

Supplementary Materials

Supplementary Table S1. PRISMA 2020 checklist	2
Supplementary Methods S1. Data-conversion equations	7
Supplementary Methods S2. Additional statistical analyses	9
Supplementary Table S2. Search Results.....	10
Supplementary Figure S1. Forest plot for VO2max.....	11
Supplementary Figure S2. Forest plot for flexibility	11
Supplementary Figure S3. Forest plot for BMI.....	11
Supplementary Figure S4. Forest plot for muscular fitness	12
Supplementary Table S3. Original reported results for psychological, psychosocial, cognitive, and behavioral outcomes.....	13
Supplementary Table S4. Quantitative data and between-group/within-group results for psychological, psychosocial, cognitive, and behavioral outcomes.....	21
Supplementary Table S5. Operational criteria used for the GRADE certainty assessment	28
Supplementary Table S6. Full subgroup analyses for muscular fitness.....	29
Supplementary Table S7. Full meta-regression analyses for muscular fitness	30
Supplementary Figure S5. Full meta-regression plots for muscular fitness.....	30
Supplementary Figure S6. Funnel plot and sunset plot for muscular fitness	31
Supplementary Figure S7. Leave-one-out sensitivity analyses	32
Supplementary Table S8. Methodological quality assessment [PEDro]	33
Supplementary Table S9. Risk-of-bias domain-level judgements.....	35
Supplementary Table S10. Sensitivity analyses for overlapping reports and cohorts.....	36
Supplementary Figure S8. Cohort-level sensitivity forest plot for muscular fitness	37
Supplementary Table S11. Cohort-level sensitivity subgroup analyses for muscular fitness.....	38
Supplementary Table S12. Cohort-level sensitivity meta-regression analyses for muscular fitness	39

Supplementary Table S1. PRISMA 2020 checklist

Section and Topic	Item#	Checklist Item	Location where item is reported
TITLE			
Title	1	Identify the report as a systematic review.	Page 1
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Page 2-3
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Page 4-6
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Page 6
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Page 9, Table 1
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Page 7
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Page 7-8, Supplementary Table S2
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Page 7-8
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Page 11
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which	Sections 2.3, 2.4, 3.2, 3.3; Tables 1-2;

		results to collect.	Supplementary Tables S3-S4
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Sections 2.4 and 2.5; Table 2
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Section 2.6
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Sections 2.5 and 2.7; Figure 2 and Table 3
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Section 2.3, 2.4, 2.7 and 3.3
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Section 2.5; Section 2.4
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Sections 2.7 and 3.2-3.3; Tables 2-3; Figures 1-2; Supplementary Figures S1-S4
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Section 2.7; Supplementary Methods S2
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Section 2.7; Supplementary Methods S2
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	Section 2.7; Supplementary

			Methods S2
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Section 2.7; Supplementary Methods S2
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Section 2.8
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Section 3.1 and Fig. 1
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Section 3.1 and Fig. 1
Study characteristics	17	Cite each included study and present its characteristics.	Section 3.2 and Table 2
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Section 3.5.2, Fig. 4, Supplementary Table S9
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Section 3.2, Table 2, Supplementary Tables S3-S4, Supplementary Figures S1-S4
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Sections 3.2, 3.3 and 3.5; Table 2, Figure 2, Figure 4 and Supplementary Table S9.
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Section 3.3.1 and Figure 2.

	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Section 3.4, Fig.3
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Section 3.6
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Section 3.6
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Section 3.7
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Section 4.1-4.6
	23b	Discuss any limitations of the evidence included in the review.	Section 4.8
	23c	Discuss any limitations of the review processes used.	Section 4.8
	23d	Discuss implications of the results for practice, policy, and future research.	Section 4.7 and 4.9
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	PROSPERO CRD42025632986
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Section 2; the review protocol is available as a peer-reviewed protocol publication: Tang et al. (2026), BMJ Open.
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	Not applicable
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Page 51
Competing interests	26	Declare any competing interests of review authors.	Page 51
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Page 51

From: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71. This work is licensed under CC BY 4.0. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>

Supplementary Methods S1. Data-conversion equations

For studies reporting outcomes by subgroups (e.g., providing separate data for boys and girls) without overall group statistics, subgroup data were combined to obtain the pooled sample size (N_{pool}), pooled mean (M_{pool}), and pooled standard deviation (SD_{pool}):

Sample size:

$$N_{pool} = n_1 + n_2$$

Mean:

$$M_{pool} = \frac{n_1 M_1 + n_2 M_2}{n_1 + n_2}$$

Standard Deviation (SD):

$$SD_{pool} = \sqrt{\frac{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2 + \frac{n_1 n_2}{n_1 + n_2} (M_1 - M_2)^2}{n_1 + n_2 - 1}}$$

Where n , M and SD denote the sample size, mean, and standard deviation for each subgroup, respectively.

