ARCHAIA JOURNAL OF SCIENCE AND TECHNOLOGY
VOL. 3, NO. 4, November 2026
https://doi.org/10.48161/ajst.v3n4a99

Creative Arts Therapy for Children with Special Needs: Educational and Health Outcomes

Marwa Mehmood[1*]

[1] Department of Physical Therapy & Rehabilitation Sciences, Faculty of Allied Health Sciences, The Superior University, Lahore, Punjab, Pakistan;

* **Corresponding author:** marwamehmood19@gmail.com.

ABSTRACT: Creative arts therapies are increasingly used in educational and paediatric services, yet controlled evidence remains limited for accessible multimodal programmes evaluated against active comparators and linked to educational, psychosocial, implementation, and family outcomes. This study estimated the effectiveness and implementation of the 12-week Creative Regulation, Expression, Agency, Togetherness, and Engagement (CREATE) programme for children aged 7–12 years with neurodevelopmental disabilities and participation or self-regulation difficulties. In a multicentre, pragmatic, hybrid Type-1 cluster-randomized trial, researchers allocated 48 classrooms in 12 inclusive and specialist schools to twice-weekly multimodal creative-arts therapy or a time- and attention-matched non-art developmental program. Twenty-four credentialed providers each delivered one classroom in each condition in different schools or terms. We assessed the 396 children at baseline, week 6, week 12, three months, six months, and 12 months. Co-primary outcomes were the standard School Function Assessment Participation criterion score, completed by masked observers, and the standard PedsQL Emotional Functioning subscale, completed primarily by parents. Intention-to-treat cross-classified multilevel models incorporated child, classroom, school, and provider effects, multiple imputation, small-sample correction, and multiplicity control. At week 12, CREATE was associated with higher school participation (+4.6 points, 95% CI 2.1–7.1; g=.32) and emotional functioning (+3.4, 95% CI 1.0–5.8; g=.24). Participation remained modestly higher at six months but was uncertain at 12 months; emotional-functioning differences were uncertain after six months. Emotion regulation, peer relationships, and educational goal attainment showed the clearest secondary differences, whereas attendance, service use, and sleep did not differ reliably. Exploratory process analyses identified associations rather than confirmed mechanisms. The findings provide preliminary programme-level support for CREATE as an educational adjunct, without modality-specific, diagnosis-specific, or durable health claims.

Keywords: creative arts therapy; neurodevelopmental disabilities; school participation; emotional functioning; mixed methods.

I. INTRODUCTION

Children with neurodevelopmental disabilities face participation barriers arising from interactions among communication, sensory regulation, motor planning, executive control, social reciprocity, classroom demands, family resources, and service accessibility. Creative-arts programs offer alternative routes to communication, relationships, and learning, but the evidence base remains methodologically uneven. A scoping review of performance and visual-arts programmes for children with disabilities reported encouraging psychosocial findings while emphasizing small samples, heterogeneous interventions, and limited evaluation of ordinary service delivery [1]. In this study, we operationalized the broad title descriptor

as documented neurodevelopmental disabilities accompanied by participation or self-regulation difficulties; the conclusions do not extend to every child receiving additional educational support.

Music therapy represents the most developed creative modality in neurodevelopmental research. Systematic reviews report possible advantages for communication and social interaction, but pooled estimates vary with diagnosis, comparator strength, dose, and study quality [2], [3]. Large controlled trials demonstrate why cautious interpretation is necessary. The TIME-A trial did not identify a group difference in broad autism symptom severity despite strong engagement [4]. Smaller randomized studies reported changes in social communication, auditory-motor connectivity, or proximal interaction following music-based intervention [5], [6]. The collective pattern favours domain-specific participation and regulation outcomes over claims of generalized developmental change.

Movement, dance, and embodied programmes provide different opportunities through rhythm, posture, imitation, shared timing, and non-verbal reciprocity. Pilot and systematic evidence supports feasibility and possible wellbeing effects while consistently noting small samples and heterogeneous delivery [7], [8]. School-based dance movement psychotherapy has shown acceptable delivery for children with emotional and behavioural difficulties [9]. Family-based movement-music programs have produced encouraging findings on social communication [10]. These studies also show that therapist expertise, family involvement, and multimodal content can complicate attribution to specific components.

Visual art, storytelling, and dramatic activity broaden the range of accessible expressive forms. Reviews across autism, attention-deficit/hyperactivity disorder, language disorders, and learning disabilities identify potential effects on communication, attention, behaviour, emotional regulation, and occupational participation, but also substantial variation in intervention definition and outcome quality [11], [12]. Peer-mediated theatre trials have reported gains in social competence and reductions in anxiety [13], [14]. Their specialist and diagnosis-specific formats limit conclusions about inclusive school implementation and children with heterogeneous communication and sensory profiles.

Passive controls do not separate creative content from adult attention, predictable routine, facilitator warmth, expectancy, peer contact, and structured reflection. Active controls provide a more conservative estimate of the added value of creative and embodied processes, but remain uncommon. Accessibility and safety matter just as much: sound, movement, visual complexity, public performance, and adult interpretation can support one child while burdening another. Child assent, the right to observe or stop, augmentative communication, sensory adaptation, and protection from compulsory disclosure are part of intervention quality, not optional accommodations.

Measurement weaknesses further limit the literature. Researchers often infer educational participation from symptom change rather than observing it directly; parent-proxy ratings may dominate several outcome domains; school-functioning items can overlap with separate participation measures; and implementation costs, adverse responses, and transfer into ordinary classroom routines are often underreported. Researchers also interpret mediation and subgroup findings with more certainty than the available numbers of schools, classrooms, and providers allow.

The present study evaluated a manualized multimodal service rather than a superior art form. CREATE combined visual art, music and rhythm, movement, and story-based dramatic activity within a repeated sequence of regulation, choice, cooperative creation, reflection, and classroom transfer. The active comparator matched adult attention, contact time, group structure, problem solving, and self-management practice without creative-arts content. Classroom allocation reduced contamination arising from shared teachers, peers, materials, and transfer routines. The study contributed an active-controlled evaluation of an accessible program using a standard masked participation outcome, a standard emotional-functioning outcome, crossed-provider modeling, 12-month observation, and integrated evidence on implementation, cost, adverse experience, and transfer. This design prioritized independent measurement and conservative interpretation over claims of broad therapeutic transformation.

II. RESEARCH PROBLEM, AIM, OBJECTIVES, AND QUESTIONS

1. RESEARCH PROBLEM

Schools and child-development services lack active-controlled evidence establishing whether an accessible multimodal creative-arts therapy programme adds educational and emotional value beyond structured attention, adult support, group routine, and non-art developmental activity. The central problem was to estimate the effect of the full CREATE program while separating provider variance from program assignment, using non-overlapping standard outcomes, documenting accessibility and adverse responses, and avoiding attribution to individual art modalities.

2. RESEARCH AIM AND OBJECTIVES

The study aimed to estimate the programme-level effect of the complete CREATE service on standard masked school participation and standard PedsQL emotional functioning at week 12, examine persistence through 12 months, and explain selective educational, developmental, family, implementation, safety, and resource findings relative to an attention- and time-matched active control. This aim was operationalized through six objectives:

- Estimate the effects of the complete CREATE programme on standard masked school participation and PedsQL Emotional Functioning at week 12, with persistence assessed at three, six, and 12 months.
- Examine communication, adaptive behaviour, emotion regulation, peer relationships, parent-reported participation, educational goal attainment, quality of life, attendance, and sleep.
- Assess caregiver stress, health and educational service use, co-interventions, programme costs, and resource implications through 12 months.
- Explore temporally ordered associations involving creative engagement, autonomy, competence, emotion regulation, and peer connectedness, together with a restricted set of prespecified contextual moderators; these analyses remained non-confirmatory.
- Evaluate reach, dose, provider and classroom fidelity, accessibility, acceptability, adaptations, contamination, cost, and adverse experiences across classrooms and schools.
- Integrate child, parent, educator, therapist, observer, educational-record, service-use, and qualitative evidence to explain transfer, divergence, negative cases, and boundary conditions.

3. RESEARCH QUESTIONS

- RQ1. What programme-level effects occurred in standard masked school participation and PedsQL Emotional Functioning at week 12, three months, six months, and 12 months?
- RQ2. What differences occurred in communication, adaptive behaviour, emotion regulation, peer relationships, parent-reported participation, educational goal attainment, quality of life, attendance, and sleep?
- RQ3. What differences occurred in caregiver stress, health, educational service use, co-interventions, program costs, and other resource outcomes through 12 months?
- RQ4. What exploratory associations linked creative engagement, autonomy, competence, emotion regulation, peer connectedness, and prespecified child or contextual characteristics with variation in outcomes?
- RQ5. How successfully was CREATE implemented with respect to reach, provider and classroom fidelity, accessibility, acceptability, adaptation, contamination, cost, and safety?
- RQ6. How did children, parents, educators, therapists, and school leaders explain benefits, limited responses, adverse experiences, and transfer into everyday educational participation?

III. LITERATURE REVIEW

1. EVIDENCE ACROSS CREATIVE MODALITIES

The literature supports a differentiated rather than uniformly positive account of creative-arts therapy. Music-based interventions have generated the largest controlled evidence base, with effects most often

concentrated in reciprocal communication, shared attention, synchrony, and proximal social behaviour. Reviews describe favourable average patterns but substantial heterogeneity in dose, therapist orientation, comparator strength, and outcome selection [2], [3]. The contrast between the null diagnostic result in the large TIME-A trial and more favourable communication findings in smaller mechanistic studies indicates that broad symptom scales may be less sensitive than participation-focused outcomes [4], [5]. A controlled comparison of active music therapy and music listening further suggests that interactional structure matters, but does not establish that music itself is uniquely responsible [6].

