ARCHAIA JOURNAL OF SCIENCE AND TECHNOLOGY
VOL. 3, NO. 3, Sep 2026
https://doi.org/10.48161/ajst.v3n3a98

Arts-Based Interventions for Reducing Stress, Anxiety, and Depression among University Students

Mohamed Merza Khalil Ebrahim Al. Rayes[1*]

[1] Department of Physiotherapy, College of Health Sciences, University of Bahrain, Ahlia University, Manama, Kingdom of Bahrain;

* **Corresponding author:** phd.mmkrayes@gmail.com.

ABSTRACT: University students frequently experience overlapping stress, anxiety, and depressive symptoms, yet many do not use conventional counselling because of stigma, limited capacity, or preference for non-clinical support. This study evaluated the comparative effectiveness, mechanisms, and implementation of structured visual-art, music, and combined movement/drama programmes. A multicentre, assessor-blinded, four-arm randomized controlled trial with an embedded longitudinal explanatory mixed-methods process evaluation allocated 448 students equally to three manualized arts programmes or an attention- and time-matched active wellbeing control. Researchers delivered ten weekly 75-minute sessions across four universities. The primary endpoint was the raw DASS-21 total score (0–63) at week 10. At week 10, adjusted differences versus control were −4.5 points for visual art, −5.3 for music, and −4.9 for combined movement/drama. Standardized effects, calculated by dividing each adjusted contrast by the pooled baseline DASS-21 standard deviation of 8.70, were 0.52, 0.61, and 0.56. Benefits remained at six months with moderate attenuation. Using a study-specific five-point response threshold, response probabilities were 68%–74% in the arts conditions and 46% in control. Mean attendance was 8.3–8.8 sessions, programme completion was 80%–89%, and mean CSQ-8 satisfaction scores were 24.9–27.3. Temporally ordered analyses identified partial statistical pathways through emotion regulation, coping self-efficacy, creative self-efficacy, and social connectedness, although arm-specific mediation remained exploratory. Qualitative findings connected improvement with symbolic distance, rhythmic regulation, integrated embodied and role-based rehearsal, attainable mastery, and reciprocal belonging. The three arts conditions produced similar average effects, and direct comparisons were inconclusive. Structured arts-based programs may be acceptable low-intensity options within university mental-health services, but not substitutes for indicated clinical care.

Keywords: arts-based intervention; university students; psychological distress; emotion regulation; mixed methods.

I. INTRODUCTION

University life combines developmental transition, academic competition, financial pressure, changing social networks, and uncertainty about future employment. Cross-national evidence indicates that common mental disorders frequently begin before or during university and can interfere with persistence, attainment, relationships, and help-seeking [1]. Surveys using symptom measures similarly show substantial depression, anxiety, and stress associated with workload, finances, sleep, social support, and academic confidence [2]. A systematic review and meta-analysis reported substantial global prevalence of depression and anxiety symptoms among college students, with marked variation by region, measurement, and context [3]. These findings position student mental health as both a health concern and an educational equity issue.

Universities have expanded counselling and digital support. Prevention and low-intensity programmes are needed for students with mild-to-moderate distress who do not require specialist treatment. A systematic

review and meta-analysis of randomized trials found that psychological interventions can reduce common mental-health problems among university and college students, although follow-up evidence and intervention optimisation remain uneven [4]. Arts-based programmes may complement these approaches because they invite action, experimentation, and social participation without requiring students to define themselves as patients.

Arts engagement can activate psychological, biological, behavioural, and social pathways simultaneously [5]. Music can influence arousal through tempo, repetition, breathing, vocalisation, and interpersonal synchrony; meta-analytic evidence indicates small-to-moderate reductions in stress-related outcomes across varied populations [6]. Dance and movement interventions support affect regulation, body awareness, agency, and interpersonal functioning, although effects vary by population and format [7]. Drama-based programmes add role distance, narrative reframing, perspective-taking, and rehearsal of alternative responses, but controlled evidence does not establish the superiority of any single drama model [8].

Visual art offers a distinctive route through image-making, sensory engagement, symbolic distance, and tangible products. Brief art-making has been associated with reductions in cortisol and perceived stress [9]. Controlled art-therapy research reports encouraging effects on adult anxiety but also substantial heterogeneity in programme content, comparator strength, and methodological quality [10]. Evidence for student-specific visual-art interventions remains comparatively limited, especially for active-control trials with follow-up and academic-functioning outcomes. The present programmes were participatory arts-in-health interventions delivered by qualified facilitators; they were not licensed creative arts therapies or substitutes for psychotherapy [11].

The evidence base nevertheless has recurring limitations. Many studies use single-group pre–post designs, small samples, immediate outcomes, or passive controls. Researchers often pool different arts modalities, making it difficult to distinguish artistic content from facilitator attention, expectancy, routine, and typical group contact. Proposed mechanisms are often measured concurrently with outcomes or inferred retrospectively. Sustained effects, adverse responses, cultural adaptation, and academic functioning receive limited attention. Measurement is also uneven: researchers sometimes label creative activity as "self-expression" when the instrument assesses confidence in creative ability, and they often measure student functioning with unvalidated adaptations.

The present study addressed these limitations through a multicentre, assessor-blinded, four-arm randomized controlled trial comparing visual art, music, and a combined movement/drama programme with an active wellbeing control matched for contact time, facilitator responsiveness, social interaction, and reflection. The raw DASS-21 total score at week 10 was the confirmatory primary endpoint. Symptom subscales, wellbeing, psychosocial mechanisms, quality-of-life domains, and exploratory academic functioning were secondary. A longitudinal explanatory mixed-methods process evaluation examined mechanism evidence, engagement, fidelity, cultural fit, academic transfer, and negative cases. The contribution lies in testing a common arts-engagement architecture while preserving modality-specific interpretation and maintaining a clear boundary between arts-in-health programming and formal psychological treatment.

1. RESEARCH PROBLEM

University mental-health services require accessible interventions that can reduce overlapping stress, anxiety, and depressive symptoms without over-medicalising mild difficulties. The primary research problem was comparative effectiveness: whether manualized visual-art, music, and combined movement/drama programmes produced greater week-10 reductions in psychological distress than an active control containing attention, social contact, psychoeducation, and reflection. Mechanism, moderation, attendance, and implementation analyses were secondary because they explained heterogeneity and feasibility but were not powered to establish modality superiority or definitive causal pathways.

2. AIM AND OBJECTIVES

The aim was to estimate the comparative effects of three structured arts-based programs on raw DASS-21 total psychological distress at week 10 (the primary endpoint) and to explain secondary outcomes, mechanisms, implementation, and six-month maintenance relative to an active wellbeing control.

- To estimate the effects of visual-art, music, and combined movement/drama programmes on raw DASS-21 total psychological distress at week 10 relative to active control, with six-month maintenance as a secondary follow-up contrast.
- To compare stress, anxiety, depression, psychological wellbeing, WHOQOL-BREF domains, and exploratory academic functioning across the four conditions.
- To estimate temporally ordered direct and indirect statistical effects through emotion regulation, coping self-efficacy, creative self-efficacy, and social connectedness; modality-specific mediation, moderation, and attendance analyses were prespecified as exploratory.
- To estimate quantitative variation through participant and facilitator accounts of engagement, psychological safety, cultural relevance, academic transfer, implementation quality, barriers, and adverse responses.

3. RESEARCH QUESTIONS

- At week 10, to what extent did visual art, music, and the combined movement/drama programme reduce raw DASS-21 total psychological distress relative to active control, and how much benefit remained at follow-up?
- How did the four groups differ in stress, anxiety, depression, wellbeing, WHOQOL-BREF domains, exploratory academic functioning, and the prespecified mechanism measures?
- To what extent did week-5 changes in emotion regulation, coping self-efficacy, creative self-efficacy, and social connectedness statistically account for week-10 distress effects, and did prespecified student characteristics modify those effects after multiplicity adjustment?
- How did students and facilitators explain stronger, weaker, or adverse responses, implementation variation, and transfer into academic and everyday life?

4. RESEARCH GAP AND ORIGINAL CONTRIBUTION

Five gaps guided the study: insufficient direct comparison of distinct arts modalities against credible active controls; limited use of a unified distress endpoint with sustained follow-up; inadequate temporal testing of mechanisms; weak integration of effectiveness with fidelity, cultural adaptation, and adverse responses; and limited use of construct-equivalent measures for academic functioning and creative processes. The study's originality rests on integrating these elements within one student-centred trial. It distinguishes confirmatory effectiveness estimates from exploratory mechanism and subgroup analyses, treats combined movement/drama as one combined programme, and separates measured creative self-efficacy from qualitative accounts of creative self-expression.

II. LITERATURE REVIEW AND CONCEPTUAL FRAMEWORK

1. COMPARATIVE EVIDENCE ACROSS MODALITIES

The three arts programmes shared group membership, facilitator attention, rehearsal, feedback, and reflection, but they organised experience differently. Visual art externalised emotion into material form and allowed private revision. Music combined auditory pattern, breathing, vocalisation, motor timing, and synchrony. The combined movement/drama programme integrated grounding, gesture, spatial awareness, role exploration, improvisation, and interpersonal rehearsal. These differences justified comparative evaluation while preventing separate conclusions about movement and drama components.

The strength of existing findings depends heavily on comparator choice. Wait-list and no-treatment comparisons cannot separate arts-specific processes from expectancy, routine, belonging, and supportive facilitation. Active controls provide a more demanding estimate of added artistic content, although they can reduce apparent effect sizes. Sustained follow-up matters just as much: immediate enjoyment can coexist with limited transfer, whereas modest post-intervention effects may remain meaningful when students continue using strategies during examinations or periods of social strain.

Adverse and ambivalent responses are theoretically informative rather than peripheral. Visual tasks may evoke unwanted exposure through personal imagery; music can trigger memories or sensory overload; movement can create pain or body-consciousness; drama can generate performance anxiety. Cultural

relevance influences material choice, public visibility, gendered interaction, emotional disclosure, touch, and meanings attached to artistic skill. Fidelity therefore requires preservation of core functions alongside bounded adaptation in pace, content, sensory intensity, physical demand, and mode of participation [5], [8],[11].

2. *ACTIVE CONTROLS AND SUSTAINED OUTCOMES*

Active-control trials are central to comparative effectiveness because attention, routine, expectancy, facilitator warmth, and group belonging can independently improve distress. They also yield smaller but more interpretable arts-specific effects than passive controls. Longer follow-up is necessary because some regulation strategies become self-sustaining, whereas contextual understanding, group support, and structured practice may weaken after sessions end. Few arts trials combine active controls, post-intervention mediation timing, adverse-event monitoring, and follow-up beyond three months [4],[10].

