, I. C. (2021). Executive functions in neurodevelopmental disorders: Comorbidity overlaps between attention deficit and hyperactivity disorder and specific learning disorders. *Frontiers in Human Neuroscience*, *15*, Article 594234. https://doi.org/10.3389/fnhum.2021.594234

Di Bartolo, C. A., Abikoff, H., McDonald, M., & Gallagher, R. (2021). Adapting an empirical clinic-based organizational skills training program for use in schools: Iterative development and pilot assessment. *Psychology in the Schools*, *58*(7), 1225–1245. https://doi.org/10.1002/pits.22530

DiPerna, J. C., & Elliott, S. N. (2000). *Academic competence evaluation scales*. The Psychological Corporation.

Dumas, J. E., Lynch, A. M., Laughlin, J. E., Phillips Smith, E., & Prinz, R. J. (2001). Promoting intervention fidelity. Conceptual issues, methods, and preliminary results from the EARLY ALLIANCE prevention trial. *American Journal of Preventive Medicine*, *20*(Suppl. 1), 38–47. https://doi.org/10.1016/S0749-3797(00)00272-5

DuPaul, G. J., Evans, S. W., Owens, J. S., Cleminshaw, C. L., Kipperman, K., Fu, Q., & Benson, K. (2021). School-based intervention for adolescents with attention-deficit/hyperactivity disorder: Effects on academic functioning. *Journal of School Psychology*, *87*, 48–63. https://doi.org/10.1016/j.jsp.2021.07.001

Dweck, C. S., & Elliott-Moskwa, E. S. (2010). Self-theories: The roots of defensiveness. In J. E. Maddux & J. P. Tangney (Eds.), *Social psychological foundations of clinical psychology* (pp. 136–153). Guilford Press.

Efron, B., & Morris, C. (1973). Stein's estimation rule and its competitors—An empirical Bayes approach. *Journal of the American Statistical Association*, *68*(341), 117–130. https://doi.org/10.2307/2284155

Eiraldi, R., Khanna, M., Jawad, A. F., Power, T. J., Cacia, J., Cabello, B., Schwartz, B. S., Swift, L., Kanine, R., Orapallo, A., McCurdy, B., & Mautone, J. A. (2020). Implementation of targeted mental health interventions in urban schools: Preliminary findings for impact of training strategy on program fidelity. *Evidence-Based Practice in Child and Adolescent Mental Health*, *5*(4), 437–451. https://doi.org/10.1080/23794925.2020.1784056

Evans, S. W., Allan, D., Xiang, J., Margherio, S. M., Owens, J. S., & Langberg, J. M. (2021). Organization interventions as a mediator of

change in grades in the Challenging Horizons Program. *Journal of School Psychology*, *87*, 18–27. https://doi.org/10.1016/j.jsp.2021.05.001

Evans, S. W., Langberg, J. M., Schultz, B. K., Vaughn, A., Altaye, M., Marshall, S. A., & Zoromski, A. K. (2016). Evaluation of a school-based treatment program for young adolescents with ADHD. *Journal of Consulting and Clinical Psychology*, *84*(1), 15–30. https://doi.org/10.1037/ccp0000057

Evans, S. W., Owens, J. S., Wymbs, B. T., & Ray, A. R. (2018). Evidence-based psychosocial treatments for children and adolescents with attention deficit/hyperactivity disorder. *Journal of Clinical Child and Adolescent Psychology*, *47*(2), 157–198. https://doi.org/10.1080/15374416.2017.1390757

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2011). *Applied longitudinal analysis* (2nd ed.). Wiley. https://doi.org/10.1002/9781119513469

Gallagher, R., Abikoff, H., & Spira, E. (2014). *Organizational skills training for children with ADHD: An empirically supported treatment*. Guilford Press.

Goldstein, S., & Naglieri, J. A. (2009). *Autism spectrum rating scales*. Multi-Health Systems.