Movement and dance interventions add embodied timing, imitation, proprioceptive feedback, and non-verbal coordination. Pilot studies show feasible engagement and possible wellbeing changes, but systematic synthesis remains limited by small samples and varied intervention definitions [7], [8]. School-based dance movement psychotherapy is especially relevant because it delivers within educational routines rather than specialist clinics [9]. Family-based movement-music work suggests that caregiver participation may strengthen transfer, although combined content and family involvement prevent clear attribution [10]. These limitations favor evaluating a coherent service package with transparent boundaries rather than making retrospective claims about a single active ingredient.

Visual art, storytelling, and theatre can reduce immediate verbal demand and permit symbolic distance from difficult experiences. Reviews of visual-art interventions report possible gains in communication, social interaction, attention, and occupational participation, but frequently identify weak comparators, incomplete intervention description, and inconsistent measurement [11], [12]. Theatre trials provide stronger evidence for rehearsed social participation and anxiety-related outcomes [13], [14]. Their structured roles and public performance demands also highlight possible adverse experiences, particularly for children with sensory sensitivity, performance anxiety, or limited communication access.

2. *ACCESSIBILITY, ASSENT, AND OUTCOME SELECTION*

Accessibility is not a secondary implementation concern. Multimodal sessions can widen participation by offering visual, auditory, movement, narrative, observational, and augmentative routes, but the same variety can increase sensory load and unpredictability. Ethical delivery therefore depends on continuing child assent, permission to observe or stop, adjustable volume and visual complexity, communication supports, adapted materials, and alternatives to public sharing. Negative responses such as fatigue, frustration, peer conflict, or perceived judgement are informative implementation outcomes rather than evidence to be omitted from an aff`irmative account of arts engagement.

Outcome selection determines what an intervention appears to achieve. Diagnostic symptom scales are often too distal to capture changes in classroom entry, sustained involvement, communication of choice, cooperative contribution, or transition. Parent report remains important but can create shared-informant bias when used across multiple primary domains. Direct masked observation provides a more independent educational endpoint, while a standard emotional-functioning subscale avoids constructing a new composite that overlaps with school participation. Goal attainment can add individual relevance when anchors are prespecified and independently rated. Attendance, sleep, caregiver stress, service use, and adverse experience remain necessary because selective educational gains do not automatically indicate general health improvement [1], [9].

3. *ACTIVE COMPARATORS, IMPLEMENTATION, AND THE REMAINING GAP*

An active comparator strengthens causal interpretation by matching attention, time, routine, choice, peer contact, and adult support. It cannot blind participants or remove all expectancy effects, but it narrows the estimated contrast to the coordinated creative, symbolic, sensory, embodied, and transfer architecture of the programme. Provider effects also require explicit treatment. When different professional groups deliver each condition, treatment and provider characteristics are confounded. A crossed-provider arrangement, with trained providers delivering both conditions under separate manuals and supervision, improves separation of programme and provider variance while preserving condition integrity [4], [6].

The remaining evidence gap concerns whether an accessible multimodal programme adds measurable educational value beyond credible non-art developmental activity, whether any effect persists after

structured delivery ends, and which implementation conditions accompany benefit or burden. The present trial addressed this gap through classroom randomization, pre-allocation recruitment, crossed providers, standard co-primary outcomes, 12-month follow-up, structured safety and cost assessment, and explanatory mixed methods. The intervention remained a single programme-level estimand: the design did not isolate visual art, music, movement, drama, or the transfer routine as independent components.

IV. CREATIVE PARTICIPATION–DEVELOPMENT FRAMEWORK

The Creative Participation–Development Framework combined three established perspectives while keeping them analytically distinct from the study-specific process measures. The participation perspective within the International Classification of Functioning treats functioning as an interaction among child capacities, activities, participation, environmental supports, and barriers. Childhood-disability scholarship similarly positions participation as both a developmental process and a meaningful outcome [15]. We therefore assessed school participation directly using a standard instrument completed by masked observers rather than inferring it from emotional change.

Self-Determination Theory identifies autonomy, competence, and relatedness as conditions supporting motivation and wellbeing [16]. CREATE operationalized these principles through choice of medium and role, adjustable challenge, permission to observe or stop, and cooperative tasks that permitted contribution without compulsory performance. The authors examined autonomy and competence as exploratory process constructs, while peer connectedness represented a relational process rather than a confirmed mediator.

Transactional developmental theory emphasizes reciprocal influence among children, caregivers, educators, peers, providers, and environments over time [17]. The same task can support or burden a child depending on sensory load, communication access, adult interpretation, classroom timing, staffing stability, and prior experience. Diagnosis was treated as one contextual descriptor rather than a sufficient explanation of response. We examined educator uptake, family resources, accessibility adaptations, provider fidelity, and negative cases as transfer and boundary conditions.

The framework specified temporal ordering without implying a deterministic causal chain. We assessed creative engagement, autonomy, and competence during early exposure; emotion regulation and peer connectedness at week 6; and outcomes at week 12 and follow-up. The process scales had been evaluated in a separate pretrial measurement sample, but trial analyses remained exploratory because the processes were not randomized and independent replication was unavailable. CREATE was evaluated as one coordinated programme; no result was attributed separately to visual art, music, movement, or dramatic storytelling.

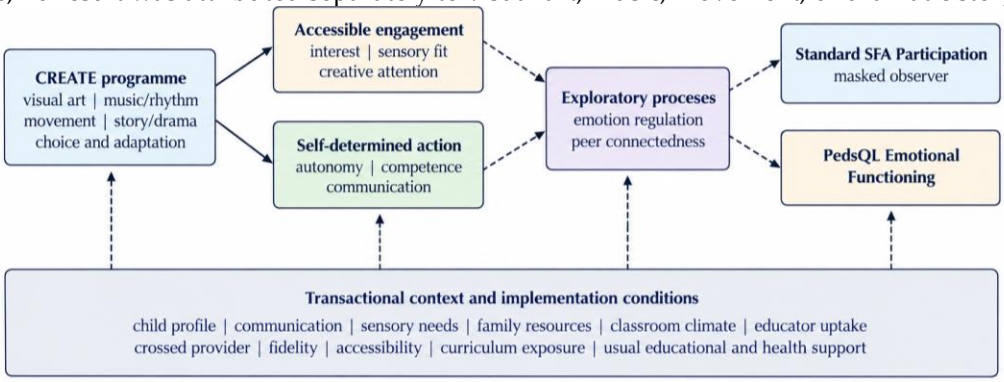


FIGURE 1. Creative Participation-Development Framework. Solid arrows identify programme exposure; dashed arrows identify exploratory process associations. No arrow establishes a unique causal mechanism.

V. METHODOLOGY

1. DESIGN, SETTING, AND REPORTING LOGIC

The study used a multicentre pragmatic hybrid Type-1 effectiveness-implementation cluster-randomized controlled design with an embedded longitudinal explanatory mixed-methods evaluation. The hybrid design prioritized effectiveness while collecting structured implementation evidence [18]. CREATE was treated as a complex, context-dependent service under contemporary Medical Research Council guidance [19]. Reporting followed the CONSORT extension for cluster-randomized trials [20] and TIDieR intervention-reporting guidance [21]. The treatment-policy estimand compared assignment to the complete CREATE programme with active control regardless of attendance, permitted adaptations, or subsequent non-study support.

Twelve schools in four education districts participated: six inclusive mainstream schools with specialist resource bases and six specialist schools. We recruited 48 classrooms serving children aged 7–12 years before allocation. Baseline enrolment, parent consent, child assent, and outcome assessment preceded randomization. Twenty-four credentialed creative-arts therapists were crossed across conditions: each delivered one CREATE classroom and one active-control classroom in a different school or academic term. We balanced condition order, and separate manuals, materials, and supervision prevented content drift. Provider identity was recorded for cross-classified analysis. Fieldwork covered one academic year of recruitment and delivery, with follow-up extending to 12 months.

Effectiveness outcomes were primary. Implementation was structured using the updated Consolidated Framework for Implementation Research [22] and interpreted through RE-AIM domains [23]. We reported findings on educational participation, emotional functioning, implementation, safety, and resources as distinct domains to avoid extending a selective program effect into a generalized health claim.

2. PARTICIPANTS, ELIGIBILITY, RECRUITMENT, AND ASSENT

Children were eligible if they were aged 7–12 years, had a documented neurodevelopmental disability or multidisciplinary developmental formulation, attended a participating classroom, and showed persistent participation or self-regulation difficulties documented by the educational team and parent. Included profiles comprised autism spectrum disorder, attention-deficit/hyperactivity disorder, developmental language disorder, and mild intellectual disability. Children with co-occurring motor, sensory, or anxiety difficulties were not excluded when safe access could be arranged.

Exclusion criteria were acute psychiatric or safeguarding crisis, uncontrolled seizure activity that precluded group participation, severe medical instability, an immediate need for a more intensive treatment programme, or inability to participate safely after reasonable adaptation. Augmentative and alternative communication, mobility support, sensory accommodations, and interpreter assistance were permitted. Families received accessible written and visual information. Parents or legal guardians provided written consent; children provided developmentally appropriate written, pictorial, or recorded assent and were reminded that they could withdraw assent at any session without affecting educational or health support.

School staff identified potentially eligible children using a standardized checklist, after which a research practitioner independently confirmed eligibility, independent of classroom assignment. Recruitment logs recorded eligibility, invitations, consent, assent, and reasons for non-participation. Children remained in follow-up if they chose not to attend particular sessions because the estimand concerned classroom assignment rather than receipt of a minimum dose.

3. RANDOMIZATION, CONCEALMENT, MASKING, AND CONTAMINATION

An independent statistician prepared the allocation schedule using variable blocks within school, stratified by school type and mean baseline support need. The secure central system implemented allocation after baseline data were locked. Therapists, children, parents, and classroom educators necessarily knew allocation because programme content was visible. Masked assessors completed standardized School Function Assessment observations, data abstractors used condition-neutral codes, and the primary statistician remained masked until the co-primary models were finalized. Assessors inferred allocation

correctly in 54% of audited visits. Early credibility and expected-helpfulness ratings did not differ materially between conditions. We nevertheless interpreted parent- and educator-reported outcomes with potential expectancy and reporting bias in mind.