3. *SAFETY, CULTURAL ADAPTATION, ACADEMIC FUNCTIONING, AND MEASUREMENT*

Adverse responses and cultural mismatch are underreported in arts-intervention research. Psychological activation, sensory overload, discomfort with touch or public performance, and inaccessible materials can reduce benefit even when attendance and technical fidelity are high. Researchers also often infer academic outcomes from distress reduction rather than measuring them directly. Existing studies rarely validate adaptations in student functioning or distinguish creative self-efficacy from creative self-expression. These limitations support conservative interpretation of academic transfer and explicit construct naming [4],[11].

4. *ARTS ENGAGEMENT–EMOTION REGULATION FRAMEWORK*

Emotion Regulation Theory provided the central mechanism model. Regulation was defined as the processes through which students noticed, tolerated, expressed, reappraised, and modified emotional responses [12]. Coping self-efficacy and social connectedness were complementary measured pathways. Creative self-efficacy represented confidence in producing and adapting creative responses; it was not treated as a direct measure of creative self-expression. Broaden-and-Build Theory and Self-Determination Theory informed interpretation of positive affect, autonomy, competence, and relatedness but were not modelled as separate causal systems [13], [14].

We retained Embodied Cognition and the Expressive Therapies Continuum as modality-specific interpretive lenses for sensory, kinesthetic, affective, symbolic, and action-based experience [15], [16]. Social-connectedness theory supported the role of reciprocal recognition and dependable group membership [17]. The resulting Arts Engagement–Emotion Regulation Framework therefore placed emotion regulation at the centre, treated coping self-efficacy, creative self-efficacy, and connectedness as complementary mediators, and used embodied processes to interpret the combined movement/drama condition rather than to claim a separately measured pathway.

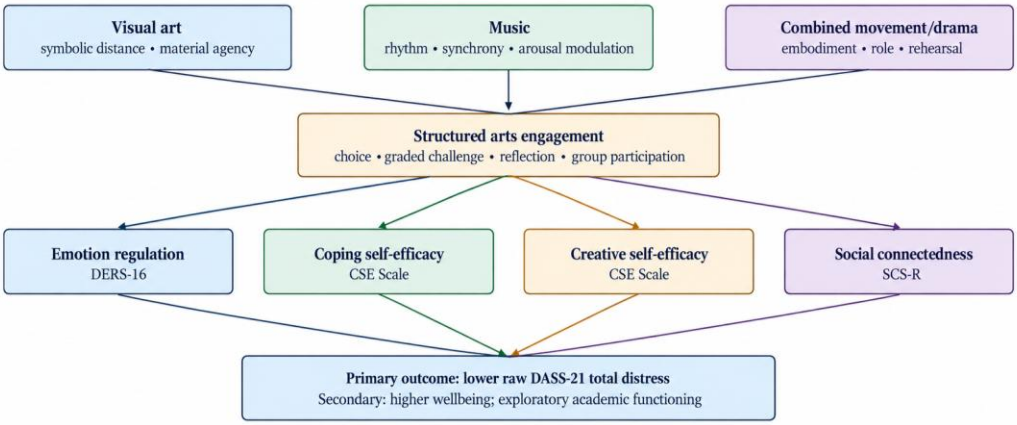


FIGURE 1. Arts Engagement–Emotion Regulation Framework. The four measured mechanisms are represented as parallel pathways; embodied and expressive processes are modality-specific explanatory lenses rather than additional statistical mediators.

https://doi.org/10.48161/ajst.v3n3a98

Table 1. Operationalization of the Arts Engagement–Emotion Regulation Framework.

Framework component	Arts process	Measured indicator	Analytical role
Emotion regulation	Expression, reappraisal, arousal modulation	DERS-16	Central mediator
Coping self-efficacy	Graded mastery and usable coping practice	Coping Self-Efficacy Scale	Complementary mediator
Creative self-efficacy	Confidence in generating and adapting creative responses	Creative Self-Efficacy Scale	Complementary mediator
Social connectedness	Synchrony, recognition, mutual contribution	SCS-R	Complementary mediator
Embodied and expressive processing	Breath, gesture, posture, rhythm, role, symbolic production	Observation and qualitative accounts	Modality-specific explanatory lens
Student functioning	Concentration, attendance, task initiation	Study-adapted WSAS	Exploratory secondary outcome

III. METHODS

1. DESIGN, SETTING, AND REPORTING

The study used a multicentre, assessor-blinded, four-arm, parallel-group randomized controlled design with an embedded longitudinal explanatory mixed-methods process evaluation. Four urban public universities with different institutional sizes and disciplinary profiles delivered the programs over one academic year. Assessments occurred at baseline (T0), week 5 (T1), week 10 (T2), three months (T3), and six months (T4). The treatment-policy estimand compared assignment to each arts condition with active control irrespective of subsequent attendance or permitted co-intervention. The trial was prospectively entered in the University Trials Register (2022-01). We finalized protocol version 2.0 (12 January 2022) and statistical analysis plan version 1.1 (30 August 2022) before unblinding; we logged deviations by date, rationale, affected arm, and analytical consequence.

Trial reporting followed CONSORT 2010 [18], protocol specification followed SPIRIT 2013 [19], and intervention reporting followed TIDieR guidance [20]. The process evaluation examined implementation, mechanisms, context, and participant response in accordance with complex-intervention guidance [21]. The raw DASS-21 total score at week 10 was the confirmatory primary endpoint. Arts-versus-control comparisons were confirmatory; six-month maintenance, arts-arm contrasts, mediation, moderation, academic functioning, and attendance analyses were secondary or exploratory.

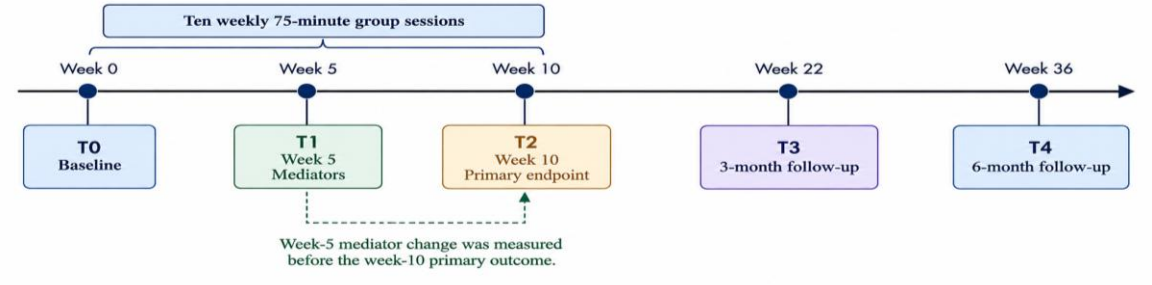


FIGURE 2. Intervention and assessment timeline from baseline through six-month follow-up. Week-5 mediator assessment preceded the week-10 primary endpoint.

2. PARTICIPANTS, ELIGIBILITY, AND RECRUITMENT

We recruited students aged 18–30 years through university health services, course announcements, student associations, accommodation networks, and digital notices. Eligibility required current enrolment, a DASS-21 total score between 15 and 42, and at least one subscale in the mild-to-moderate range. Students needed sufficient language proficiency to complete validated measures and capacity to participate safely in any randomized condition. Exclusion criteria were acute suicide risk, current psychiatric crisis, severe cognitive impairment, unstable medical conditions, a mobility or sensory need that could not be safely accommodated across all conditions, concurrent enrolment in structured creative arts therapy, or inability to provide informed consent. We used screening combined questionnaires, a brief safety interview, a treatment and medication review, and referral when higher-intensity care was indicated.

3. SAMPLE SIZE

The target sample was 448 students, providing 112 participants per arm. The calculation assumed three confirmatory arts-versus-control comparisons at week 10, a standardized difference of 0.35, 80% power, two-sided familywise alpha of 0.05 controlled by Holm adjustment, a baseline-to-post correlation of 0.55, a delivery-group intraclass correlation of 0.02, a mean group size of eight, and 18% attrition. The study was powered for modest arts-versus-control effects, not small differences among arts programmes, facilitator-level variance, four-mediator arm-specific mediation, or interaction effects. The 56 groups were nested within 20 facilitators, and facilitators were assigned exclusively to one intervention arm.

4. RANDOMIZATION, ALLOCATION CONCEALMENT, AND BLINDING

Following baseline assessment, an independent statistician generated the allocation sequence using variable blocks of four and eight, stratified by university and baseline distress severity. A central, password-protected system made allocation inaccessible to recruiters, facilitators, assessors, and local coordinators until enrolment was confirmed. Outcome assessors and the primary analyst remained masked to condition. Participants and facilitators could not be blinded, but all conditions were presented as credible group wellbeing programmes. Assessors documented accidental disclosure and recorded allocation guesses after each post-intervention assessment.

5. INTERVENTION CONDITIONS

Each programme comprised ten weekly 75-minute sessions delivered to groups of seven to nine students. We matched contact time, room quality, group size, facilitator attention, opening and closing routines, guided reflection, and home-practice invitations. The combined movement/drama arm allocated approximately 55% of core activity time to grounding, mirroring, movement imagery, posture, and spatial work and 45% to role exploration, improvisation, dialogue, and ensemble scenes. Sessions 1–4 were movement-led, sessions 5–8 were drama-led, and sessions 9–10 integrated both components. The design therefore supported conclusions about the combined programme only. Students could observe, modify, or decline any activity without withdrawing.

Table 2. Intervention content and matched delivery characteristics.

Condition	Core content	Dose and structure	Primary process emphasis
Visual art	Drawing, collage, clay, image dialogue, visual metaphor, collaborative installation	10 × 75 min; six individual and four collaborative sessions	Symbolic distance, sensory regulation, creative agency
Music	Rhythmic entrainment, group singing, accessible instruments, improvisation, shared composition	10 × 75 min; progressive rhythm, voice, improvisation, and composition	Arousal modulation, mastery, synchrony, belonging

Condition	Core content	Dose and structure	Primary process emphasis
Combined movement/drama	Grounding, mirroring, movement imagery, role exploration, improvisation, ensemble scenes	10 × 75 min; ~55% movement and 45% drama; sessions 1–4 movement-led, 5–8 drama-led, 9–10 integrated	Embodied regulation, perspective-taking, interpersonal rehearsal
Active wellbeing control	Psychoeducation, structured discussion, non-art problem solving, study-wellbeing planning	10 × 75 min; contact, reflection, and group structure matched	Attention, reflection, routine, social contact

6. FACILITATORS, FIDELITY, AND SAFETY

Fifty-six delivery groups were formed, fourteen per condition, with two or three groups delivered by each facilitator. Groups were nested within facilitators, and facilitators were nested within intervention arms. Twenty facilitators were assigned exclusively to one condition: five visual-art, five music, five combined movement/drama, and five active-control facilitators. The arts team comprised visual artists or art educators, music practitioners, and combined movement/drama practitioners; 12 held relevant postgraduate qualifications and three held accredited professional diplomas. The control team comprised three counsellors and two student-health-promotion specialists. Median prior group-facilitation experience was six years in both the arts and control teams. All facilitators completed 16 hours of standardized training and 90-minute monthly supervision covering the manual, psychological safety, graded challenge, accessibility, confidentiality, contamination prevention, expectancy neutrality, and adverse-event response.