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, *42*(2), 377–381. https://doi.org/10.1016/j.jbi.2008.08.010

Kelley, M. L., Heffer, R. W., Gresham, F. M., & Elliott, S. N. (1989). Development of a modified treatment evaluation inventory. *Journal of Psychopathology and Behavioral Assessment*, *11*(3), 235–247. https://doi.org/10.1007/BF00960495

Kind, A. J. H., & Buckingham, W. R. (2018). Making neighborhood disadvantage metrics accessible—The Neighborhood Atlas. *The New England Journal of Medicine*, *378*(26), 2456–2458. https://doi.org/10.1056/NEJMp1802313

Kofler, M. J., Irwin, L. N., Soto, E. F., Groves, N. B., Harmon, S. L., & Sarver, D. E. (2019). Executive functioning heterogeneity in pediatric ADHD. *Journal of Abnormal Child Psychology*, *47*(2), 273–286. https://doi.org/10.1007/s10802-018-0438-2

Langberg, J. M., Dvorsky, M. R., & Evans, S. W. (2013). What specific facets of executive function are associated with academic functioning in youth with attention-deficit/hyperactivity disorder? *Journal of Abnormal Child Psychology*, *41*(7), 1145–1159. https://doi.org/10.1007/s10802-013-9750-z

Langberg, J. M., Dvorsky, M. R., Molitor, S. J., Bourchtein, E., Eddy, L. D., Smith, Z. R., Oddo, L. E., & Eadeh, H. M. (2018). Overcoming the research-to-practice gap: A randomized trial with two brief homework and organization interventions for students with ADHD as implemented by school mental health providers. *Journal of Consulting and Clinical Psychology*, *86*(1), 39–55. https://doi.org/10.1037/ccp0000265

Langberg, J. M., Epstein, J. N., Becker, S. P., Girio-Herrera, E., & Vaughn, A. J. (2012). Evaluation of the homework, organization, and planning skills (HOPS) intervention for middle school students with attention deficit hyperactivity disorder as implemented by school mental health providers. *School Psychology Review*, *41*(3), 342–364. https://doi.org/10.1080/02796015.2012.12087514

Langberg, J. M., Molina, B. S. G., Arnold, L. E., Epstein, J. N., Altaye, M., Hinshaw, S. P., Swanson, J. M., Wigal, T., & Hechtman, L. (2011). Patterns and predictors of adolescent academic achievement and performance in a sample of children with attention-deficit/hyperactivity disorder. *Journal of Clinical Child and Adolescent Psychology*, *40*(4), 519–531. https://doi.org/10.1080/15374416.2011.581620

Loux, T., & Huang, Y. (2023). The uses of propensity scores in randomized controlled trials. *Observational Studies*, *9*(1), 77–85. https://doi.org/10.1353/obs.2023.0007

Lund, J. I., Toombs, E., Radford, A., Boles, K., & Mushquash, C. (2020). Adverse childhood experiences and executive function difficulties in children: A systematic review. *Child Abuse & Neglect*, *106*, Article 104485. https://doi.org/10.1016/j.chiabu.2020.104485

Midgley, C., Kaplan, A., Middleton, M., Maehr, M. L., Urdan, T., Anderman, L. H., Anderman, E., & Roeser, R. (1998). Development and validation of scales assessing students' achievement goal orientation. *Contemporary Educational Psychology*, *23*(2), 113–131. https://doi.org/10.1006/ceps.1998.0965

Molina, B. S. G., Hinshaw, S. P., Swanson, J. M., Arnold, L. E., Vitiello, B., Jensen, P. S., Epstein, J. N., Hoza, B., Hechtman, L., Abikoff, H. B., Elliott, G. R., Greenhill, L. L., Newcorn, J. H., Wells, K. C., Wigal, T., Gibbons, R. D., Hur, K., Houck, P. R., & the MTA Cooperative Group. (2009). The MTA at 8 years: Prospective follow-up of children treated for combined-type ADHD in a multisite study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *48*(5), 484–500. https://doi.org/10.1097/CHI.0b013e31819c23d0

O'Hearn, K., Asato, M., Ordaz, S., & Luna, B. (2008). Neurodevelopment and executive function in autism. *Development and Psychopathology*, *20*(4), 1103–1132. https://doi.org/10.1017/S0954579408000527

Pfiffner, L. J., Rooney, M., Haack, L., Villodas, M., Delucchi, K., & McBurnett, K. (2016). A randomized controlled trial of a school-implemented school-home intervention for attention-deficit/hyperactivity disorder symptoms and impairment. *Journal of the American Academy of Child & Adolescent Psychiatry*, *55*(9), 762–770. https://doi.org/10.1016/j.jaac.2016.05.023

Power, T. J., Dombrowski, S. C., Watkins, M. W., Mautone, J. A., & Eagle, J. W. (2007). Assessing children's homework performance: Development of multi-dimensional, multi-informant rating scales. *Journal of School Psychology*, *45*(3), 333–348. https://doi.org/10.1016/j.jsp.2007.02.002

Power, T. J., Werba, B. E., Watkins, M. W., Angelucci, J. G., & Eiraldi, R. B. (2006). Patterns of parent-reported homework problems among ADHD-referred and non-referred children. *School Psychology Quarterly*, *21*(1), 13–33. https://doi.org/10.1521/scpq.2006.21.1.13

Reynolds, C., & Kamphaus, R. (2015). *Behavior assessment for children* (3rd ed.). Pearson.