CREATE and active-control delivery used condition-specific manuals, materials, preparation checklists, and supervision. Crossed providers delivered the two conditions in different schools or terms and did not prepare both conditions concurrently. Ordinary curriculum art and music continued in both conditions, and we prospectively documented them. We recorded cross-condition staff movement, shared materials, external creative therapies, and newly organized arts activities. Sensitivity analyses adjusted for curriculum exposure and excluded control classrooms with substantial non-study creative activity.

4. SAMPLE-SIZE DETERMINATION

The target was 384 children across 48 classrooms. Calculations assumed eight children per classroom, classroom intraclass correlation of .05, correlation of .60 among repeated observations, 15% attrition by six months, two co-primary week-12 outcomes tested at two-sided alpha=.025, and adjustment for baseline score, school, and school type. The design provided approximately 80% power to detect d=.30 at week 12 [24]. The final sample of 396 exceeded the individual target but did not increase the numbers of classrooms or providers. The trial was powered for the two primary programme contrasts, not for mediation, provider heterogeneity, moderation, or diagnosis-specific effects; those estimates were restricted to exploratory reporting.

5. INTERVENTION AND ACTIVE COMPARATOR

CREATE comprised 24 group sessions delivered twice weekly for 12 weeks, each lasting 45 minutes. Groups contained six to nine children and were led by a credentialed creative-arts therapist with a classroom support practitioner. Sessions followed four stable functions: an accessible regulation entry; child-led selection of medium or role; individual or cooperative creation; and a brief transfer routine linked to a classroom participation target. The study integrated visual art, rhythm and music, movement, and story-based dramatic activity as a single manualized service rather than separable treatment components.

The active control matched session frequency, duration, provider contact, group size, predictable opening and closing routines, cooperative activity, choice, and guided reflection. Content comprised non-art problem solving, structured reading, tabletop cooperative games, organization routines, and age-adapted self-management practice. The study prohibited creative production, music-making, rhythmic entrainment, movement improvisation, role play, image-making, and therapeutic storytelling. Usual educational, occupational-therapy, speech-language, psychological, medical, and family support continued in both conditions. Because the modalities were bundled, every effectiveness estimate concerned the complete CREATE service and not an individual art form.

Table 1. Programme and active-control specifications.

Domain	CREATE programme	Active developmental control
Therapeutic architecture	Regulation entry; choice; creation; reflection; transfer	Predictable entry; choice of non-art task; problem solving; reflection
Core content	Visual art, music/rhythm, movement, story and drama	Reading, cooperative games, planning, non-art self-management
Dose	24 × 45 min over 12 weeks	24 × 45 min over 12 weeks
Group format	6–9 children; crossed credentialed therapist + classroom support practitioner	6–9 children; same crossed-provider pool + classroom support practitioner
Accessibility	AAC, visual schedule, sensory breaks, adapted tools, seated/standing options, observation role	Equivalent communication, sensory, motor, and pacing adaptations

Domain	CREATE programme	Active developmental control
Child agency	Choice of medium, role, partner, privacy, sharing, and stopping	Choice of task, partner, response mode, and stopping
Prohibited practices	Diagnostic interpretation of artwork; compulsory disclosure or performance; display without consent	Creative arts activities, guided imagery, therapeutic storytelling, performance
Transfer	Teacher-selected classroom cue and weekly participation target	Teacher-selected classroom cue and weekly participation target
Fidelity threshold	≥85% core-function adherence and no prohibited practice	≥85% core-function adherence and no prohibited practice
Provider competence / supervision	Credentialed therapist; one CREATE and one control classroom in different schools/terms; condition-specific supervision	Same credentialed provider pool; non-art manual; comparator-separation supervision

Both conditions were delivered alongside usual educational and health support. Adaptations were permitted when they preserved core functions and condition separation.

6. THERAPIST TRAINING, ACCESSIBILITY, FIDELITY, AND SAFETY

Twenty-four credentialed creative-arts therapists delivered both conditions under the crossed-provider arrangement; each provider led one CREATE classroom and one active-control classroom in different schools or terms. Providers held postgraduate or accredited professional qualifications in at least one creative-arts therapy discipline. The program protocol restricted cross-modal activities to manualized, low-risk participatory exercises. Providers did not present themselves as licensed in another discipline, interpret creative products diagnostically, or deliver modality-specific psychotherapy outside their competence. All completed 18 hours of manual, safeguarding, communication-access, comparator-separation, and adverse-response training, followed by condition-specific fortnightly supervision. Classroom support practitioners completed a six-hour orientation.

Fidelity combined provider checklists, session logs, independent observation of 25% of sessions, material audits, educator transfer records, and cross-condition drift checks. Observers rated core-function delivery, accessibility, child choice, pacing, emotional safety, scope-of-practice boundaries, and prohibited content. Twenty percent of observation records were independently double-rated. We coded adaptations as sensory, communication, motor, cognitive, cultural, relational, or scheduling changes. Adverse experiences included sensory overload, fatigue, frustration, perceived judgement, peer conflict, emotional activation, physical discomfort, and safeguarding concerns. We reviewed serious or unexpected events independently of program delivery.

7. OUTCOMES, MEASURES, AND DATA SOURCES

The first co-primary outcome was the School Function Assessment Participation scale completed by trained masked observers after repeated standardized classroom observations. The instrument's standard criterion score was used on its native 0–100 metric; no study-specific rescaling or transformation was applied. This score summarizes participation across routine school settings rather than isolated task performance. Its construct validity and ability to differentiate school participation have been documented in elementary-school samples [25]. Double-scored observations assessed inter-rater reliability. The parent-rated PEM-CY school section was retained as a supportive secondary perspective.

The second co-primary outcome was the standard PedsQL Emotional Functioning subscale, scored from 0 to 100 according to the published instrument algorithm. Parents were the principal informants; child report was collected when reliable and analyzed as a complementary sensitivity perspective. The standard subscale avoided overlap with the School Function Assessment and avoided constructing a new emotional-social composite. Evidence supports cautious adapted PedsQL use with intellectual and developmental disabilities

[26]. No population-specific minimum clinically important difference was assumed for this subscale; interpretation relied on adjusted mean differences, standardized effects, confidence intervals, and prespecified responder distributions.

Secondary child outcomes included the Children's Communication Checklist-2 General Communication Composite, Vineland-3 Adaptive Behavior Composite, BRIEF-2 Emotion Regulation Index, PROMIS Pediatric Peer Relationships Short Form 8a, standardized Goal Attainment Scaling, parent-rated PEM-CY school participation, PedsQL total score, PROMIS Pediatric Sleep Disturbance Parent Proxy Short Form 8a, and school attendance. Goal Attainment Scaling used explicit five-level anchors and independent blinded scoring because goal quality can influence estimates [27]. Parent outcomes included the Parenting Stress Index-Fourth Edition Short Form. Service records covered speech-language, occupational therapy, psychology, paediatric, emergency, medication, and non-study therapy contacts.

Exploratory process measures assessed creative engagement, perceived autonomy, competence, emotion regulation, and peer connectedness using visual response options, supported communication, and observer ratings. Before the trial, a separate measurement team evaluated the scales in 218 children from eight non-participating schools, with no overlap in children, schools, or providers. Confirmatory factor analysis, two-week test-retest reliability, convergent correlations with established regulation and peer measures, and invariance across communication mode and school type supported preliminary measurement adequacy. The scales remained exploratory because we did not establish external replication, criterion validity, or diagnosis-specific invariance.

Table 2. Outcomes, primary informants, scoring, and measurement schedule.

Construct	Measure / source	Direction	Timing	Analytical role
School participation	School Function Assessment Participation criterion score; masked observer primary; PEM-CY parent report supportive	Standard criterion score; 0–100 higher better	T0–T5	Co-primary at T2
Emotional functioning	PedsQL Emotional Functioning subscale; parent proxy primary; child report supportive	Standard score; 0–100 higher better	T0–T5	Co-primary at T2
Communication	CCC-2 General Communication Composite; parent/educator informants	Standard score; higher better	T0, T2, T4, T5	Secondary
Adaptive behaviour	Vineland-3 Adaptive Behavior Composite; parent interview	Standard score; higher better	T0, T2, T4, T5	Secondary
Emotion regulation	BRIEF-2 Emotion Regulation Index; parent/educator	T-score; lower better	T0–T5	Secondary
Peer relationships	PROMIS Pediatric Peer Relationships SF 8a; child/parent	T-score; higher better	T0–T5	Secondary
Educational goals	Goal Attainment Scaling; independent blinded rating	T-score; higher better	T0, T2–T4	Secondary
Quality of life	PedsQL total; parent proxy and accessible child report	0–100; higher better	T0–T5	Secondary
Sleep	PROMIS Pediatric Sleep Disturbance Parent Proxy SF 8a	T-score; lower better	T0, T2, T4, T5	Secondary
Caregiver stress	PSI-4-SF; parent report	Higher worse	T0, T2, T4, T5	Family outcome

Construct	Measure / source	Direction	Timing	Analytical role
Attendance / service use	School, health, and education-service records	Days / contacts	Continuous through T5	Secondary
Implementation	Provider logs, audits, independent observations, costing	Reach, dose, fidelity, access, cost, safety	Continuous	Implementation
Exploratory process constructs	Externally separated pretrial validation sample; trial child/support/observer formats	Standardized scores	Weeks 3, 6, 12	Exploratory only

T0: baseline; T1: week 6; T2: week 12; T3: three months; T4: six months; T5: 12 months. Co-primary outcomes were tested at T2.