Fidelity combined facilitator checklists, session logs, and independent observation of 20% of sessions. Observers rated core-component delivery from 0 (absent) to 2 (complete), adaptation quality from 1 to 5, and prohibited cross-condition content as present or absent. Assessors recorded attendance, late arrival, early departure, home practice, non-study arts participation, psychological treatment, medication changes, and deviations. Expectancy was measured after the first session, before substantial intervention exposure, and assessors remained masked to these ratings. Adverse events were classified by severity, seriousness, expectedness, and intervention relatedness; assessors independently reviewed serious events.

7. OUTCOMES AND MEASURES

Psychological distress was assessed with the DASS-21 [22]. Each of the 21 items was scored 0–3. Raw item scores were summed without doubling: total psychological distress ranged from 0 to 63, and each seven-item stress, anxiety, and depression subscale ranged from 0 to 21. The prespecified total was used as a unified transdiagnostic endpoint because the study targeted overlapping symptoms rather than diagnostic severity, the three subscales were strongly correlated at baseline, and longitudinal factor checks supported a stable higher-order distress construct. Doubled subscale conventions used in some clinical severity tables were not applied. We measured psychological wellbeing with the WEMWBS [23], emotion-regulation difficulty with the DERS-16 [24], and coping self-efficacy with the 26-item Coping Self-Efficacy Scale [25]. We assessed creative self-efficacy with the Creative Self-Efficacy Scale [26] and social connectedness with the SCS-R [27].

Academic functioning was assessed with a study-adapted WSAS in which “work” was replaced by “study” [28]. Expert review, cognitive interviews, and internal-consistency checks supported comprehensibility, but criterion validity was not established; all academic-functioning analyses were therefore exploratory. We measured quality of life with the WHOQOL-BREF [29] and reported scores for the physical, psychological, social-relationships, and environmental domains. No non-standard overall total was calculated. To reduce respondent burden, we administered the DASS-21 and wellbeing at all five occasions. We assessed mediators at baseline, week 5, week 10, and six months; we assessed academic functioning and WHOQOL-BREF at baseline, week 10, three months, and six months. Treatment expectancy used the

Credibility/Expectancy Questionnaire, and alliance used WAI-SR [30], [31]. Group engagement used the GCQ-S Engagement subscale, and satisfaction used CSQ-8 [32], [33].

Table 3. Measurement and process-evaluation schedule.

Measure/data source	Construct / scoring	Timing	Role
DASS-21 raw total and subscales	Total 0–63; subscales 0–21; no doubling	T0–T4	Primary total at T2; subscales secondary
WEMWBS	Psychological wellbeing	T0–T4	Secondary
DERS-16	Emotion-regulation difficulty	T0, T1, T2, T4	Central mediator
Coping Self-Efficacy Scale	Coping confidence	T0, T1, T2, T4	Mediator
Creative Self-Efficacy Scale	Confidence in creative capability	T0, T1, T2, T4	Mediator; not direct self-expression
SCS-R	Social connectedness	T0, T1, T2, T4	Mediator
Study-adapted WSAS	Academic-role impairment; higher=worse	T0, T2–T4	Exploratory secondary
WHOQOL-BREF	Physical, psychological, social, environmental domains	T0, T2–T4	Secondary; no overall total
Credibility/Expectancy Questionnaire	Credibility and expectancy	After session 1	Expectancy process measure
WAI-SR	Participant–facilitator alliance	T1 and T2	Process measure
GCQ-S Engagement subscale	Constructive group engagement	T1 and T2	Process measure
CSQ-8	Programme satisfaction, 8–32	T2	Process measure
Attendance, fidelity, co-interventions, adverse events	Implementation and safety	Continuous	Process and safety
Interviews, focus groups, journals, observations, logs	Experience, context, transfer, negative cases	T1–T4	Qualitative explanation

8. PRACTICAL IMPORTANCE AND RESPONDER DEFINITION

No universally established minimally important difference exists for the raw DASS-21 total in comparable university samples. The study prespecified a five-point decrease as a study-specific practical-response threshold. The baseline standard deviation was 8.7, so one-half standard deviation was 4.35 points; rounding upward produced the five-point threshold. Reliable change was calculated from the baseline standard deviation and total-score omega (ω=0.91): RCI threshold = $1.96 \times SD \times \sqrt{[2(1-\omega)]} = 1.96 \times 8.7 \times \sqrt{0.18} = 7.23$. We therefore defined reliable improvement and deterioration as decreases or increases of at least eight raw points. These distribution-based indicators supported interpretation within this study and were not presented as universal clinical cut-offs. We pooled responder outcomes across imputed datasets on the log-odds scale to obtain marginal probabilities, risk differences, and numbers needed to treat.

9. QUALITATIVE PROCESS EVALUATION

Maximum-variation purposive sampling selected 56 students after week 10, with fourteen from each condition and variation in university, attendance, study level, gender identity, living situation, prior arts experience, satisfaction, and outcome trajectory. Twenty-eight completed individual interviews lasting 42–70 minutes, and 28 participated in seven focus groups lasting 75–95 minutes. Twenty facilitator interviews, reflective journals, structured observations, and session logs supplied delivery and longitudinal perspectives. Sampling included improvement, limited change, deterioration, low attendance, ambivalent satisfaction, and adverse-response cases.

We transcribed audio recordings verbatim, checked a 20% sample against the audio, and managed them in NVivo 12 Plus. Two researchers without intervention-delivery roles led the analysis and documented their disciplinary positions, prior arts experience, and assumptions about non-clinical support. A versioned coding framework, case summaries, and a case-by-theme matrix linked every quotation to source type, arm, university, participant code, and outcome trajectory. Reflexive thematic analysis guided interpretive development [34], while Framework Analysis supported systematic comparison [35]. We selected quotations to represent central patterns, variation, and negative cases; counts indicated substantive contribution to a theme rather than prevalence.

10. ETHICS AND DATA GOVERNANCE

The University Research Ethics Consortium approved the protocol before recruitment (10 February 2022). All students received plain-language information and provided written informed consent. Consent separately covered trial participation, audio recording, de-identified quotation, academic-functioning information, and follow-up contact. Participation was voluntary, students could decline individual activities without penalty, and withdrawal did not affect academic standing or access to services. Identifying information was separated from coded records. Electronic files were encrypted, role-restricted, and audit logged; paper materials were held in locked facilities. Quotations were screened to minimise deductive identification, and risk-sensitive referral pathways were active throughout.

11. STATISTICAL ANALYSIS

Analyses followed the intention-to-treat principle. Quantitative analyses used R 4.2.2 with lme4 1.1-31, clubSandwich 0.5.8, mice 3.15.0, emmeans 1.8.3, and lavaan 0.6-14. Baseline characteristics, recruitment, observed sample sizes, fidelity, process indicators, co-interventions, adverse events, and missingness were summarised by randomized arm. McDonald's omega assessed internal consistency. Longitudinal measurement checks evaluated configural, metric, and scalar stability of the DASS-21 total and mediator measures.

For participant i in delivery group g, delivered by facilitator f within randomized arm a, at occasion t, the primary model was $Y_{igfat} = \beta0 + \beta1Arm_i + \beta2Time_t + \beta3(Arm\times Time)_{it} + \beta4Baseline_i + \beta5University_i + \beta6Severity_i + u_i + b_{f(a)} + c_{g(f,a)} + \varepsilon_{igfat}$. Random intercepts u_i, $b_{f(a)}$, and $c_{g(f,a)}$ represented participants, facilitators nested within arms, and delivery groups nested within facilitators and arms; within-participant residuals followed an AR(1) structure. Restricted maximum likelihood and Kenward–Roger small-sample degrees-of-freedom correction were used. The three week-10 arts-versus-control comparisons were adjusted by Holm's procedure, and secondary outcome families and 15 moderator tests used Benjamini–Hochberg false-discovery-rate control. Standardized effects were calculated as the adjusted mean difference divided by the pooled baseline standard deviation of the corresponding outcome. For the primary DASS-21 total, the pooled baseline standardizer was 8.70 points: d = |adjusted mean difference|/8.70. Adjusted and unadjusted estimates, 95% confidence intervals, standardized effects, corrected degrees of freedom, exact p values, and q values were retained. Alternative covariance structures, variance-inflation factors, residual diagnostics, influence statistics, and singularity checks were examined.

Forty multiple-imputation datasets were generated by chained equations using arm, university, outcomes, demographic variables, study level, living situation, counselling, medication, process indicators, and recorded reasons for missingness. Trace plots, distributional overlap, Monte Carlo error, relative efficiency, and fractions of missing information were inspected. Complete-case, per-protocol, alternative

covariance, university-random-effect, protocol-deviation, and delta-adjusted missing-not-at-random analyses assessed robustness.

Parallel-mediator models estimated total, direct, specific indirect, and combined indirect effects from randomized assignment to week-5 changes in emotion regulation, coping self-efficacy, creative self-efficacy, and social connectedness, followed by week-10 distress. The pooled arts-versus-control mechanism model was primary; arm-specific decompositions were exploratory because 56 delivery groups and 20 facilitators provided limited higher-level precision. Delivery-group clustering was handled with cluster-robust standard errors and 5,000 cluster bootstrap resamples. Model fit, mediator residual correlations, leave-one-facilitator-out analyses, and alternative group-centering specifications were reported. Interpretation required consistency, no unmeasured mediator–outcome confounding, correct temporal ordering, and no exposure-induced mediator–outcome confounder; these assumptions were treated as strong and unverifiable [36]. Cluster-level uncertainty was interpreted conservatively [37]. Attendance was analysed separately as a post-randomization exposure and not causally.

IV. RESULTS

1. PARTICIPANT FLOW AND BASELINE CHARACTERISTICS

Of 734 students screened, 448 met eligibility criteria, completed baseline assessment, and were randomized equally across the four conditions. Week-5 data were observed for 421 students, week-10 data for 396, three-month data for 383, and six-month data for 368. The primary intention-to-treat model included all 448 randomized students through multiple imputation. Attrition was slightly higher in active control, but the arm difference was not statistically significant. Lower attendance, first-year status, and living off campus were associated with missing follow-up; baseline distress and randomized arm were not independently associated with missingness after adjustment.