For studies reporting means with 95% confidence intervals (CIs) rather than SDs, SDs were derived using a t -distribution rather than the standard normal distribution to account for small sample sizes:

$$SD = \frac{\sqrt{N} \times (CI_{upper} - CI_{lower})}{2 \times t_{df}}$$

Where CI_{upper} and CI_{lower} are the upper and lower limits of the 95% CI, N is the group sample size, and t_{df} is the critical t -value with $N - 1$ degrees of freedom.

Pre-to-post intervention change scores were used to calculate the standardized mean

differences (SMDs). When the mean change (M_{change}) and its standard deviation (SD_{change}) were not reported, they were calculated as follows:

$$M_{change} = M_{post} - M_{pre}$$

$$SD_{change} = \sqrt{SD_{pre}^2 + SD_{post}^2 - (2 \times r \times SD_{pre} \times SD_{post})}$$

Where SD_{pre} and SD_{post} are the standard deviations at baseline and post-intervention, respectively. When unavailable, the pre-post correlation coefficient (r) was set to 0.5.

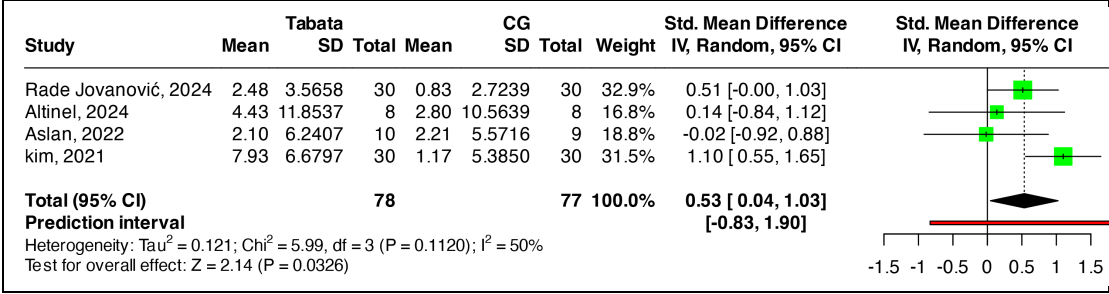
Supplementary Methods S2. Additional statistical analyses

Component	Additional specification	References
Model structure	Two-level models were used for VO ₂ max, BMI, and flexibility because one effect size was included per study. The three-level model for muscular fitness partitioned variance into sampling, within-study, and between-study components. Confidence intervals and hypothesis tests were based on the <i>t</i> -distribution.	[39–43]
Heterogeneity	I ² values of < 25%, 25%–75%, and > 75% were interpreted as low, moderate, and high heterogeneity, respectively. Prediction intervals were calculated for the quantitatively synthesized outcomes.	[44–48]
Subgroup analyses	Analyses for muscular fitness examined gender, population type, intervention duration, Tabata protocol type, age, and muscular-fitness classification. The age cut-off of 15.84 years was based on the median study age and was treated as exploratory. Detailed definitions and coding criteria are provided in Supplementary Table S6.	—
Meta-regression	Analyses examined session frequency, session time, pure Tabata time per session, and total number of sessions. Cluster-robust variance correction was applied to analyses involving dependent effects.	—
Statistical-power assessment	Study-level statistical power for muscular fitness was assessed using a sunset plot.	[49]
Funnel-plot asymmetry	Assessment was restricted to muscular fitness because it was the only outcome with at least 10 study reports. A contour-enhanced funnel plot and multilevel Egger regression with standard error as the moderator were used.	[50–52]
Sensitivity analyses	Leave-one-out analyses were conducted for all quantitatively synthesized outcomes. For muscular fitness, dependent effects were first aggregated at the study level. Additional analyses examined potentially overlapping reports and cohorts for BMI and muscular fitness; report and cohort groupings are provided in Supplementary Table S10.	[33, 38, 42, 53–55]