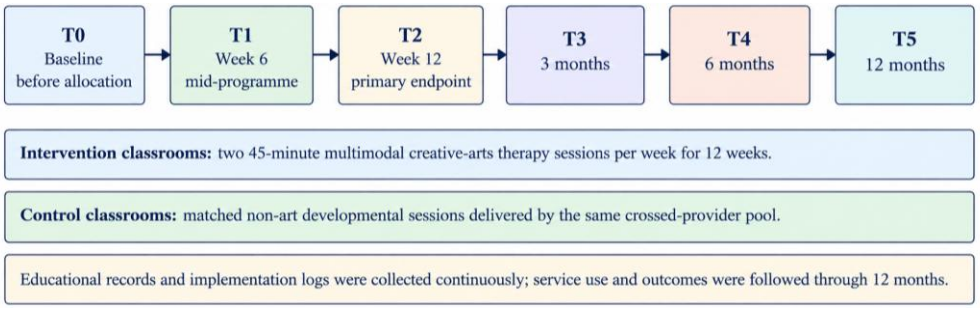


FIGURE 2. Cluster allocation, crossed-provider delivery, and assessment schedule. Baseline enrolment and assessment preceded classroom randomization; follow-up extended to 12 months.

8. *QUANTITATIVE ANALYSIS*

Analyses followed the intention-to-treat principle. For child i in classroom j, school k, and provider h at occasion t, the principal continuous-outcome model included allocation, categorical time, allocation-by-time interaction, baseline outcome, school fixed effects, school type, age, communication support, and baseline support need. Random intercepts represented child, classroom, and provider, with provider crossed with condition and school; within-child residuals followed an autoregressive structure. Kenward-Roger degrees of freedom were used for the 48 randomized classrooms. Binary and count outcomes used generalized mixed models with appropriate links. Effects are reported as adjusted mean differences, Hedges g, 95% confidence intervals, exact p values, classroom and provider variance components, and observed denominators.

Each co-primary week-12 contrast used two-sided alpha=.025. Secondary child outcomes formed one false-discovery-rate family; parent, family, service, and resource outcomes formed a separate descriptive family. A five-point School Function Assessment difference was prespecified as a pragmatic group-level interpretation benchmark. No clinical threshold was imposed on the PedsQL Emotional Functioning subscale. Forty multilevel chained-equation imputations incorporated school, classroom, crossed provider, allocation, all outcome occasions, child and family characteristics, support needs, attendance, co-interventions, assessment mode, and recorded reasons for missingness [28], [29]. We reviewed trace plots, distributional overlap, Monte Carlo error, relative efficiency, and fractions of missing information.

Exploratory process models examined temporal associations from early creative engagement, autonomy, and competence to week-6 emotion regulation and peer connectedness and then to week-12 outcomes. Indirect estimates used cluster-robust small-sample uncertainty and were interpreted under the strong, unverifiable assumptions required for mediation analysis [30], [31]. Moderator analyses were restricted to diagnosis group, communication support, sensory sensitivity, baseline support need, prior arts experience, family resource strain, and classroom fidelity. The 48-classroom design did not support confirmatory mediation, provider-effect, moderator, or diagnosis-specific inference; complete estimates were reported only in the Supplementary Material.

Sensitivity analyses included complete cases, no-imputation models, classroom fixed effects, school random effects, crossed-provider omission and alternative variance structures, per-protocol contrasts,

curriculum-arts adjustment, exclusion of contaminated control classrooms, informant-specific models, inverse-probability retention weighting, and leave-one-classroom-out estimation. We examined model convergence, residual distributions, school, classroom, and provider variance, collinearity, covariance fit, imputation diagnostics, and influential clusters before interpretation.

9. QUALITATIVE AND MIXED-METHODS EVALUATION

Maximum-variation sampling included 48 children, 32 parents, 28 educators, all 24 crossed providers, and 12 school leaders from all 12 schools. Child interviews used visual elicitation, object-based prompts, scaling cards, drawing choice, augmentative and alternative communication, and segmented sessions; median total interview time was 23 minutes (IQR 16–31). Median durations were 42 minutes for parents (IQR 34–53), 39 minutes for educators (IQR 31–48), 55 minutes for providers (IQR 46–66), and 48 minutes for school leaders (IQR 40–59). Sampling varied diagnosis, age, communication mode, sensory profile, support need, school type, provider, fidelity, attendance, and quantitative response, including limited change and adverse experience. Information power guided sample adequacy [32].

Framework Analysis combined structured comparison with openness to unanticipated processes [33]. A child-development psychologist, an arts-therapy researcher, a special-education researcher, and a parent-advisory analyst independently coded 16 contrasting cases before developing a versioned matrix linking child, family, classroom, provider, implementation, and outcome evidence. Reflexive memos recorded disciplinary assumptions and positive expectations about arts engagement. A communication-access specialist audited supported interviews, and a second analyst checked 20% of quotations against audio, AAC output, observation notes, and contextual records. Published quotations represented all 12 schools, both school types, several communication modes, and positive, neutral, and adverse trajectories. Negative cases altered the final interpretation by treating refusal, sensory fit, and educator transfer as active programme conditions rather than exceptions.

The quantitative and qualitative strands were connected through purposive sampling, common case identifiers, interim case summaries, and joint displays of research questions. Joint displays examined convergence, complementarity, expansion, and disagreement [34], and integration followed established mixed-methods principles [35]. Qualitative evidence explained experience and context but did not replace effect estimates. Reporting followed qualitative and mixed-methods standards [36], [37].

10. ETHICS, SAFEGUARDING, CONFIDENTIALITY, AND DATA GOVERNANCE

The coordinating Child Development and Education Research Ethics Committee approved the protocol before recruitment (approval 041), and participating education authorities completed governance review. Parents or legal guardians provided written informed consent, and children provided developmentally appropriate written, pictorial, verbal, or recorded assent. Consent and assent covered questionnaires, educational records, observations, interviews, audio recording, and de-identified quotation; separate permission governed photography or display of creative products. Participation was voluntary, refusal did not affect education or care, and children could observe, change an activity, take a sensory break, or stop while remaining in outcome follow-up.

We stored names and contact details separately from coded research records. Electronic files were encrypted on access-controlled institutional systems; paper records were held in locked storage. Role-based permissions, audit logs, retention schedules, and secure transfer procedures governed access. Quotations and descriptions of creative products were screened for deductive identification. Providers followed safeguarding, adverse-response, and referral procedures, and an independent clinician reviewed serious or unexpected events without access to outcome trends. The cluster trial was prospectively registered in the Institutional Education and Child Development Trials Register. The protocol and analysis plan prespecified the 12-month assessment.

https://doi.org/10.48161/ajst.v3n4a99

VI. RESULTS

1. CLUSTER AND PARTICIPANT FLOW

All 48 classrooms completed the study. Of 524 children screened, 422 met eligibility criteria, 26 families declined, and 396 completed baseline assessment before allocation. Twenty-four classrooms containing 199 children were assigned to CREATE and 24 classrooms containing 197 children to active control. We observed week-6 data for 380 children (96.0%), week-12 data for 365 (92.2%), three-month data for 347 (87.6%), six-month data for 331 (83.6%), and 12-month data for 304 (76.8%). Twelve-month observations included 153 CREATE and 151 control children. Missing follow-up was associated with school transfer, higher baseline support need, and unstable family contact, but not independently with allocation or baseline co-primary scores. All 396 children contributed to multiply imputed intention-to-treat models.

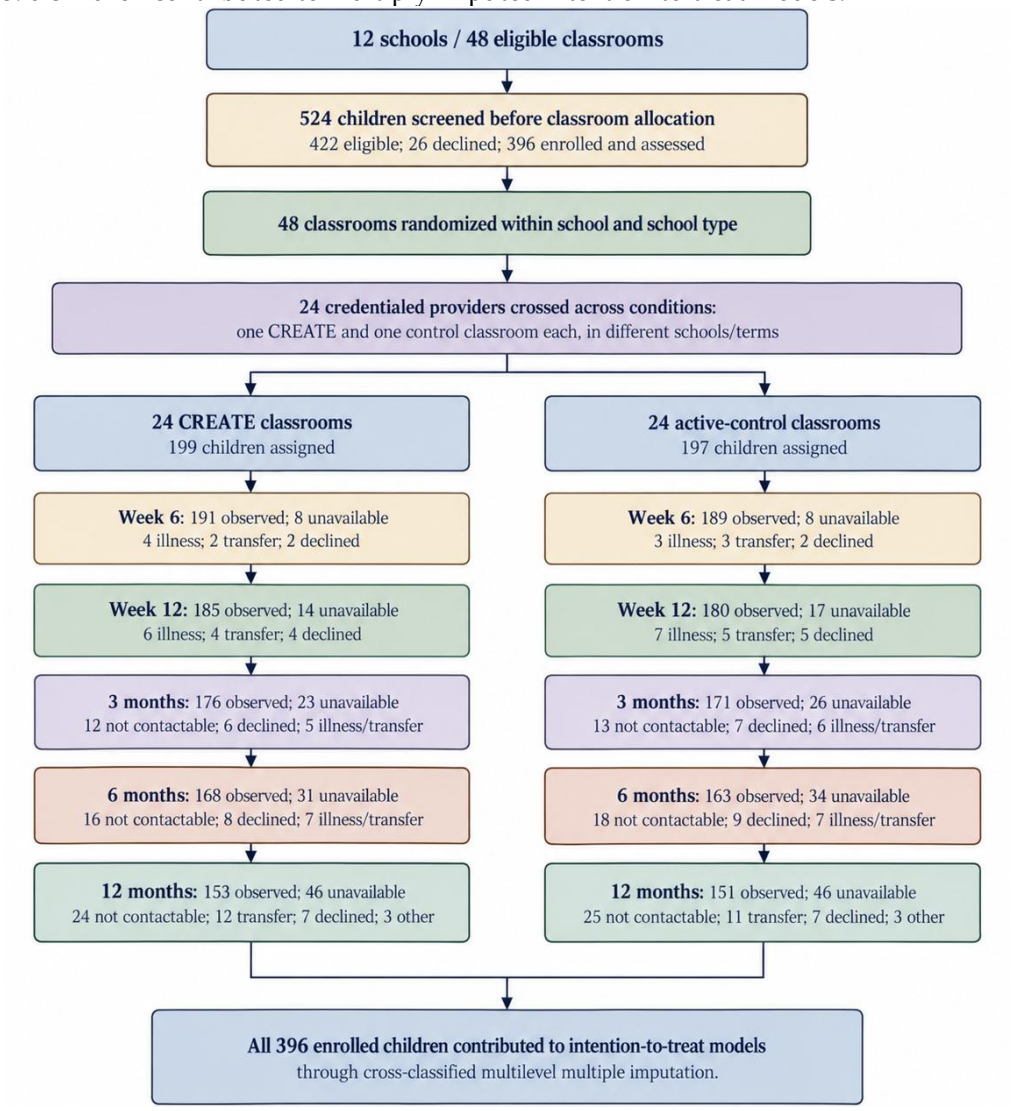


FIGURE 3. Cluster-randomized study flow. We recruited children and conducted baseline assessments before classroom allocation; all enrolled children remained in the intention-to-treat estimand.