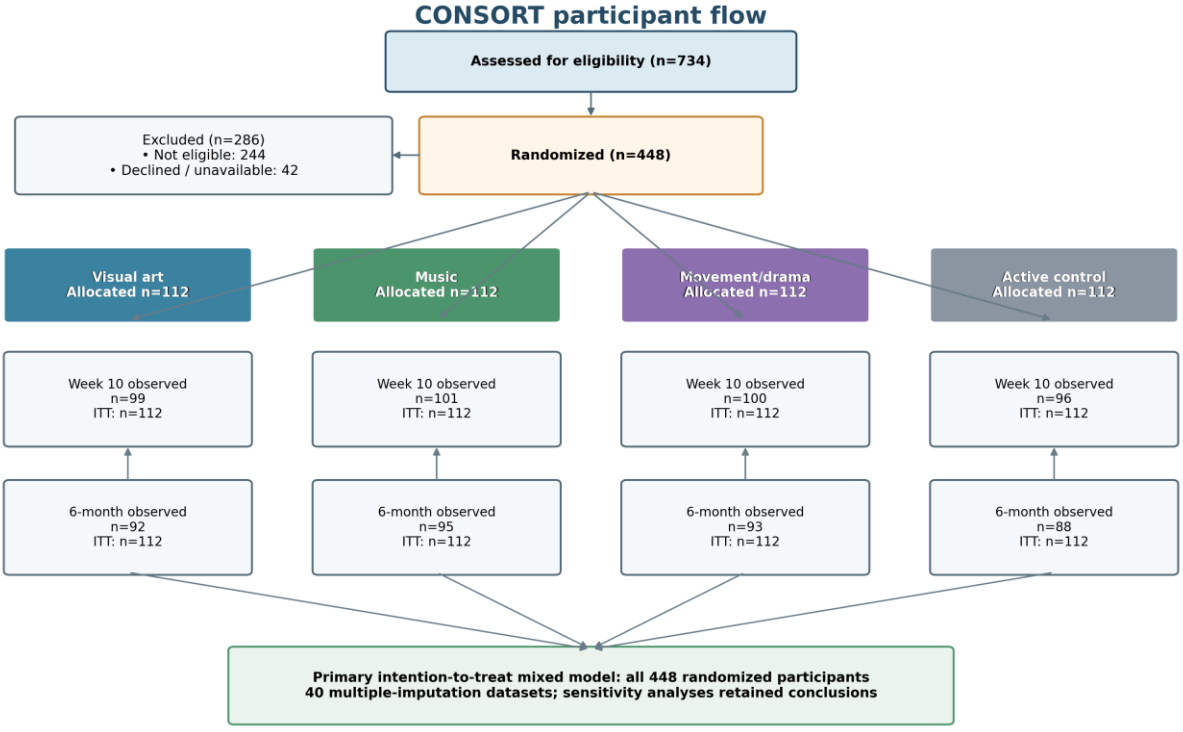


FIGURE 3. CONSORT participant flow through screening, randomization, repeated assessment, and intention-to-treat analysis.

Table 4. Combined Movement/Drama

Characteristic	Visual art (n=112)	Music (n=112)	Combined movement/drama (n=112)	Control (n=112)
Age, mean (SD)	21.6 (2.8)	21.7 (2.9)	21.5 (2.7)	21.8 (2.9)
Women, n (%)	72 (64.3)	70 (62.5)	73 (65.2)	71 (63.4)
Men, n (%)	35 (31.3)	37 (33.0)	34 (30.4)	36 (32.1)
Non-binary / another identity, n (%)	5 (4.5)	5 (4.5)	5 (4.5)	5 (4.5)
First-year students, n (%)	42 (37.5)	41 (36.6)	43 (38.4)	42 (37.5)
University accommodation, n (%)	49 (43.8)	50 (44.6)	48 (42.9)	49 (43.8)
Previous organized arts experience, n (%)	38 (33.9)	40 (35.7)	39 (34.8)	38 (33.9)
Current counseling, n (%)	15 (13.4)	14 (12.5)	16 (14.3)	15 (13.4)
Psychotropic medication, n (%)	11 (9.8)	12 (10.7)	11 (9.8)	12 (10.7)
Stable medication ≥6 weeks, n (%)	9 (8.0)	10 (8.9)	9 (8.0)	10 (8.9)
Raw DASS-21 total, mean (SD)	29.6 (8.7)	29.4 (8.6)	29.7 (8.8)	29.5 (8.7)
Raw stress subscale, mean (SD)	11.3 (3.8)	11.2 (3.7)	11.4 (3.9)	11.3 (3.8)
Raw anxiety subscale, mean (SD)	8.7 (3.6)	8.6 (3.5)	8.8 (3.7)	8.7 (3.6)
Raw depression subscale, mean (SD)	9.6 (3.9)	9.6 (3.8)	9.5 (3.9)	9.5 (3.8)
WEMWBS, mean (SD)	39.4 (7.9)	39.7 (8.0)	39.3 (7.8)	39.6 (8.1)
DERS-16, mean (SD)	49.8 (10.2)	49.5 (10.0)	50.1 (10.4)	49.7 (10.1)
Coping self-efficacy, mean (SD)	112.1 (25.8)	111.7 (26.1)	111.2 (25.9)	112.4 (26.0)
Creative self-efficacy, mean (SD)	14.7 (3.1)	14.8 (3.0)	14.6 (3.2)	14.7 (3.1)
SCS-R, mean (SD)	39.8 (8.1)	40.1 (8.0)	39.6 (8.3)	39.9 (8.2)
Study-adapted WSAS, mean (SD)	17.1 (5.6)	16.9 (5.5)	17.2 (5.7)	17.0 (5.6)

Characteristic	Visual art (n=112)	Music (n=112)	Combined movement/drama (n=112)	Control (n=112)
WHOQOL physical domain, mean (SD)	57.9 (11.8)	58.1 (11.6)	57.6 (11.9)	58.0 (11.7)
WHOQOL psychological domain, mean (SD)	51.1 (12.3)	51.4 (12.1)	50.9 (12.4)	51.2 (12.2)
WHOQOL social domain, mean (SD)	55.2 (14.0)	55.6 (13.8)	55.0 (14.2)	55.4 (13.9)
WHOQOL environmental domain, mean (SD)	60.2 (10.7)	60.4 (10.5)	60.0 (10.8)	60.3 (10.6)
Any baseline item missing, n (%)	2 (1.8)	2 (1.8)	3 (2.7)	2 (1.8)

2. MEASUREMENT QUALITY, BLINDING, AND IMPLEMENTATION

Omega coefficients ranged from 0.82 to 0.94 across assessment occasions. Longitudinal checks supported stable interpretation of the raw DASS-21 total, coping self-efficacy, creative self-efficacy, and connectedness. Two DERS-16 item intercepts were freed to achieve partial scalar invariance; summed-score and factor-score sensitivity analyses produced materially equivalent estimates. Assessors remained unaware of allocation in 93.1% of completed week-10 visits, and allocation guesses did not exceed chance after "unsure" responses were retained.

All 56 delivery groups commenced. Mean attendance ranged from 8.3 to 8.8 sessions, and 80.4%–89.3% of participants completed at least eight sessions. Core-component fidelity exceeded 90% in every arm, and cross-condition contamination was below 2% of observed components. CEQ credibility and expectancy were balanced after session 1. WAI-SR alliance and GCQ-S engagement were stronger in the arts conditions than in control, while contact time and recorded facilitator responsiveness were comparable. CSQ-8 satisfaction means ranged from 24.9 in control to 27.3 in music. Co-interventions and medication changes were uncommon and balanced. We documented 20 protocol deviations; none altered eligibility, randomization, or the primary estimand. Because facilitator background and post-randomization relational processes could contribute to treatment differences, exploratory sensitivity models added session-1 expectancy and then week-5 alliance and engagement. Expectancy adjustment changed arts-versus-control contrasts by no more than 0.2 points; additional alliance and engagement adjustment attenuated them by 0.3–0.5 points, while all three remained directionally consistent. These models were not treated as primary because alliance and engagement were measured after randomization.

Table 5. Delivery, expectancy, process outcomes, co-interventions, and fidelity.

Implementation indicator	Visual art	Music	Combined movement/drama	Control
Delivery groups / facilitators	14 / 5	14 / 5	14 / 5	14 / 5
Sessions attended, mean (SD)	8.5 (1.8)	8.8 (1.6)	8.6 (1.7)	8.3 (2.0)

https://doi.org/10.48161/ajst.v3n3a98

Implementation indicator	Visual art	Music	Combined movement/drama	Control
Completed ≥8 sessions, % (95% CI)	84.8 (77.3–91.0)	89.3 (82.5–94.2)	87.5 (80.3–92.9)	80.4 (72.0–87.3)
Core-component fidelity, % (95% CI)	92.4 (89.8–95.0)	93.8 (91.4–96.2)	92.9 (90.2–95.6)	91.5 (88.6–94.4)
CEQ credibility, 1–9, mean (SD)	6.7 (1.3)	6.8 (1.2)	6.7 (1.3)	6.6 (1.4)
CEQ expectancy, 0–100, mean (SD)	63.8 (15.2)	65.1 (14.7)	64.3 (15.0)	62.7 (15.8)
WAI-SR alliance, 1–5, mean (SD)	4.10 (0.54)	4.18 (0.51)	4.12 (0.55)	3.83 (0.62)
GCQ-S engagement, 0–6, mean (SD)	4.52 (0.72)	4.68 (0.68)	4.59 (0.70)	4.13 (0.81)
CSQ-8 satisfaction, 8–32, mean (SD)	26.8 (3.7)	27.3 (3.4)	27.0 (3.6)	24.9 (4.2)
Any co-intervention, n (%)	18 (16.1)	17 (15.2)	19 (17.0)	20 (17.9)
Medication change, n (%)	5 (4.5)	4 (3.6)	5 (4.5)	6 (5.4)
Protocol deviations, n	5	4	6	5

3. PRIMARY OUTCOME AND PRACTICAL RESPONSE

Raw DASS-21 total scores decreased in all conditions, with larger reductions in the three arts programs. At week 10, adjusted mean differences versus active control were −4.5 points for visual art (95% CI −6.1 to −2.9; Holm-adjusted $p<0.001$), −5.3 for music (−6.9 to −3.7; $p<0.001$), and −4.9 for the combined movement/drama program (−6.5 to −3.3; $p<0.001$). Using the pooled baseline DASS-21 standard deviation of 8.70 as the prespecified standardizer, the corresponding benefit-oriented standardized effects were 0.52, 0.61, and 0.56. Kenward–Roger corrected degrees of freedom were 45.8, 46.1, and 45.6, respectively. Unadjusted contrasts differed from adjusted estimates by no more than 0.3 points. Direct arts-arm contrasts were small and uncertain: music versus visual art −0.8 points (95% CI −2.2 to 0.6), music versus combined movement/drama −0.4 (−1.8 to 1.0), and combined movement/drama versus visual art −0.4 (−1.8 to 1.0). The three arts conditions therefore produced similar average effects, and direct comparisons were inconclusive.