Rinsky, J. R., & Hinshaw, S. P. (2011). Linkages between childhood executive functioning and adolescent social functioning and psychopathology in girls with ADHD. *Child Neuropsychology*, *17*(4), 368–390. https://doi.org/10.1080/09297049.2010.544649

Ryan, S., Power, T. J., Pendergast, L., Poznanski, B., Nissley-Tsiopinis, J., Abikoff, H., Gallagher, R., Tremont, K., Cacia, J., & Mautone, J. A. (2024). Multi-informant assessment of organizational skills: Psychometric characteristics of the Children's Organizational Skills Scale (COSS). *School Mental Health*, *16*(2), 354–369. https://doi.org/10.1007/s12310-024-09646-w

Schultz, B. K., Evans, S. W., Langberg, J. M., & Schoemann, A. M. (2017). Outcomes for adolescents who comply with long-term psychosocial treatment for ADHD. *Journal of Consulting and Clinical Psychology*, *85*(3), 250–261. https://doi.org/10.1037/ccp0000172

Sibley, M. H., Graziano, P. A., Kuriyan, A. B., Coxe, S., Pelham, W. E., Rodriguez, L., Sanchez, F., Derefinko, K., Helseth, S., & Ward, A. (2016). Parent-teen behavior therapy + motivational interviewing for adolescents with ADHD. *Journal of Consulting and Clinical Psychology*, *84*(8), 699–712. https://doi.org/10.1037/ccp0000106

Slick, D. J., Lautzenhiser, A., Sherman, E. M. S., & Eyrl, K. (2006). Frequency of scale elevations and factor structure of the Behavior Rating Inventory of Executive Function (BRIEF) in children and adolescents with intractable epilepsy. *Child Neuropsychology*, *12*(3), 181–189. https://doi.org/10.1080/09297040600611320

Sonuga-Barke, E. J., Brandeis, D., Cortese, S., Daley, D., Ferrin, M., Holtmann, M., Stevenson, J., Danckaerts, M., van der Oord, S., Döpfner, M., Dittmann, R. W., Simonoff, E., Zuddas, A., Banaschewski, T., Buitelaar, J., Coghill, D., Hollis, C., Konofal, E., Lecendreux, M., … the European ADHD Guidelines Group. (2013). Nonpharmacological interventions for ADHD: Systematic review and meta-analyses of

randomized controlled trials of dietary and psychological treatments. *The American Journal of Psychiatry*, *170*(3), 275–289. https://doi.org/10.1176/appi.ajp.2012.12070991

Sugai, G., & Horner, R. R. (2006). A promising approach for expanding and sustaining school-wide positive behavior support. *School Psychology Review*, *35*(2), 245–259. https://doi.org/10.1080/02796015.2006.12087989

Swanson, J. M., Kraemer, H. C., Hinshaw, S. P., Arnold, L. E., Conners, C. K., Abikoff, H. B., Clevenger, W., Davies, M., Elliott, G. R., Greenhill, L. L., Hechtman, L., Hoza, B., Jensen, P. S., March, J. S., Newcorn, J. H., Owens, E. B., Pelham, W. E., Schiller, E., Severe, J. B., … Wu, M. (2001). Clinical relevance of the primary findings of the MTA: Success rates based on severity of ADHD and ODD symptoms at the end of treatment. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*(2), 168–179. https://doi.org/10.1097/00004583-200102000-00011

U.S. Census Bureau. (2023). *Census Bureau releases new educational attainment data*. https://www.census.gov/newsroom/press-releases/2023/educational-attainment-data.html

Witt, J. C., & Elliott, S. N. (1985). Acceptability of classroom intervention strategies. In T. R. Kratochwill (Ed.), *Advances in school psychology* (Vol. 4, pp. 251–288). Lawrence Erlbaum.

Received January 2, 2024
Revision received June 4, 2024
Accepted June 18, 2024 ■