2. BASELINE CHARACTERISTICS AND MEASUREMENT QUALITY

Children had a mean age of 9.6 years (SD=1.6); 34.6% were girls, 63.6% boys, and 1.8% used another description or preferred not to respond. Autism was the principal documented diagnosis for 42.2%, ADHD

for 24.7%, developmental language disorder for 18.2%, and mild intellectual disability for 14.9%. Co-occurring sensory and anxiety difficulties were common, and 18% used AAC. Baseline child, family, classroom, school-type, and provider-distribution characteristics were closely balanced, with absolute standardized mean differences below .10. Any-item baseline missingness was 3.0% in CREATE and 2.5% in control.

Omega coefficients for established multi-item measures ranged from .79 to .93. Masked School Function Assessment ratings showed inter-rater ICC=.84 across double-scored observations, and blinded Goal Attainment anchors showed ICC=.82. The independent pretrial process-scale sample showed acceptable factor fit (CFI=.945, TLI=.938, RMSEA=.044, SRMR=.046), omega coefficients of .78-.89, two-week ICCs of .72-.84, and small-to-moderate convergent correlations with established regulation and peer measures. Configural and metric invariance were retained across communication mode and school type. These findings supported exploratory use but did not establish external replication, diagnostic invariance, or criterion validity.

Table 3. Baseline classroom, child, family, and provider characteristics.

Characteristic	CREATE (n=199)	Active control (n=197)	SMD
Classrooms / schools	24 / 12	24 / 12	—
Children, n	199	197	—
Age, mean (SD)	9.6 (1.6)	9.7 (1.5)	−0.06
Girls, n (%)	70 (35.2)	67 (34.0)	0.03
Autism, n (%)	85 (42.7)	82 (41.6)	0.02
ADHD, n (%)	48 (24.1)	50 (25.4)	−0.03
Developmental language disorder	36 (18.1)	36 (18.3)	−0.01
Mild intellectual disability	30 (15.1)	29 (14.7)	0.01
AAC use, n (%)	37 (18.6)	35 (17.8)	0.02
High sensory sensitivity	76 (38.2)	75 (38.1)	0.00
Moderate/high support need	158 (79.4)	154 (78.2)	0.03
Standard SFA Participation criterion score, mean (SD)	46.8 (14.3)	47.1 (14.0)	−0.02
PedsQL Emotional Functioning, mean (SD)	59.6 (14.5)	60.0 (14.2)	−0.03
BRIEF-2 emotion regulation T	67.1 (9.4)	66.8 (9.6)	0.03
Prior organized arts participation	71 (35.7)	69 (35.0)	0.01
Family resource strain, high	58 (29.1)	60 (30.5)	−0.03
Inclusive / specialist classrooms	12 / 12	12 / 12	—
Mean classroom enrolment, mean (SD)	8.3 (1.1)	8.2 (1.0)	0.09
Crossed credentialed providers, n	24	24	—
Any baseline item missing, n (%)	6 (3.0)	5 (2.5)	0.03

SMD: standardized mean difference. Absolute values below .10 indicate close balance.

3. *EXPOSURE, FIDELITY, ACCESSIBILITY, COST, AND SAFETY*

CREATE children attended a median of 22 of 24 sessions (IQR 20-24); 167 of 199 (83.9%) attended at least 20. Control children attended a median of 21 sessions (IQR 19-23), and total contact time differed by less than

4%. Independent observers rated 144 of 576 sessions in each condition. Mean core-function fidelity was 89% for CREATE and 88% for control, with 21 of 24 intervention and 20 of 24 control classrooms meeting the prespecified 85% threshold. Double-rated fidelity records showed ICC=.86. Provider-condition drift was identified in 1.7% of audited sessions and corrected through supervision. Three lower-fidelity CREATE classrooms experienced staff absence, room changes, and delayed communication supports.

A total of 173 CREATE adaptations were documented: 58 sensory, 42 communication, 31 motor, 28 pacing or attention, and 14 cultural or family-context adaptations. Common changes included quieter instruments, visual schedules, AAC response boards, adapted grips, seated movement, smaller groups, and alternatives to public sharing. Acceptability averaged 4.3/5 among children able to rate directly, 4.1/5 among parents, and 4.0/5 among educators. Lower ratings clustered around unpredictable noise, transition demands, fear of performance, and fatigue-sensitive scheduling.

Fourteen CREATE children (7.0%) and ten control children (5.1%) experienced a session-related burden or adverse response. CREATE events comprised transient sensory overload, fatigue, frustration with materials, peer conflict, and performance anxiety. All were mild or moderate and resolved through stopping, environmental adaptation, or support. No serious adverse event was attributed to either program.

Table 4. Implementation, accessibility, fidelity, acceptability, cost, and safety.

Indicator	CREATE	Active control
Classrooms initiating program	24/24	24/24
Median sessions attended	22 (IQR 20–24)	21 (IQR 19–23)
Children attending ≥20 sessions	167/199 (83.9%)	154/197 (78.2%)
Core-function fidelity	89%	88%
Classrooms meeting ≥85% fidelity	21/24	20/24
Observed genuine choice opportunity	92%	89%
Documented accessibility adaptations	173	126
Child acceptability	4.3/5 (n=165)	4.0/5 (n=160)
Parent acceptability	4.1/5 (n=181)	3.9/5 (n=178)
Teacher acceptability	4.0/5 (n=73)	3.8/5 (n=70)
Program-related burden/adverse response	14/199 (7.0%)	10/197 (5.1%)
Incremental provider cost	£436 per assigned child	Not applicable
Sessions independently observed	144/576 (25.0%)	144/576 (25.0%)
Fidelity inter-rater ICC	.86	.84
Crossed providers / classrooms per provider	24 / 1 CREATE + 1 control	24 / 1 CREATE + 1 control
Provider-condition drift in audited sessions	1.7%	1.7%

4. CONFIRMATORY CO-PRIMARY OUTCOMES

At week 12, the adjusted mean standard School Function Assessment Participation score was 57.4 in CREATE and 52.8 in active control. The adjusted difference was +4.6 points (95% CI 2.1-7.1; g=.32; p=.001), close to the prespecified five-point pragmatic group benchmark. The week-6 contrast was +2.2 (95% CI 0.4-4.0), the three-month contrast +3.7 (95% CI 1.1-6.3), the six-month contrast +2.7 (95% CI 0.2-5.2), and the 12-month contrast +1.7 (95% CI -0.8-4.2). Parent-rated PEM-CY school participation showed the same direction, with a smaller standardized effect at week 12. The 12-month estimate did not show durable improvement in participation.

For the standard PedsQL Emotional Functioning subscale, the week-12 adjusted mean was 68.0 in CREATE and 64.6 in control, producing a +3.4-point difference (95% CI 1.0-5.8; g=.24; p=.007). The week-6

difference was +1.2 (95% CI -0.7 to 3.1), the three-month difference +2.5 (95% CI 0.1 to 4.9), the six-month difference +1.7 (95% CI -0.7 to 4.1), and the 12-month difference +0.7 (95% CI -1.8 to 3.2). Accessible child-report sensitivity estimates were directionally consistent but less precise. Because no population-specific minimum clinically important difference was assumed, the small week-12 effect was interpreted through its confidence interval and standardized magnitude; later estimates were uncertain.

Complete-case, classroom-fixed-effect, school-random-effect, curriculum-arts-adjusted, informant-specific, retention-weighted, and alternative crossed-provider models were similar in direction. Provider variance was small but non-zero (ICC=.016 for participation and .013 for emotional functioning), and including it changed week-12 contrasts by less than 0.2 points. Excluding three control classrooms with substantial non-study arts exposure changed estimates by less than 0.4 points. All primary models converged, school and classroom variance remained estimable, and no single classroom changed either contrast by more than 14%.

Table 5. Multilevel estimates for the co-primary outcomes.