At six months, adjusted differences versus control were −3.2 points for visual art, −4.0 for music, and −3.6 for combined movement/drama. Using the study-specific five-point response threshold, week-10 response probabilities were 68.2%, 73.5%, and 71.1% in the arts conditions and 45.8% in control. We estimated a reliable improvement of at least eight raw points in 48.1%, 54.0%, 51.2%, and 29.0%, respectively. Reliable deterioration of at least eight points was uncommon. These thresholds supported practical interpretation but were not treated as externally validated clinical cut-offs.

https://doi.org/10.48161/ajst.v3n3a98

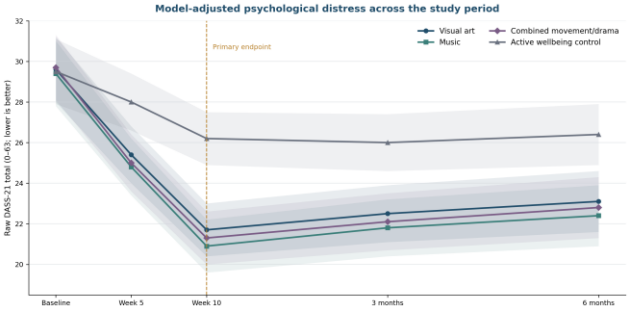


FIGURE 4. Model-adjusted raw DASS-21 total trajectories with 95% confidence intervals across five assessment occasions. Lower scores indicate less distress; observed sample sizes are reported in Table 6.

Table 6. Model-adjusted raw DASS-21 total means and observed sample sizes.

Time point	Visual art adjusted mean (95% CI)	Music adjusted mean (95% CI)	Combined movement/drama adjusted mean (95% CI)	Control adjusted mean (95% CI)
Baseline	29.6 (28.0–31.2); n=112	29.4 (27.8–31.0); n=112	29.7 (28.1–31.3); n=112	29.5 (27.9–31.1); n=112
Week 5	25.4 (24.0–26.8); n=106	24.8 (23.4–26.2); n=107	25.0 (23.6–26.4); n=105	28.0 (26.6–29.4); n=103
Week 10	21.7 (20.4–23.0); n=99	20.9 (19.6–22.2); n=101	21.3 (20.0–22.6); n=100	26.2 (24.9–27.5); n=96
3 months	22.5 (21.1–23.9); n=96	21.8 (20.4–23.2); n=98	22.1 (20.7–23.5); n=97	26.0 (24.6–27.4); n=92
6 months	23.1 (21.6–24.6); n=92	22.4 (20.9–23.9); n=95	22.8 (21.3–24.3); n=93	26.4 (24.9–27.9); n=88

Table 7. Pooled marginal study-specific response and reliable-change estimates at week 10; percentages are model-based, not raw counts.

Responder indicator	Visual art	Music	Combined movement/drama	Control
Study-specific practical response: ≥5-point decrease	68.2% (59.4–77.0)	73.5% (65.1–81.9)	71.1% (62.5–79.7)	45.8% (36.6–55.0)
Risk difference vs control	22.4% (9.7–35.1)	27.7% (15.3–40.1)	25.3% (12.8–37.8)	Reference
Number needed to treat	4.5 (2.8–10.3)	3.6 (2.5–6.5)	4.0 (2.6–7.8)	Reference
Reliable improvement: ≥8-point decrease	48.1% (38.8–57.4)	54.0% (44.6–63.4)	51.2% (41.8–60.6)	29.0% (20.6–37.4)
Reliable deterioration: ≥8-point increase	3.5% (0.3–6.7)	2.6% (0.0–5.5)	4.4% (0.6–8.2)	6.1% (1.7–10.5)

4. SECONDARY OUTCOMES

Secondary outcomes were coherent with the primary finding but varied by proximity to the intervention targets. Stress, wellbeing, emotion regulation, coping self-efficacy, creative self-efficacy, and connectedness showed the clearest differences. Anxiety and depression improved more modestly. WHOQOL-BREF psychological-domain effects were small but more consistent than physical, social, or environmental-domain effects. The study-adapted academic-functioning score improved modestly across arts conditions, but these

https://doi.org/10.48161/ajst.v3n3a98

findings remained exploratory. Music produced the largest mean stress reduction, the combined movement/drama program the largest emotion-regulation improvement, and visual art the largest creative self-efficacy effect; direct comparisons among arts arms were not conclusive. Standardized effects in Table 8 use the pooled baseline standard deviation for each corresponding outcome, and Figure 5 shows benefit-oriented signs.

Table 8. Adjusted week-10 effects versus active control. Standardized effects use pooled baseline SD. Negative differences indicate improvement; academic functioning is exploratory.

Outcome	Visual art vs control	Music vs control	Combined movement/drama vs control
Raw DASS-21 total	−4.5 (−6.1, −2.9); d=0.52; p<0.001	−5.3 (−6.9, −3.7); d=0.61; p<0.001	−4.9 (−6.5, −3.3); d=0.56; p<0.001
Raw stress subscale	−1.8 (−2.6, −1.0); d=0.47; q=0.001	−2.2 (−3.0, −1.4); d=0.58; q<0.001	−2.0 (−2.8, −1.2); d=0.53; q<0.001
Raw anxiety subscale	−1.2 (−1.9, −0.5); d=0.33; q=0.011	−1.5 (−2.2, −0.8); d=0.42; q=0.004	−1.4 (−2.1, −0.7); d=0.39; q=0.006
Raw depression subscale	−1.4 (−2.2, −0.6); d=0.36; q=0.008	−1.6 (−2.4, −0.8); d=0.42; q=0.004	−1.5 (−2.3, −0.7); d=0.39; q=0.006
WEMWBS wellbeing	4.1 (2.3, 5.9); d=0.52; q=0.002	5.0 (3.2, 6.8); d=0.63; q<0.001	4.7 (2.9, 6.5); d=0.59; q<0.001
DERS-16 emotion dysregulation	−4.0 (−5.8, −2.2); d=0.39; q=0.002	−4.5 (−6.3, −2.7); d=0.44; q<0.001	−4.8 (−6.6, −3.0); d=0.47; q<0.001
Coping self-efficacy	11.8 (6.2, 17.4); d=0.45; q=0.003	12.4 (6.8, 18.0); d=0.48; q=0.002	13.2 (7.6, 18.8); d=0.51; q=0.001
Creative self-efficacy	1.7 (0.9, 2.5); d=0.55; q<0.001	1.3 (0.5, 2.1); d=0.42; q=0.003	1.5 (0.7, 2.3); d=0.48; q=0.001
Social connectedness	3.7 (1.5, 5.9); d=0.45; q=0.009	4.8 (2.6, 7.0); d=0.59; q=0.002	4.3 (2.1, 6.5); d=0.53; q=0.004
Exploratory academic functioning	−1.9 (−3.2, −0.6); d=0.34; q=0.022	−2.1 (−3.4, −0.8); d=0.38; q=0.017	−2.4 (−3.7, −1.1); d=0.43; q=0.009
WHOQOL physical domain	2.0 (0.1, 3.9); d=0.17; q=0.094	2.5 (0.6, 4.4); d=0.21; q=0.061	2.3 (0.4, 4.2); d=0.20; q=0.071
WHOQOL psychological domain	3.4 (1.4, 5.4); d=0.28; q=0.028	4.1 (2.1, 6.1); d=0.33; q=0.014	3.8 (1.8, 5.8); d=0.31; q=0.019
WHOQOL social domain	1.7 (−0.6, 4.0); d=0.12; q=0.182	2.4 (0.1, 4.7); d=0.17; q=0.098	2.1 (−0.2, 4.4); d=0.15; q=0.126
WHOQOL environmental domain	1.3 (−0.5, 3.1); d=0.12; q=0.205	1.8 (0.0, 3.6); d=0.17; q=0.142	1.6 (−0.2, 3.4); d=0.15; q=0.164

https://doi.org/10.48161/ajst.v3n3a98

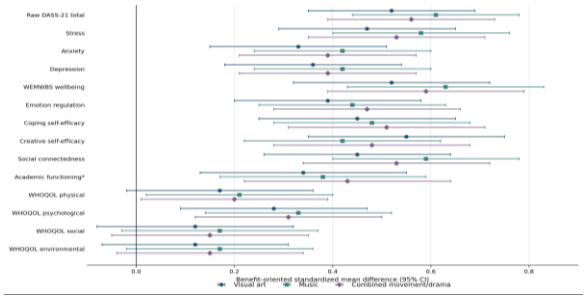


FIGURE 5. Benefit-oriented standardized week-10 effects versus active control. Standardized effects use the pooled baseline standard deviation of each outcome; values to the right of zero favor the arts condition. Academic functioning is exploratory.

5. *MEDIATION, MODERATION, AND ATTENDANCE*

The pooled parallel-mediator model supported partial statistical indirect pathways. For the pooled arts-versus-control contrast, the total week-10 effect was −4.90 raw DASS-21 points. The combined indirect effect was −2.65 points (95% cluster-bootstrap CI −3.41 to −1.89), leaving a residual direct effect of −2.25 points (−3.15 to −1.37). Specific pooled indirect effects were −0.81 through emotion regulation, −0.62 through coping self-efficacy, −0.58 through creative self-efficacy, and −0.64 through social connectedness. Model fit was acceptable (CFI=.946, TLI=.938, RMSEA=.048, SRMR=.052), and residual mediator correlations ranged from .22 to .46. The pooled model was the primary mechanism analysis. Arm-specific decompositions were exploratory because 56 groups and 20 facilitators provided limited higher-level precision; leave-one-facilitator-out analyses preserved direction but widened several confidence intervals. The results therefore described temporally ordered statistical pathways rather than definitive causal mediation.

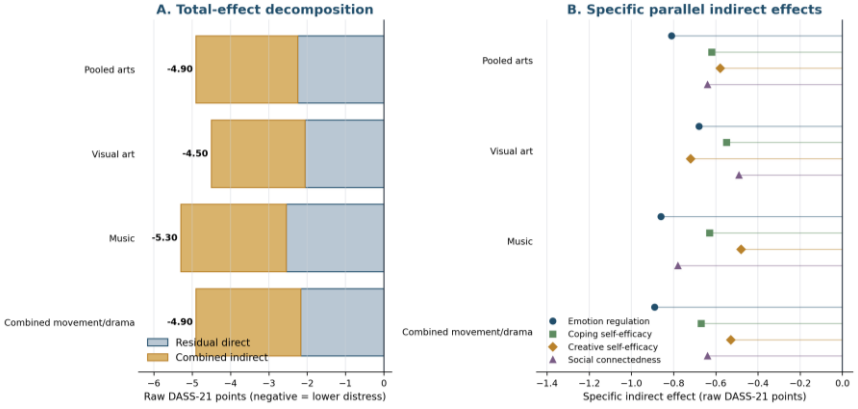


FIGURE 6. Temporally ordered parallel-mediator results. Panel A decomposes total effects into combined indirect and residual direct effects; Panel B shows the four specific indirect effects. Negative values indicate lower week-10 raw DASS-21 distress.