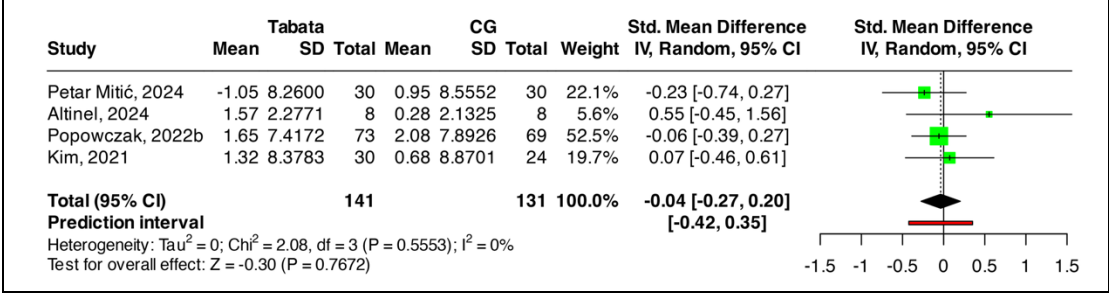
Supplementary Table S2. Search Results

Database	Search Query	Result
Pubmed	((("Adolescent"[Mesh] OR "Child"[Mesh] OR "Students"[Mesh]) OR ("middle school"[Title/Abstract] OR "secondary school"[Title/Abstract] OR "junior high school"[Title/Abstract] OR student*[Title/Abstract] OR child*[Title/Abstract] OR adolescen*[Title/Abstract] OR teenager*[Title/Abstract] OR youth*[Title/Abstract]))) AND ("Tabata"[Title/Abstract] OR "Tabata training"[Title/Abstract] OR "Tabata exercise"[Title/Abstract] OR "Tabata protocol"[Title/Abstract]))	30
Cochrane Library	([mh Adolescent] OR [mh Child] OR [mh Students] OR ("middle school" OR "secondary school" OR "junior high school" OR student* OR child* OR adolescen* OR teenager* OR youth*):ti,ab,kw) AND ("Tabata" OR "Tabata training" OR "Tabata exercise" OR "Tabata protocol"):ti,ab,kw	19
Web of Science	TS=("middle school" OR "secondary school" OR "junior high school" OR student* OR child* OR adolescen* OR teenager* OR youth*) AND TS=("Tabata" OR "Tabata training" OR "Tabata exercise" OR "Tabata protocol")	38
Scopus	TITLE-ABS-KEY("middle school" OR "secondary school" OR "junior high school" OR student* OR child* OR adolescen* OR teenager* OR youth*) AND TITLE-ABS-KEY("Tabata" OR "Tabata training" OR "Tabata exercise" OR "Tabata protocol")	54

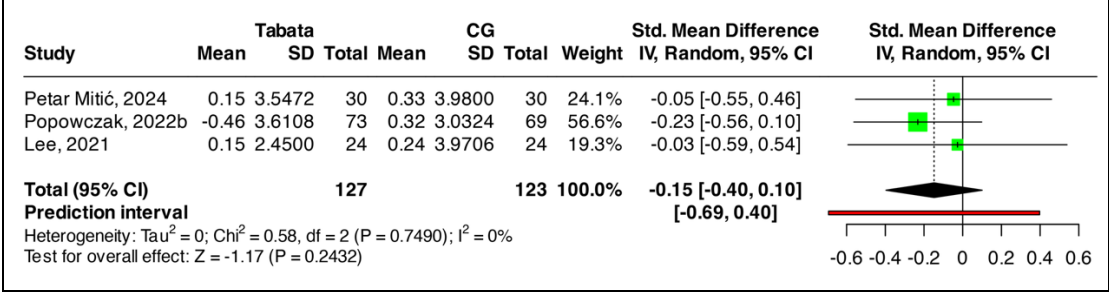
Supplementary Figure S1. Forest plot for VO2max



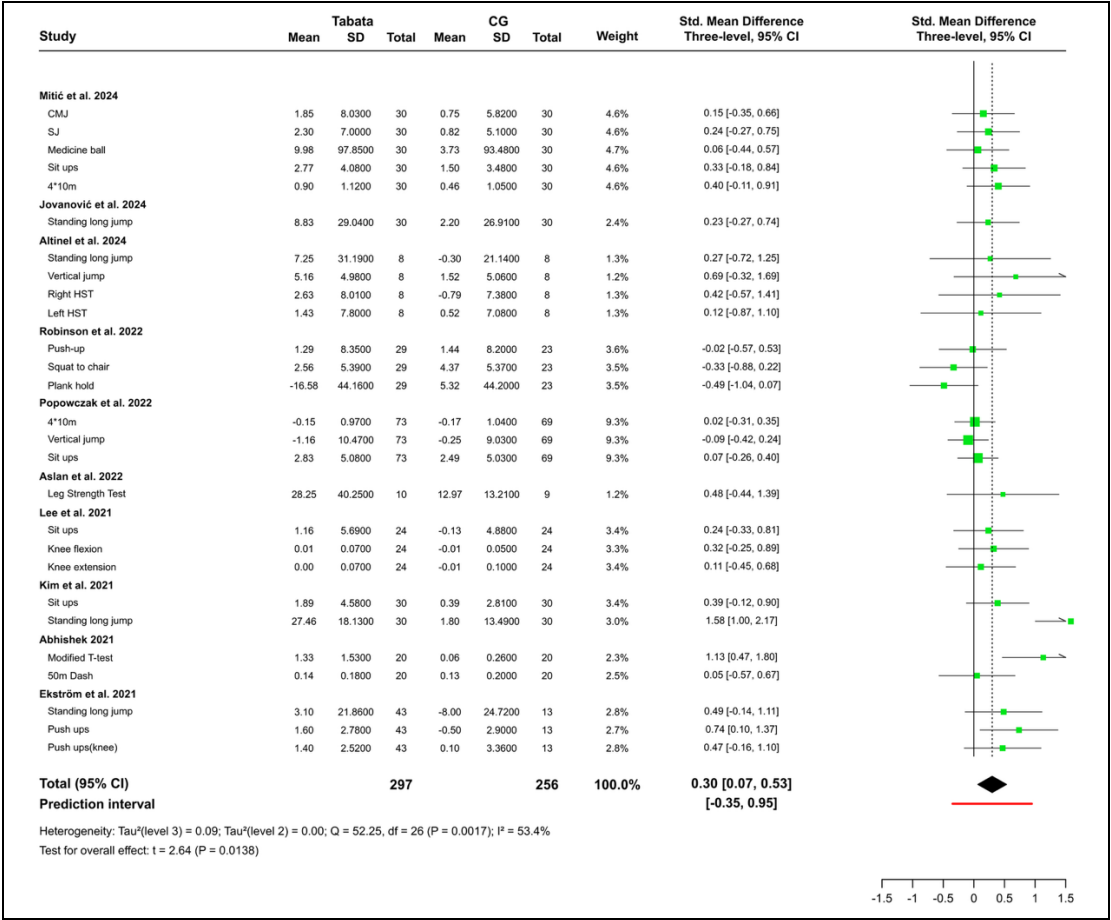
Supplementary Figure S2. Forest plot for flexibility