Outcome / occasion	Adjusted means I / C	Difference (95% CI)	SE	KR df	Classroom ICC	Hedges g	Interpretation
Standard SFA Participation: week 6	52.2 / 50.1	+2.2 (0.4, 4.0)	0.92	44.1	.052	.15	Below threshold
Standard SFA Participation: week 12	57.4 / 52.8	+4.6 (2.1, 7.1)	1.27	43.6	.052	.32	Near 5-point threshold
Standard SFA Participation: 3 months	56.0 / 52.3	+3.7 (1.1, 6.3)	1.32	43.1	.052	.25	Modest retained effect
Standard SFA Participation: 6 months	54.2 / 51.5	+2.7 (0.2, 5.2)	1.28	42.7	.052	.18	Small retained effect
PedsQL Emotional Functioning: week 6	63.1 / 61.9	+1.2 (−0.7, 3.1)	0.97	44.6	.039	.09	Uncertain
PedsQL Emotional Functioning: week 12	68.0 / 64.6	+3.4 (1.0, 5.8)	1.22	44.0	.039	.24	Small; no MCID assumed
PedsQL Emotional Functioning: 3 months	66.2 / 63.7	+2.5 (0.1, 4.9)	1.22	43.5	.039	.18	Small
PedsQL Emotional Functioning: 6 months	64.8 / 63.1	+1.7 (−0.7, 4.1)	1.22	43.0	.039	.12	Uncertain
Standard SFA Participation: 12 months	52.9 / 51.2	+1.7 (−0.8, 4.2)	1.28	42.3	.052	.12	Uncertain
PedsQL Emotional Functioning: 12 months	63.5 / 62.8	+0.7 (−1.8, 3.2)	1.28	42.1	.039	.05	Uncertain

Positive differences favor CREATE. Week 12 was the confirmatory endpoint; each co-primary test used alpha=.025. No MCID was assumed for PedsQL Emotional Functioning. KR: Kenward-Roger.

https://doi.org/10.48161/ajst.v3n4a99

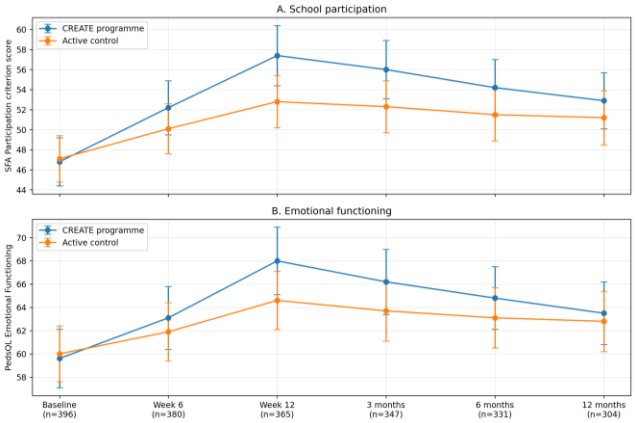


FIGURE 4. Model-based trajectories for standard SFA Participation and PedsQL Emotional Functioning. Error bars represent 95% confidence intervals; observed sample sizes are shown beneath each occasion.

5. SECONDARY EDUCATIONAL, DEVELOPMENTAL, HEALTH, AND FAMILY OUTCOMES

At week 12, after false-discovery-rate adjustment, the clearest secondary child differences were lower BRIEF-2 emotion-regulation difficulty, stronger PROMIS peer relationships, and higher educational Goal Attainment scores. Communication showed a small adjusted difference that did not remain reliable after multiplicity control; adaptive behavior was uncertain. Parent-rated PEM-CY school participation and PedsQL total quality of life favored CREATE modestly, whereas sleep and attendance were close to no effect.

At six and 12 months, secondary child estimates were generally attenuated. Goal attainment remained directionally higher at six months (+3.1 T points, 95% CI 0.7-5.5) but was uncertain at 12 months (+1.6, 95% CI -0.9-4.1). Emotion regulation and peer relationships retained small six-month differences that were not reliable at 12 months. Communication, adaptive behavior, parent-rated participation, sleep, and attendance were uncertain. The pattern did not support generalized or durable educational or health recovery.

Caregiver stress decreased by 2.4 points at week 12, but the adjusted confidence interval reached no difference. Health and educational service contacts were similar between groups through 12 months (rate ratio .98, 95% CI .77-1.25), and medication changes and initiation of non-study therapy were balanced. We report detailed family, service-use, and resource results in the Supplementary Material.

Table 6. Week-12 secondary child outcomes.

Outcome	Adjusted effect (95% CI)	FDR q	Std. effect	Interpretation
BRIEF-2 emotion regulation T	−2.8 (−4.4, −1.2)	.012	.30	Reliable small effect
PROMIS peer relationships T-score	+2.3 (0.8, 3.8)	.016	.27	Reliable small effect
Goal Attainment T-score	+4.9 (2.6, 7.2)	.004	.36	Most distinct educational effect
CCC-2 General Communication Composite	+2.1 (0.3, 3.9)	.064	.18	Small; not FDR reliable
Vineland-3 Adaptive Behavior Composite	+1.7 (−0.1, 3.5)	.119	.14	Uncertain
PedsQL total	+2.6 (0.5, 4.7)	.048	.20	Small
PROMIS sleep disturbance T-score	−1.1 (−2.6, 0.4)	.214	.10	Uncertain

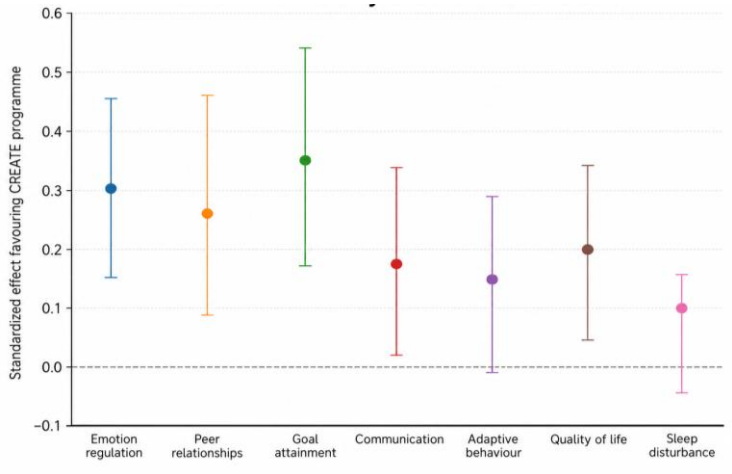


FIGURE 5. Standardized week-12 secondary child outcome effects. Positive values indicate effects favoring CREATE after orienting lower-is-better measures. We use a single neutral color because color does not encode outcome categories.

6. EXPLORATORY PROCESS AND MODERATOR FINDINGS

CREATE assignment was associated with higher early creative engagement, autonomy, and competence. Week-6 changes in emotion regulation and peer connectedness were correlated with week-12 participation and emotional outcomes. Indirect estimates were small, correlated, and sensitive to crossed-provider and classroom-fidelity adjustment; direct program estimates remained. The 48-classroom design did not support stable pathway partitioning, and we did not use a proportion-mediated estimate as a substantive conclusion. Complete exploratory estimates are reported in the Supplementary Material.

We detected no reliable diagnosis-specific effect modification after multiplicity correction. Sensory sensitivity, AAC use, family resource strain, prior arts experience, age, gender, provider, and fidelity interactions were small or imprecise. These estimates were not used to define treatment-responsive subgroups. The pattern supported a cautious examination of accessibility and implementation conditions rather than matching treatment to diagnosis.

7. QUALITATIVE FINDINGS AND NEGATIVE CASES

Six themes explained variation in response. First, communication before or alongside words captured how images, rhythm, gesture, and characters allowed children to initiate meaning without immediate verbal demand. One child used short spoken phrases to explain, “I could show it first, then the words came” (C17, inclusive school, CREATE). Teachers described the creative product as a shared reference that made later classroom communication easier, particularly when adults resisted interpreting it for the child.

Second, selectable challenge supported competence. Children valued tasks with visible progress and multiple acceptable endings. “There was not one right picture, so I could finish it my way” (C31, specialist school, CREATE). Competence diminished when adults corrected aesthetic choices or moved too quickly from exploration to performance. Negative cases led analysts to distinguish mastery from praise: praise sometimes felt like pressure, whereas control over difficulty and the ability to stop supported continued engagement.

Third, rhythm and rehearsal created usable regulation routines. Repeated musical patterns, movement sequences, and story cues helped some children anticipate transitions and re-enter tasks after frustration. A teacher stated, “The beat became a bridge back to the lesson, but only because we used the same cue outside the group” (T09, inclusive school). Children with sound sensitivity benefited when teachers could adjust volume, instrument choice, and distance; two children withdrew from early rhythm activities until quieter alternatives were introduced.

Fourth, peer recognition helped when contributions did not require public performance. Cooperative murals, shared sound patterns, and jointly developed stories created roles of different intensity. "I did the background, and he did the talking. It was still ours" (C08, specialist school). Group visibility became burdensome for children who feared evaluation or were placed in roles by adults. The option to observe, work beside the group, or contribute privately was central to accessibility.

Fifth, transfer depended on educator uptake and environmental continuity. Teachers who incorporated visual schedules, rhythmic transition cues, and child-created characters into ordinary lessons reported clearer participation gains. Where sessions were treated as isolated therapy, children described them as enjoyable but disconnected from classroom demands. Parent accounts showed a similar pattern: transferable routines persisted when they were simple, child-chosen, and compatible with home resources.

Sixth, sensory load, fatigue, and the expectation to create defined the principal boundary conditions. One parent noted, "He loved the story but came home exhausted when the room was crowded" (P22, parent). Therapists described shortening sessions, separating materials, lowering visual complexity, and accepting non-production as valid participation. Negative cases changed the final framework by positioning accessibility and refusal as active therapeutic processes rather than delivery exceptions.

Table 7. Qualitative themes, convergence, and negative cases.

Theme	Illustrative evidence	Convergent evidence	Boundary / divergence
Communication before words	"I could show it first, then the words came" (C17).	Small communication effect; teacher observations	Benefit weakened when adults interpreted for the child
Selectable challenge and competence	"There was not one right picture" (C31).	Autonomy/competence process scores; attendance	Correction and excessive praise created pressure
Rhythm and rehearsal for regulation	"The beat became a bridge back to the lesson" (T09).	BRIEF-2 improvement; transfer logs	Noise sensitivity required alternatives
Peer recognition without forced performance	"I did the background, and he did the talking" (C08).	Peer relationships; cooperative observation	Public performance burdened some children
Transfer through educator uptake	Teacher reuse of cues and creative products	Fidelity–participation pattern	Isolated sessions produced limited classroom transfer
Sensory, fatigue, and refusal boundaries	"He loved the story but came home exhausted" (P22).	Adaptation and adverse-response logs	Crowding, pace, and expectation reduced access

8. MIXED-METHODS INTEGRATION

Integration produced four principal interpretations. First, the participation effect was more educationally specific and more persistent than the emotional-functioning effect, but neither outcome showed a reliable 12-month contrast. Second, emotion regulation and peer connection offered plausible explanatory associations but did not establish mechanisms. Third, we detected no reliable diagnosis-specific or provider-specific effect modification; variation was more clearly associated with sensory fit, communication access, child choice, delivery quality, and educator transfer. Fourth, strong average acceptability did not imply universal benefit. Negative cases and mild burdens identified practical limits that mean scores alone could not represent.