Table 9. Cluster-adjusted total, combined indirect, specific indirect, residual direct, and proportion-mediated estimates (95% cluster-bootstrap confidence intervals).

Panel A: effect decomposition	Total effect	Combined indirect	Residual direct	Proportion mediated	Analysis n
Pooled arts vs control	−4.90 (−6.02, −3.79)	−2.65 (−3.41, −1.89)	−2.25 (−3.15, −1.37)	54% (39–68)	448 ITT

Panel A: effect decomposition	Total effect	Combined indirect	Residual direct	Proportion mediated	Analysis n
Visual art	−4.50 (−6.10, −2.90)	−2.44 (−3.18, −1.72)	−2.06 (−3.01, −1.17)	54% (38–69)	448 ITT
Music	−5.30 (−6.90, −3.70)	−2.75 (−3.57, −1.94)	−2.55 (−3.43, −1.70)	52% (37–68)	448 ITT
Combined movement/drama	−4.90 (−6.50, −3.30)	−2.73 (−3.54, −1.92)	−2.17 (−3.07, −1.28)	56% (40–70)	448 ITT
Panel B: specific indirect effects	Emotion regulation	Coping self-efficacy	Creative self-efficacy	Social connectedness	Residual r range
Pooled arts vs control	−0.81 (−1.21, −0.46)	−0.62 (−0.99, −0.31)	−0.58 (−0.94, −0.27)	−0.64 (−1.03, −0.32)	.22–.46
Visual art	−0.68 (−1.12, −0.30)	−0.55 (−0.96, −0.22)	−0.72 (−1.16, −0.33)	−0.49 (−0.88, −0.18)	.22–.44
Music	−0.86 (−1.34, −0.44)	−0.63 (−1.07, −0.28)	−0.48 (−0.86, −0.16)	−0.78 (−1.26, −0.36)	.24–.46
Combined movement/drama	−0.89 (−1.39, −0.45)	−0.67 (−1.13, −0.30)	−0.53 (−0.93, −0.19)	−0.64 (−1.08, −0.27)	.23–.45

Baseline severity showed somewhat larger average arts effects among students with greater initial distress, but none of the arm-specific interactions remained below q=.05 after false-discovery-rate adjustment. Previous arts experience, study level, living situation, and the three-level gender-identity test did not consistently modify effects. Gender identities were not collapsed into a men/other category; sparse-cell estimates were interpreted through global interaction tests. Attendance had a positive, non-linear association with week-10 improvement through about eight sessions, then plateaued. Because attendance occurred after randomization, we interpreted it as descriptive process evidence rather than a causal dose effect.

Table 10. Exploratory effect-modification estimates for week-10 raw DASS-21 total. q values control the false-discovery rate across the 15 prespecified interaction tests.

Prespecified moderator	Visual art interaction	Music interaction	Combined movement/drama interaction
Baseline severity, per SD	−0.58 [−1.08, −0.08]; p=.024; q=.12	−0.69 [−1.20, −0.18]; p=.011; q=.08	−0.63 [−1.14, −0.12]; p=.016; q=.10
Previous arts experience	−0.21 [−1.42, 1.00]; q=.84	−0.35 [−1.55, 0.85]; q=.79	−0.18 [−1.38, 1.02]; q=.87
First year vs later study	−0.34 [−1.51, 0.83]; q=.78	−0.40 [−1.57, 0.77]; q=.76	−0.52 [−1.69, 0.65]; q=.69
Gender identity, global 3-level test	p=.61; q=.83	p=.54; q=.78	p=.58; q=.82
University accommodation	0.12 [−0.98, 1.22]; q=.91	−0.08 [−1.18, 1.02]; q=.94	0.05 [−1.05, 1.15]; q=.96

6. QUALITATIVE FINDINGS

We developed five cross-case themes and did not treat them as exhaustive or equally prevalent. First, symbolic or performative distance made difficult emotion approachable without compulsory disclosure. A visual-art participant explained, "I could put the pressure into the image without having to explain the whole story" (VA-P14, interview). A music participant described rhythm as "a way to settle before words arrived" (MU-P09, focus group). In combined movement/drama, one student valued "trying the feeling through a role, not announcing that it was mine" (MD-P05, interview). Choice over disclosure and public visibility distinguished manageable activation from withdrawal.

Second, attainable mastery restored agency. Students valued completing a drawing, maintaining a rhythm, revising a scene, or contributing to a collective piece. "Finishing something small made the rest of the week feel less impossible" (VA-P03, interview), while a music participant said, "I was not the best player, but the group needed my beat" (MU-P17, focus group). Mastery depended less on artistic quality than on clear progression, usable feedback, and recognition of contribution. Third, regulation became embodied and repeatable. A combined movement/drama participant reported, "I noticed my shoulders before the panic reached my thoughts" (MD-P11, interview), and a visual-art participant described tactile focus as "a pause button I could use before an exam" (VA-P08, focus group).

Fourth, reciprocal belonging differed from simple attendance. Students valued being noticed, depended upon, and able to contribute at different skill levels. "When I missed a week, people asked where I was; that was different from sitting in a lecture" (MU-P04, interview). A control participant also valued ordinary social contact: "The planning was useful because I realized other people were struggling with the same deadlines" (CT-P12, focus group). Fifth, transfer required personal relevance and continuity. "The breathing and stepping sequence came with me into presentations" (MD-P02, interview), whereas another participant said that practice disappeared when "assignments took every spare hour" (VA-P19, focus group).

Negative cases identified performance anxiety, unwanted exposure, sensory overload, discomfort with movement, cultural unfamiliarity, and frustration with group pacing. One student described improvisation as "another performance I could fail" (MD-P13, interview). A music participant withdrew from one activity because layered sound felt "too crowded to think" (MU-P21, journal). Some students improved on distress scores while reporting limited enjoyment, and a small group in active control reported substantial benefit from planning and discussion. These accounts indicated contributions from routine, expectation, attention, and ordinary social contact and prevented qualitative explanation from being treated as validation of statistical mediation.

Table 11. Case-by-theme matrix for the 56 participant process-evaluation cases; counts indicate substantive contribution rather than prevalence.

Cross-case theme	Visual art cases	Music cases	Combined movement/drama cases	Control cases	Primary interpretive content
Safe expression without compulsory disclosure	11	12	12	4	Choice, indirect expression, graded visibility
Attainable mastery and agency	10	11	10	5	Progression, feedback, usable accomplishment
Embodied and sensory regulation	8	12	13	3	Rhythm, breath, tactile focus, posture, movement
Reciprocal belonging	10	12	11	9	Mutual recognition, dependable contribution
Academic and everyday transfer	9	10	10	7	Examination coping, study routines, peer situations

Cross-case theme	Visual art cases	Music cases	Combined movement/drama cases	Control cases	Primary interpretive content
Boundary or negative-case evidence	6	5	7	5	Anxiety, overload, pain, cultural or pacing mismatch

7. MIXED-METHODS INTEGRATION AND SENSITIVITY

Quantitative and qualitative evidence converged on regulation, mastery, creative confidence, and connection but did not establish a modality hierarchy. Statistical mediation described temporally ordered associations; participant accounts explained how activities were experienced; fidelity and adverse-event records documented implementation; and the framework supplied theoretical interpretation. Visual art's stronger creative self-efficacy pathway aligned with material externalization and revision. Music's numerical advantage on stress was consistent with rapid rhythmic access and low verbal demand. Combined movement/drama accounts clarified how integrated bodily activation and role rehearsal supported regulation while also creating performance and accessibility barriers. Divergent cases indicated that common group factors remained relevant.

Table 12. Mixed-methods joint display distinguishing statistical mediation, participant explanation, implementation evidence, and theoretical interpretation.

Integrated finding	Statistical evidence	Participant explanation	Implementation / negative-case evidence	Interpretive status
Visual art strengthened creative confidence	Largest creative self-efficacy indirect estimate	Images supported symbolic distance and revision	Exposure concerns required control over visibility	Convergent but model-dependent
Music had the largest mean distress estimate	Largest total and stress effects; arts-arm contrasts inconclusive	Rhythm provided rapid regulation with low verbal demand	Sensory overload limited participation for a minority	Complementary; superiority not established
Combined movement/drama supported embodied coping	Largest regulation and exploratory academic-functioning estimates	Grounding and role rehearsal linked bodily cues with action	Combined arm and performance anxiety limited attribution	Modality explanation, not separate combined movement/drama effect
Attendance related to improvement	Non-linear association through approximately eight sessions	Trust and skills accumulated through repetition	Attendance was post-randomization and self-selected	Process association, not causal dose
Not all students benefited	Limited-change and deterioration subgroup	Exposure, overload, discomfort, mismatch	Adaptations and opt-out protected safety	Boundary condition and divergence

The primary model converged without singularity. Estimated variance components were 18.4 for the participant intercept, 2.1 for delivery groups, 0.7 for facilitators, and 72.5 for the within-participant residual, yielding a model-based marginal standard deviation of 9.68. Delivery-group and facilitator intraclass

correlations were 2.2% and 0.7%, respectively, and the within-participant AR(1) correlation was 0.64. Kenward–Roger degrees of freedom for the three confirmatory contrasts ranged from 45.6 to 46.1. Variance-inflation factors were below 2.1; residual plots showed no material heteroscedasticity or non-linearity; and alternative covariance structures produced similar information criteria and effects. Imputation distributions overlapped observed values, Monte Carlo error remained below 5% of standard errors, relative efficiency exceeded 0.98, and the largest fraction of missing information was 0.16. Complete-case, per-protocol, alternative covariance, random-university, protocol-deviation, delta-adjusted, expectancy-adjusted, and relational-process sensitivity analyses retained the direction and substantive interpretation of all primary comparisons.

Fourteen mild intervention-related adverse events were recorded: transient emotional activation during visual-art or music sessions, short-lived musculoskeletal discomfort during movement activities, and performance-related anxiety during drama activities. Events resolved through rest, adaptation, follow-up, or referral advice. No serious intervention-related adverse event occurred. The combined arm recorded the activity type associated with each event, but its design did not support separate estimates of movement versus drama risk.

Table 13. Sensitivity-analysis, diagnostic, missing-data, and adverse-event summary.