Table 8. Mixed-methods joint display by research question.

Question	Quantitative finding	Qualitative explanation	Divergence / boundary	Integrated interpretation
RQ1: co-primary effectiveness	Participation +4.6; emotional functioning +3.4 at week 12	Children described agency, regulation, and recognition	Both 12-month contrasts uncertain	Selective short-term program-level advantage
RQ2: secondary outcomes	Emotion regulation, peer connection, and GAS clearest	Rhythm, rehearsal, shared roles, and attainable mastery	Sleep, attendance, and adaptive behavior uncertain	Effects were selective, not generalized recovery
RQ3: family/services	Caregiver stress small; service use unchanged	Transfer was easier with simple home routines	Resource strain reduced maintenance	No short-term service-saving claim
RQ4: processes/moderators	Small unstable associations; no reliable subgroup or provider effects	Access depended on choice, sensory fit, communication support	Forced performance and high sensory load constrained benefit	Contextual signals did not establish treatment matching
RQ5: implementation	89% fidelity; 84% high exposure; £436/child	Stakeholders valued flexibility and educator coordination	Three classrooms had staffing-related fidelity difficulty	Feasible with trained staff, adaptation resources, and protected scheduling
RQ6: explanation	Quantitative heterogeneity across classrooms	Six themes and negative cases explained variation	Enjoyment sometimes occurred without classroom transfer	Transfer required continuity beyond the therapy session

VII. DISCUSSION

1. PRINCIPAL FINDINGS

CREATE was associated with a small-to-moderate improvement in standard masked classroom participation and a smaller improvement in standard parent-rated emotional functioning at the week-12 primary endpoint. Participation approached the prespecified group-level benchmark and remained modestly higher at six months, but the 12-month contrast was uncertain. Emotional functioning showed a small week-12 difference and uncertain estimates after six months. The results indicate a selective, short-term, program-level educational effect rather than broad or durable health recovery.

The active comparator strengthens interpretation because both groups received adult attention, predictable routines, cooperative tasks, choice, reflection, and self-management practice. The contrast therefore estimates the added value of CREATE's coordinated creative, sensory, symbolic, embodied, and transfer architecture more conservatively than a passive-control design. The effect sizes were smaller than many uncontrolled reports and resembled the modest domain-specific pattern in rigorous music studies [4]-[6]. Comparable restraint is supported by controlled movement and theatre research [7], [9], [13].

The distinction between direct educational participation and emotional functioning was theoretically and methodologically important. We used the School Function Assessment criterion score without study-specific transformation and had masked observers complete it. We scored the PedsQL Emotional Functioning subscale using its standard algorithm, and it did not include school-functioning items. This design reduced shared-informant and construct-overlap concerns, although parent expectancy could still influence the emotional outcome. Children and educators described creative cues, shared products, and flexible roles as

ways to enter routines, communicate choices, and contribute with peers; these educational changes can occur without generalized emotional improvement.

Secondary results were selective. Emotion regulation, peer relationships, and individualized educational goals showed the clearest week-12 differences. Communication and quality of life changed modestly, whereas adaptive behavior, sleep, attendance, caregiver stress, and service use were uncertain. Therefore, statistical significance did not imply comprehensive educational or clinical significance. The Goal Attainment result was educationally relevant, while the absence of attendance or service-use effects limited claims about system-level health benefit or efficiency.

Exploratory process results were compatible with the Creative Participation-Development Framework but did not verify causal pathways. A separate pretrial sample supported preliminary reliability and structure of the process scales, yet external replication and criterion validation remained unavailable. In the trial, the mediators were not randomized, indirect estimates were unstable, and provider and classroom fidelity affected results. The framework therefore offers a provisional relational interpretation rather than a confirmed mechanism model.

The study detected no reliable diagnosis-specific, provider-specific, or communication-mode effect modification. This absence does not establish equal effectiveness across diagnoses, providers, or AAC use because the study had limited higher-level power. Practical variation was more consistently visible in sensory fit, communication access, child choice, classroom stability, and educator transfer. Subgroup and provider estimates were therefore retained as uncertainty descriptions rather than treatment-matching evidence.

Implementation findings distinguish feasibility from effortless scalability. Exposure and fidelity were high, but delivery required qualified providers, cross-modal scope boundaries, supervision, adapted materials, classroom coordination, and protected scheduling. Crossing providers across conditions reduced provider-treatment confounding, although each provider contributed only one classroom per condition and provider variance remained imprecise.

Safety and accessibility findings counter an exclusively affirmative account of arts participation. Sensory overload, fatigue, frustration, peer conflict, and performance anxiety were uncommon and non-serious but educationally meaningful. Child assent, permission to observe or stop, quieter and less visually complex options, AAC, adapted tools, and alternatives to public sharing were active ingredients of acceptable delivery. A technically faithful session could still be poorly matched to a child.

Mixed-methods integration added explanatory value by connecting effect estimates with transfer and negative cases. Quantitative models established magnitude and uncertainty; qualitative evidence showed that gains were strongest when creative cues crossed into ordinary classroom routines. Enjoyment without educator uptake often remained isolated from participation. This convergence supports a transactional interpretation in which program content, provider practice, classroom context, and child agency jointly shape response.

The theoretical contribution is a participation-centered account of creative-arts therapy. CREATE was not conceptualized as treatment for a diagnostic category. We provisionally linked accessible creative engagement to autonomy, competence, regulation, and peer contribution within educational environments. The program's four modalities remained inseparable in the trial, so the theoretical interpretation concerns their coordinated service architecture rather than a superior art form.

2. STRENGTHS AND LIMITATIONS

Strengths included recruitment and baseline assessment before classroom allocation, an attention- and time-matched active control, crossed providers, standard co-primary measures, masked direct participation assessment, 12-month observation, prespecified multiplicity control, accessibility and adverse-response monitoring, and integration of child, family, educator, provider, observational, and record data.

Several limitations constrain inference. Children, educators, parents, and providers could not be blinded, so expectancy, performance, and reporting effects remain possible. Equivalent early credibility ratings and masked observation reduced but did not remove bias, and assessors inferred allocation in 54% of audited

visits. Parent proxy remained the primary emotional-functioning informant; accessible child report was incomplete and less precise.

CREATE deliberately bundled visual art, music, movement, and story-based drama. The results apply to the coordinated program and cannot identify the contribution of any modality, sequence, transfer routine, or interaction among components. Cross-modal training was restricted to manualized low-risk activities, but professional scope and disciplinary identity remain relevant implementation concerns. Component attribution requires a different design and is not supported by these findings.

The sample included several neurodevelopmental profiles, strengthening pragmatic relevance but limiting condition-specific conclusions. Forty-eight classrooms and 24 crossed providers were adequate for the primary program contrasts but insufficient for definitive provider effects, mediation, moderation, or diagnosis-specific analyses. Crossing providers improved separation of program and provider variance, yet each provider delivered only one classroom per condition and precision remained limited.

The 12-month follow-up extended beyond the immediate and six-month assessments, but the observed 12-month contrasts were uncertain, and attrition reached 23.2%. School transfers and instability in family contact contributed to missingness. Multiple imputation and retention weighting supported robustness, but missing-not-at-random mechanisms cannot be excluded. The study-specific process measures had separate pretrial evaluation, yet independent external replication across communication modes, diagnoses, languages, and settings remained absent.

The findings apply to children aged 7–12 years with specified neurodevelopmental disabilities and documented participation or self-regulation difficulties in inclusive and specialist schools; they do not support generalization to all childhood disabilities or educational settings.

3. IMPLICATIONS FOR PRACTICE AND EVIDENCE DEVELOPMENT

Schools and child-development services may consider CREATE as an adjunct to strengthen classroom participation, emotion regulation, peer contribution, and individualized educational goals. Implementation requires qualified providers, discipline-appropriate supervision, communication and sensory resources, continuing child assent, non-performance options, educator involvement in transfer, and routine monitoring of burden. Adoption decisions also require local appraisal of provider cost and workforce capacity.

The principal evidence requirements are independent validation of process measures, larger numbers of schools and cross-provider comparisons, designs that can separate program components, and formal economic evaluation.

VIII. CONCLUSION

Among children aged 7–12 years with neurodevelopmental disabilities and documented participation or self-regulation difficulties, assignment to the 12-week CREATE program was associated with modestly higher standard masked school participation and a smaller short-term improvement in standard parent-rated emotional functioning than an attention- and time-matched non-art developmental program. Participation remained modestly higher at six months, but neither co-primary contrast was reliable at 12 months. The clearest secondary differences concerned emotion regulation, peer relationships, and individualized educational goals; adaptive behavior, sleep, attendance, caregiver stress, and service use did not differ reliably. Exploratory process, provider, and subgroup analyses did not establish causal mechanisms or treatment-responsive subgroups. CREATE is therefore best interpreted as an accessible, context-dependent program-level adjunct whose observed value concerned short-term participation and selected psychosocial domains, not an individual art modality, a diagnosis-specific treatment, or evidence of durable health recovery.

Author Contributions

The author conducted the conceptualization, methodology, data analysis, investigation, writing, review, editing, and final approval of the manuscript.

Funding

This research received no external funding.

Data Availability

The dataset will be available from the author upon reasonable request.

Conflicts of Interest

The author declares no conflict of interest.

REFERENCES

[1]. B. M. Edwards, E. Smart, G. King, C. J. Curran, and S. Kingsnorth, "Performance and visual arts-based programs for children with disabilities: A scoping review focusing on psychosocial outcomes," Disability and Rehabilitation, vol. 42, no. 4, pp. 574–585, 2020, doi: 10.1080/09638288.2018.1503734.