Analysis or safety indicator	Visual art	Music	Combined movement/drama	Control / overall
Complete-case week-10 AMD	−4.3 (−5.9, −2.7)	−5.1 (−6.7, −3.5)	−4.7 (−6.3, −3.1)	Reference
Per-protocol week-10 AMD	−4.8 (−6.5, −3.1)	−5.6 (−7.3, −3.9)	−5.2 (−6.9, −3.5)	Reference
Delta-adjusted MNAR AMD	−4.0 (−5.7, −2.3)	−4.7 (−6.4, −3.0)	−4.4 (−6.1, −2.7)	Reference
Expectancy-adjusted AMD	−4.4	−5.1	−4.7	Post-randomization sensitivity
Alliance/engagement-adjusted AMD	−4.1	−4.7	−4.4	Post-randomization sensitivity
Variance components	—	—	—	Participant 18.4; group 2.1; facilitator 0.7; residual 72.5
Kenward–Roger df	45.8	46.1	45.6	Confirmatory contrasts
Group / facilitator ICC	—	—	—	0.022 / 0.007
Within-participant AR(1)	—	—	—	0.64
Largest VIF / FMI	—	—	—	2.1 / 0.16
Mild related adverse events	4	3	6	1
Serious related adverse events	0	0	0	0

V. DISCUSSION

This multicentre comparative trial found that structured visual-art, music, and combined movement/drama programs reduced raw DASS-21 total psychological distress more than an active wellbeing control matched for attention, group contact, routine, and reflection. Effects were moderate at week 10 and remained evident six months later, although they attenuated after facilitated sessions ended. The three arts conditions produced similar average effects, and direct comparisons were inconclusive. This conclusion is more defensible than a claim of modality superiority because the trial was not powered for small arts-arm differences and movement and drama were delivered as one integrated condition. The standardized effects are reproducible from the reported formula and pooled baseline DASS-21 standard deviation rather than from an unspecified model denominator.

The practical-response analysis added information that standardized effects alone could not provide. The five-point study-specific threshold correctly represented approximately one-half of the baseline standard deviation after rounding, while the eight-point reliable-change threshold incorporated baseline reliability. Individual response probabilities were substantially higher in the arts conditions than in the control condition, although both thresholds were distribution-based and not externally validated clinical cut-offs. They contextualized average effects in a population selected for mild-to-moderate distress.

Outcomes closest to the intervention targets—emotion regulation, coping self-efficacy, creative self-efficacy, connectedness, stress, and wellbeing—showed larger and more consistent effects than the more distal WHOQOL-BREF domains. Anxiety and depression changed modestly, which was compatible with the low-intensity, non-therapy scope. Exploratory academic-functioning effects suggested possible transfer to concentration and task initiation, but the adapted WSAS did not justify strong claims about academic performance.

Temporally ordered mediation supplied a mechanism-oriented contribution, but its evidential status remained secondary. Week-5 changes in regulation, coping self-efficacy, creative self-efficacy, and connectedness jointly accounted statistically for approximately half of the week-10 effect. The corrected decompositions showed exact consistency between total, combined indirect, and direct effects. However, randomization protected the assignment–outcome relation, not the mediator–outcome relation. 56 groups and 20 facilitators limited higher-level precision, and qualitative themes explained participant experience without validating the causal mediation assumptions.

The effect magnitudes were broadly consistent with previous arts and student-intervention syntheses, while the active comparator provided a more demanding test than wait-list or no-treatment designs [4]–[6]. Modality-specific reviews provide convergent evidence for movement, drama, and visual art [7]–[9]. Art-therapy and arts-in-health reviews further emphasize heterogeneity, facilitator competence, and scope [10], [11]. Improvement in control confirmed that psychoeducation, routine, reflection, facilitator responsiveness, and social contact were active ingredients. The added arts effects therefore emerged beyond common group processes but did not imply that artistic content operated independently of those processes. The moderate attenuation by six months also aligns with evidence that sustained outcomes depend on continued practice and contextual reinforcement.

The Arts Engagement–Emotion Regulation Framework was refined in two ways. First, emotion regulation remained the central measured pathway, with coping self-efficacy, creative self-efficacy, and connectedness as complementary resources. Second, modality-specific lenses explained differences in experience without assigning exclusive mechanisms. Image-making, rhythm, movement, and role were not interchangeable, yet none operated through a single pathway. This hierarchical account is more parsimonious than theoretical stacking and more precise than a generic “arts participation” explanation.

Qualitative negative cases showed that psychological safety and cultural fit were components of intervention integrity rather than optional enhancements. A manual could be delivered accurately while failing functionally if students experienced activities as exposing, inaccessible, physically uncomfortable, or culturally alien. Facilitator preparation therefore requires graded participation, opt-out procedures, mobility and sensory adaptation, culturally responsive material selection, boundary management, adverse-event recognition, and referral pathways.

For urban public-university services serving students with mild-to-moderate distress, the findings support arts programs as potentially acceptable low-intensity complements within stepped care. Similar mean effects across programs favor offering informed choice rather than assigning students based on demographic assumptions. Nevertheless, facilitators with different professional backgrounds delivered the arts and control conditions, and alliance, engagement, and satisfaction were higher in the arts conditions. Sensitivity analyses indicated that expectancy and relational-process adjustment attenuated but did not eliminate the arts-versus-control contrasts. The added effects therefore cannot be attributed exclusively to artistic content; facilitation, credibility, group climate, and participant preference remain plausible nonspecific contributors. The study did not test digital or hybrid delivery; concerns about synchrony, materials, privacy, and embodied co-presence are implementation inferences rather than effectiveness findings.

1. STRENGTHS AND LIMITATIONS

Strengths included random allocation, centralized concealment, assessor blinding, an active comparator, multicentre delivery, repeated follow-up, explicit week-10 primary-outcome specification, raw DASS-21 scoring transparency, delivery-group and facilitator adjustment, matched contact time, temporal mediator measurement, fidelity assessment, adverse-event monitoring, representative quotations, negative-case sampling, and integrated interpretation. The statistical approach reported model structure, multiplicity control, covariance and imputation diagnostics, and robustness analyses.

Several limitations remain. Participants and facilitators could not be blinded, so expectancy and self-report effects were possible despite balanced early credibility. Facilitator qualifications differed by condition, and post-randomization alliance and engagement were stronger in the arts arms; sensitivity models reduced but did not remove this concern. The four urban public universities limit transferability to rural, private, distance-learning, and culturally distinct settings. The study excluded students with severe symptoms or acute risk. The intervention conditions bundled multiple activities, and the combined movement/drama arm prevented separate conclusions about movement and drama. The study-adapted academic-functioning measure remained exploratory, and WHOQOL-BREF effects were domain-specific and small. The mediation analysis established temporal order but not causal identification; 56 groups and 20 facilitators limited arm-specific precision. Attrition was moderate and may have remained informative despite imputation and sensitivity analyses.

VI. CONCLUSION

Structured visual-art, music, and combined movement/drama programs reduced raw DASS-21 total psychological distress among university students more than an active wellbeing control and retained part of their benefit six months after the intervention. The findings did not support a definitive ranking or separate conclusions about movement and drama. Instead, they indicated a shared architecture of emotion regulation, coping self-efficacy, creative self-efficacy, and social connection, with different experiential emphases across image-making, rhythm, and integrated bodily and role-based activity.

Implementation quality influenced whether students could use the programs safely and meaningfully. Benefit was associated with voluntary participation, achievable challenge, facilitator competence, cultural relevance, sensory and physical accessibility, and transfer into academic life. Performance anxiety, unwanted exposure, overload, and timetable pressure constrained engagement for some students. Arts-based programs should therefore be positioned as structured, monitored additions to university wellbeing provision, not as generic recreation or replacements for indicated clinical care.

The principal contribution is a comparative, mechanism-informed account of arts participation for students with mild-to-moderate distress in urban public universities. Within that scope, the programmes represented potentially acceptable low-intensity options when delivered by qualified facilitators with fidelity and safety monitoring. Claims about other university sectors, severe disorders, separate movement and drama effects, or digital delivery extend beyond the evidence.

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]. R. P. Auerbach et al., "Mental disorders among college students in the World Health Organization World Mental Health Surveys," Psychological Medicine, vol. 46, no. 14, pp. 2955–2970, 2016, doi: 10.1017/S0033291716001665.

[2]. R. Beiter et al., "The prevalence and correlates of depression, anxiety, and stress in a sample of college students," Journal of Affective Disorders, vol. 173, pp. 90–96, 2015, doi: 10.1016/j.jad.2014.10.054.

[3]. W. Li, Z. Zhao, D. Chen, Y. Peng, and Z. Lu, "Prevalence and associated factors of depression and anxiety symptoms among college students: A systematic review and meta-analysis," Journal of Child Psychology and Psychiatry, vol. 63, no. 11, pp. 1222–1230, 2022, doi: 10.1111/jcpp.13606.

[4]. J. Huang, Y. T. Nigatu, R. Smail-Crevier, X. Zhang, and J. Wang, "Interventions for common mental health problems among university and college students: A systematic review and meta-analysis of randomized controlled trials," Journal of Psychiatric Research, vol. 107, pp. 1–10, 2018, doi: 10.1016/j.jpsychires.2018.09.018.

[5]. D. Fancourt, H. Aughterson, S. Finn, E. Walker, and A. Steptoe, "How leisure activities affect health: A narrative review and multi-level theoretical framework of mechanisms of action," The Lancet Psychiatry, vol. 8, no. 4, pp. 329–339, 2021, doi: 10.1016/S2215-0366(20)30384-9.

[6]. M. de Witte, A. Spruit, S. van Hooren, X. Moonen, and G. J. J. M. Stams, "Effects of music interventions on stress-related outcomes: A systematic review and two meta-analyses," Health Psychology Review, vol. 14, no. 2, pp. 294–324, 2020, doi: 10.1080/17437199.2019.1627897.

[7]. S. C. Koch, R. F. F. Riege, K. Tisborn, J. Biondo, L. Martin, and A. Beelmann, "Effects of dance movement therapy and dance on health-related psychological outcomes: A meta-analysis update," Frontiers in Psychology, vol. 10, art. 1806, 2019, doi: 10.3389/fpsyg.2019.01806.

[8]. H. Orkibi, S. Keisari, N. L. Sajnani, and M. de Witte, "Effectiveness of drama-based therapies on mental health outcomes: A systematic review and meta-analysis of controlled studies," Psychology of Aesthetics, Creativity, and the Arts, advance online publication, Apr. 27, 2023, doi: 10.1037/aca0000582.

[9]. G. Kaimal, K. Ray, and J. Muniz, "Reduction of cortisol levels and participants' responses following art making," Art Therapy, vol. 33, no. 2, pp. 74–80, 2016, doi: 10.1080/07421656.2016.1166832.

[10]. A. Abbing, A. Ponstein, S. van Hooren, L. de Sonneville, H. Swaab, and E. Baars, "The effectiveness of art therapy for anxiety in adults: A systematic review of randomized and non-randomized controlled trials," PLOS ONE, vol. 13, no. 12, art. e0208716, 2018, doi: 10.1371/journal.pone.0208716.

[11]. A. Jensen and L. O. Bonde, "The use of arts interventions for mental health and wellbeing in health settings," Perspectives in Public Health, vol. 138, no. 4, pp. 209–214, 2018, doi: 10.1177/1757913918772602.

[12]. J. J. Gross, "Emotion regulation: Current status and future prospects," Psychological Inquiry, vol. 26, no. 1, pp. 1–26, 2015, doi: 10.1080/1047840X.2014.940781.

[13]. B. L. Fredrickson, "The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions," American Psychologist, vol. 56, no. 3, pp. 218–226, 2001, doi: 10.1037/0003-066X.56.3.218.

[14]. 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.

[15]. L. W. Barsalou, "Grounded cognition," Annual Review of Psychology, vol. 59, pp. 617–645, 2008, doi: 10.1146/annurev.psych.59.103006.093639.

[16]. V. B. Lusebrink, "Assessment and therapeutic application of the Expressive Therapies Continuum: Implications for brain structures and functions," Art Therapy, vol. 27, no. 4, pp. 168–177, 2010, doi: 10.1080/07421656.2010.10129380.
[17]. J. Holt-Lunstad, T. B. Smith, M. Baker, T. Harris, and D. Stephenson, "Loneliness and social isolation as risk factors for mortality: A meta-analytic review," Perspectives on Psychological Science, vol. 10, no. 2, pp. 227–237, 2015, doi: 10.1177/1745691614568352.
[18]. K. F. Schulz, D. G. Altman, and D. Moher, "CONSORT 2010 statement: Updated guidelines for reporting parallel group randomized trials," BMJ, vol. 340, art. c332, 2010, doi: 10.1136/bmj.c332.
[19]. A.-W. Chan et al., "SPIRIT 2013 statement: Defining standard protocol items for clinical trials," BMJ, vol. 346, art. e7586, 2013, doi: 10.1136/bmj.e7586.
[20]. 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.
[21]. 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.
[22]. J. D. Henry and J. R. Crawford, "The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample," British Journal of Clinical Psychology, vol. 44, no. 2, pp. 227–239, 2005, doi: 10.1348/014466505X29657.
[23]. R. Tennant et al., "The Warwick–Edinburgh Mental Well-being Scale (WEMWBS): Development and UK validation," Health and Quality of Life Outcomes, vol. 5, art. 63, 2007, doi: 10.1186/1477-7525-5-63.
[24]. J. Bjureberg et al., "Development and validation of a brief version of the Difficulties in Emotion Regulation Scale: The DERS-16," Journal of Psychopathology and Behavioral Assessment, vol. 38, no. 2, pp. 284–296, 2016, doi: 10.1007/s10862-015-9514-x.
[25]. M. A. Chesney, D. B. Neilands, D. B. Chambers, J. M. Taylor, and S. Folkman, "A validity and reliability study of the Coping Self-Efficacy Scale," British Journal of Health Psychology, vol. 11, no. 3, pp. 421–437, 2006, doi: 10.1348/135910705X53155.
[26]. P. Tierney and S. M. Farmer, "Creative self-efficacy: Its potential antecedents and relationship to creative performance," Academy of Management Journal, vol. 45, no. 6, pp. 1137–1148, 2002, doi: 10.2307/3069429.
[27]. R. M. Lee, M. Draper, and S. Lee, "Social connectedness, dysfunctional interpersonal behaviors, and psychological distress: Testing a mediator model," Journal of Counseling Psychology, vol. 48, no. 3, pp. 310–318, 2001, doi: 10.1037/0022-0167.48.3.310.
[28]. J. C. Mundt, I. M. Marks, M. K. Shear, and J. M. Greist, "The Work and Social Adjustment Scale: A simple measure of impairment in functioning," British Journal of Psychiatry, vol. 180, no. 5, pp. 461–464, 2002, doi: 10.1192/bjp.180.5.461.
[29]. S. M. Skevington, M. Lotfy, and K. A. O'Connell, "The World Health Organization's WHOQOL-BREF quality of life assessment: Psychometric properties and results of the international field trial," Quality of Life Research, vol. 13, no. 2, pp. 299–310, 2004, doi: 10.1023/B:QURE.0000018486.91360.00.
[30]. G. J. Devilly and T. D. Borkovec, "Psychometric properties of the Credibility/Expectancy Questionnaire," Journal of Behavior Therapy and Experimental Psychiatry, vol. 31, no. 2, pp. 73–86, 2000, doi: 10.1016/S0005-7916(00)00012-4.
[31]. R. L. Hatcher and J. A. Gillaspy, "Development and validation of a revised short version of the Working Alliance Inventory," Psychotherapy Research, vol. 16, no. 1, pp. 12–25, 2006, doi: 10.1080/10503300500352500.
[32]. J. E. Johnson, D. Pulsipher, S. L. Ferrin, G. M. Burlingame, D. R. Davies, and R. Gleave, "Measuring group processes: A comparison of the GCQ and CCI," Group Dynamics: Theory, Research, and Practice, vol. 10, no. 2, pp. 136–145, 2006, doi: 10.1037/1089-2699.10.2.136.
[33]. D. L. Larsen, C. C. Attkisson, W. A. Hargreaves, and T. D. Nguyen, "Assessment of client/patient satisfaction: Development of a general scale," Evaluation and Program Planning, vol. 2, no. 3, pp. 197–207, 1979, doi: 10.1016/0149-7189(79)90094-6.
[34]. V. Braun and V. Clarke, "One size fits all? What counts as quality practice in (reflexive) thematic analysis?" Qualitative Research in Psychology, vol. 18, no. 3, pp. 328–352, 2021, doi: 10.1080/14780887.2020.1769238.
[35]. 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.
[36]. 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.
[37]. D. McNeish and L. M. Stapleton, "Modeling clustered data with very few clusters," Multivariate Behavioral Research, vol. 51, no. 4, pp. 495–518, 2016, doi: 10.1080/00273171.2016.1167008.

https://doi.org/10.48161/ajst.v3n3a98

APPENDIX A. CONDENSED INTERVENTION DESCRIPTION

This appendix provides a concise TIDieR-oriented extract. The separate supplementary package includes the full session manuals, facilitator scripts, adaptation boundaries, and home-practice materials.

Table A1. Condensed intervention and interpretation boundaries.

Condition	Core content	Dose	Facilitator profile	Interpretive boundary
Visual art	Drawing, collage, clay, visual metaphor, collaborative installation	10 × 75 min	Qualified visual-art/art-education facilitators	Supports conclusions about the complete visual-art program
Music	Rhythm, voice, accessible instruments, improvisation, shared composition	10 × 75 min	Qualified music practitioners	Supports conclusions about the complete music program
Combined movement/drama	~55% grounding, mirroring and movement; ~45% role, dialogue and ensemble scenes	10 × 75 min	Qualified combined movement/drama practitioners	Does not support separate movement or drama effects
Active wellbeing control	Psychoeducation, structured discussion, non-art problem solving and planning	10 × 75 min	Counseling and health-promotion facilitators	Controls attention, routine, reflection and social contact

APPENDIX B. DASS-21 SCORING AND RESPONSE CALCULATIONS

DASS-21 item scores were summed without doubling. The raw total ranged from 0 to 63, and each subscale ranged from 0 to 21. We prespecified the practical-response and reliable-change thresholds as study-specific, distribution-based indicators.

Table B1. DASS-21 scoring, responder thresholds, and primary effect-size standardization.

Indicator	Calculation	Threshold
Primary score	Sum of 21 items scored 0–3	Raw DASS-21 total 0–63
Practical response	0.5 × baseline SD = 0.5 × 8.70 = 4.35; rounded upward	≥5-point decrease
Reliable change	$1.96 \times 8.70 \times \sqrt{[2(1-0.91)]} = 7.23$; rounded upward	≥8-point decrease or increase
Primary standardized effect	d = |adjusted mean difference| / pooled baseline SD	Standardizer = 8.70

APPENDIX C. PRIMARY MODEL AND DIAGNOSTIC EXTRACT

Participants were nested within delivery groups, groups within facilitators, and facilitators within randomized arms. Arm-specific mediation decompositions were exploratory; the pooled arts-versus-control mediation model was primary.

Diagnostic	Result	Interpretation
Variance components	Participant 18.4; group 2.1; facilitator 0.7; residual 72.5	Small but non-zero group and facilitator clustering
Corrected df	45.6–46.1 for confirmatory contrasts	Uncertainty reflects higher-level sample size
ICC / AR(1)	Group .022; facilitator .007; within-participant .64	Nested and repeated dependence retained

Diagnostic	Result	Interpretation
Imputation	40 datasets; maximum FMI .16; relative efficiency >.98	Adequate imputation precision
Pooled mediation	Total −4.90; indirect −2.65; direct −2.25	Primary mechanism estimate
Relational sensitivity	Arts contrasts attenuated 0.3–0.5 points after alliance/engagement adjustment	Nonspecific relational effects remain plausible

Table C1. Essential model, clustering, missing-data, and mediation diagnostics.

APPENDIX D. QUALITATIVE AND SAFETY TRACEABILITY EXTRACT

We linked representative quotations to intervention arm, source type, participant code, university, attendance, and outcome trajectory. Theme counts denote substantive contribution rather than prevalence. We classified adverse events as mild, moderate, or serious, and by expectedness and relatedness; no serious related events occurred. The complete codebook, transcript extracts, quotation matrix, and adverse-event procedure are included in the separate supplementary package.

Table D1. Condensed qualitative quotation and safety traceability matrix.

Theme / safety domain	Representative source	Boundary evidence	Traceability field
Safe indirect expression	VA-P14 interview; MU-P09 focus group; MD-P05 interview	Unwanted exposure or performance anxiety	Arm, university, source, trajectory
Attainable mastery	VA-P03 interview; MU-P17 focus group	Comparison and perceived failure	Arm, attendance, source, theme
Embodied regulation	MD-P11 interview; VA-P08 focus group	Pain or sensory overload	Activity type and adverse-response status
Reciprocal belonging	MU-P04 interview; CT-P12 focus group	Uneven pacing or exclusion	Group, facilitator and case memo

1. SAFETY CLASSIFICATION EXTRACT

Table D2. Condensed adverse-event classification and response requirements.

Category	Operational definition	Required response
Mild related event	Transient emotional, sensory, musculoskeletal, or performance discomfort requiring brief adaptation	Pause or adapt activity; document and follow up
Moderate related event	Symptoms requiring session withdrawal, clinical consultation, or sustained modification	Coordinator review, safety assessment, referral advice, and monitoring
Serious adverse event	Death, life-threatening event, hospitalization, persistent disability, or urgent safeguarding event	Immediate clinical or safeguarding response and independent review
Expectedness and relatedness	Manual-listed risk and reasonable temporal/mechanistic link	Classify, document rationale, and review mitigation