[2]. H. Mayer-Benarous, X. Benarous, F. Vonthron, and D. Cohen, "Music therapy for children with autism spectrum disorder and/or other neurodevelopmental disorders: A systematic review," Frontiers in Psychiatry, vol. 12, art. 643234, 2021, doi: 10.3389/fpsyt.2021.643234.

[3]. X. Ke, W. Song, M. Yang, J. Li, and W. Liu, "Effectiveness of music therapy in children with autism spectrum disorder: A systematic review and meta-analysis," Frontiers in Psychiatry, vol. 13, art. 905113, 2022, doi: 10.3389/fpsyt.2022.905113.

[4]. Ł. Bieleninik et al., "Effects of improvisational music therapy vs enhanced standard care on symptom severity among children with autism spectrum disorder: The TIME-A randomized clinical trial," JAMA, vol. 318, no. 6, pp. 525–535, 2017, doi: 10.1001/jama.2017.9478.

[5]. M. Sharda et al., "Music improves social communication and auditory–motor connectivity in children with autism," Translational Psychiatry, vol. 8, art. 231, 2018, doi: 10.1038/s41398-018-0287-3.

[6]. T. Rabeyron et al., "A randomized controlled trial of 25 sessions comparing music therapy and music listening for children with autism spectrum disorder," Psychiatry Research, vol. 293, art. 113377, 2020, doi: 10.1016/j.psychres.2020.113377.

[7]. S. Aithal, V. Karkou, S. Makris, T. Karaminis, and J. Powell, "A dance movement psychotherapy intervention for the wellbeing of children with an autism spectrum disorder: A pilot intervention study," Frontiers in Psychology, vol. 12, art. 588418, 2021, doi: 10.3389/fpsyg.2021.588418.

[8]. S. Aithal, Z. Moula, V. Karkou, T. Karaminis, J. Powell, and S. Makris, "A systematic review of the contribution of dance movement psychotherapy towards the well-being of children with autism spectrum disorders," Frontiers in Psychology, vol. 12, art. 719673, 2021, doi: 10.3389/fpsyg.2021.719673.

[9]. Z. Moula, J. Powell, S. Brocklehurst, and V. Karkou, "Feasibility, acceptability, and effectiveness of school-based dance movement psychotherapy for children with emotional and behavioral difficulties," Frontiers in Psychology, vol. 13, art. 883334, 2022, doi: 10.3389/fpsyg.2022.883334.

[10]. H. Ren, G. Ren, Y. Zhan, and Z. Jia, "Examining the efficacy of dance movement and music mixed treatment on social communication impairment in children with autism—Based on family parent-child situation," Frontiers in Psychology, vol. 13, art. 937564, 2022, doi: 10.3389/fpsyg.2022.937564.

[11]. V. Martínez-Vérez, P. Gil-Ruíz, and S. Domínguez-Lloria, "Interventions through art therapy and music therapy in autism spectrum disorder, ADHD, language disorders, and learning disabilities in pediatric-aged children: A systematic review," Children, vol. 11, no. 6, art. 706, 2024, doi: 10.3390/children11060706.

[12]. A. Bernier, K. Ratcliff, C. Hilton, P. Fingerhut, and C.-Y. Li, "Art interventions for children with autism spectrum disorder: A scoping review," American Journal of Occupational Therapy, vol. 76, no. 5, art. 7605205030, 2022, doi: 10.5014/ajot.2022.049320.

[13]. B. A. Corbett et al., "Improvement in social competence using a randomized trial of a theatre intervention for children with autism spectrum disorder," Journal of Autism and Developmental Disorders, vol. 46, no. 2, pp. 658–672, 2016, doi: 10.1007/s10803-015-2600-9.

[14]. B. A. Corbett, S. D. Blain, S. Ioannou, and M. Balser, "Changes in anxiety following a randomized controlled trial of a theatre-based intervention for youth with autism spectrum disorder," Autism, vol. 21, no. 3, pp. 333–343, 2017, doi: 10.1177/1362361316643623.

[15]. C. Imms, B. Adair, D. Keen, A. Ullenhag, P. Rosenbaum, and M. Granlund, "Participation, both a means and an end: A conceptual analysis of processes and outcomes in childhood disability," Developmental Medicine & Child Neurology, vol. 59, no. 1, pp. 16–25, 2017, doi: 10.1111/dmcn.13237.

[16]. E. L. Deci and R. M. Ryan, "The 'what' and 'why' of goal pursuits: Human needs and the self-determination of behavior," Psychological Inquiry, vol. 11, no. 4, pp. 227–268, 2000, doi: 10.1207/S15327965PLI1104_01.

[17]. A. Sameroff, "A unified theory of development: A dialectic integration of nature and nurture," Child Development, vol. 81, no. 1, pp. 6–22, 2010, doi: 10.1111/j.1467-8624.2009.01378.x.

[18]. S. J. Landes, S. A. McBain, and G. M. Curran, "An introduction to effectiveness-implementation hybrid designs," Psychiatry Research, vol. 280, art. 112513, 2019, doi: 10.1016/j.psychres.2019.112513.

[19]. K. Skivington et al., "A new framework for developing and evaluating complex interventions: Update of Medical Research Council guidance," BMJ, vol. 374, art. n2061, 2021, doi: 10.1136/bmj.n2061.

[20]. M. K. Campbell et al., "CONSORT 2010 statement: Extension to cluster randomized trials," BMJ, vol. 345, art. e5661, 2012, doi: 10.1136/bmj.e5661.

[21]. T. C. Hoffmann et al., "Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide," BMJ, vol. 348, art. g1687, 2014, doi: 10.1136/bmj.g1687.

[22]. L. J. Damschroder, C. M. Reardon, M. A. O. Widerquist, and J. Lowery, "The updated Consolidated Framework for Implementation Research based on user feedback," Implementation Science, vol. 17, art. 75, 2022, doi: 10.1186/s13012-022-01245-0.

[23]. R. E. Glasgow et al., "RE-AIM planning and evaluation framework: Adapting to new science and practice with a 20-year review," Frontiers in Public Health, vol. 7, art. 64, 2019, doi: 10.3389/fpubh.2019.00064.

[24]. K. Hemming, S. Eldridge, G. Forbes, C. Weijer, and M. Taljaard, "How to design efficient cluster randomized trials," BMJ, vol. 358, art. j3064, 2017, doi: 10.1136/bmj.j3064.

[25]. J.-L. Hwang, P. L. Davies, M. P. Taylor, and W. J. Gavin, "Validation of School Function Assessment with elementary school children," OTJR: Occupation, Participation and Health, vol. 22, no. 2, pp. 48–58, 2002, doi: 10.1177/153944920202200202.

[26]. M. A. Viecili and J. A. Weiss, "Reliability and validity of the Pediatric Quality of Life Inventory with individuals with intellectual and developmental disabilities," American Journal on Intellectual and Developmental Disabilities, vol. 120, no. 4, pp. 289–301, 2015, doi: 10.1352/1944-7558-120.4.289.

[27]. K. Harpster, A. Sheehan, E. A. Foster, E. Leffler, S. M. Schwab, and J. M. Angeli, "The methodological application of goal attainment scaling in pediatric rehabilitation research: A systematic review," Disability and Rehabilitation, vol. 41, no. 24, pp. 2855–2864, 2019, doi: 10.1080/09638288.2018.1474952.

[28]. C. K. Enders, S. A. Mistler, and B. T. Keller, "Multilevel multiple imputation: A review and evaluation of joint modeling and chained equations imputation," Psychological Methods, vol. 21, no. 2, pp. 222–240, 2016, doi: 10.1037/met0000063.

[29]. K. DiazOrdaz, M. G. Kenward, M. Gomes, and R. Grieve, "Multiple imputation methods for bivariate outcomes in cluster randomized trials," Statistics in Medicine, vol. 35, no. 20, pp. 3482–3496, 2016, doi: 10.1002/sim.6935.

[30]. T. J. VanderWeele, "Mediation analysis: A practitioner's guide," Annual Review of Public Health, vol. 37, pp. 17–32, 2016, doi: 10.1146/annurev-publhealth-032315-021402.

[31]. C. Leyrat, K. E. Morgan, B. Leurent, and B. C. Kahan, "Cluster randomized trials with a small number of clusters: Which analyses should be used?" International Journal of Epidemiology, vol. 47, no. 1, pp. 321–331, 2018, doi: 10.1093/ije/dyx169.

[32]. K. Malterud, V. D. Siersma, and A. D. Guassora, "Sample size in qualitative interview studies: Guided by information power," Qualitative Health Research, vol. 26, no. 13, pp. 1753–1760, 2016, doi: 10.1177/1049732315617444.

[33]. N. K. Gale, G. Heath, E. Cameron, S. Rashid, and S. Redwood, "Using the framework method for the analysis of qualitative data in multi-disciplinary health research," BMC Medical Research Methodology, vol. 13, art. 117, 2013, doi: 10.1186/1471-2288-13-117.

[34]. M. T. McCrudden, G. Marchand, and P. A. Schutz, "Joint displays for mixed methods research in psychology," Methods in Psychology, vol. 5, art. 100067, 2021, doi: 10.1016/j.metip.2021.100067.

[35]. T. C. Guetterman, S. Fàbregues, and R. Sakakibara, "Visuals in joint displays to represent integration in mixed methods research: A methodological review," Methods in Psychology, vol. 5, art. 100080, 2021, doi: 10.1016/j.metip.2021.100080.

[36]. H. M. Levitt et al., "Journal article reporting standards for qualitative primary, qualitative meta-analytic, and mixed methods research in psychology," American Psychologist, vol. 73, no. 1, pp. 26–46, 2018, doi: 10.1037/amp0000151.

[37]. B. C. O'Brien, I. B. Harris, T. J. Beckman, D. A. Reed, and D. A. Cook, "Standards for reporting qualitative research: A synthesis of recommendations," Academic Medicine, vol. 89, no. 9, pp. 1245–1251, 2014, doi: 10.1097/ACM.0000000000000388.