Supplementary Figure S3. Forest plot for BMI



Supplementary Figure S4. Forest plot for muscular fitness



Supplementary Table S3. Original reported results for psychological, psychosocial, cognitive, and behavioral outcomes

Study	Outcome	Measurement instrument	Original reported result	P value	Effect direction	Statistical significance
Mitić 2025	Stress management	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales, 6-point Likert scale.	Stress management 3.68 ± 0.93 to 3.76 ± 0.91 in Tabata ($p=0.42$); CG 3.70 ± 1.10 to 3.87 ± 0.88 ($p=0.49$).	Tabata $p=0.42$; CG $p=0.49$; between-group not separately reported.	Null	Non-significant
	Revitalization	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales, 6-point Likert scale.	Revitalization 4.07 ± 0.92 to 3.89 ± 0.96 in Tabata ($p=0.57$); CG 3.98 ± 1.01 to 3.84 ± 0.98 ($p=0.28$).	Tabata $p=0.57$; CG $p=0.28$; between-group not separately reported.	Null	Non-significant
	Enjoyment (exercise motivation subscale)	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales, 6-point Likert scale.	Enjoyment 4.04 ± 0.97 to 3.94 ± 0.98 in Tabata ($p=0.80$); CG 4.06 ± 1.12 to 4.13 ± 0.91 ($p=0.83$).	Tabata $p=0.80$; CG $p=0.83$; between-group not separately reported.	Null	Non-significant
	Challenge	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales, 6-point Likert scale.	Challenge 3.92 ± 0.83 to 4.08 ± 0.90 in Tabata ($p=0.12$); CG 4.11 ± 1.12 to 4.22 ± 0.81 ($p=0.76$).	Tabata $p=0.12$; CG $p=0.76$; between-group not separately reported.	Null	Non-significant
	Social recognition	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales, 6-point Likert scale.	A significant time $\times$ group interaction was observed for social recognition ($p = 0.04$; partial eta squared $\eta^2p = 0.079$, medium effect), indicating that Tabata had a superior effect compared to CG.	time $\times$ group $p=0.04$; Tabata within-group $p=0.02$; CG within-group $p=0.92$ in text and 0.97 in Table 2	Positive	Significant
	Affiliation	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales, 6-point Likert scale.	A significant main effect for time was found for affiliation ($F(1,58) = 6.39$, $\eta^2p = 0.05$; $p = 0.02$).	time effect $p=0.02$; Tabata $p=0.02$; CG $p=0.35$; no significant time $\times$ group	Positive	Mixed

		Affiliation showed a significant improvement in the Tabata group (%difference = 10.83%), while no significant changes were noted in the CG.	marker reported		
Competition	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales, 6-point Likert scale.	A significant main effect for time was found for competition ($F(1,58) = 5.78, \eta^2p = 0.04; p = 0.02$). Competition significantly increased in the Tabata group (%difference = 12.14%), whereas no significant differences were found in the CG.	time effect $p=0.02$; Tabata $p=0.02$; CG $p=0.74$; no significant time x group marker reported	Positive	Mixed
Health pressures	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales, 6-point Likert scale.	Health pressures 3.12 ± 0.96 to 3.17 ± 0.99 in Tabata ($p=0.79$); CG 3.19 ± 1.23 to 3.41 ± 1.06 ($p=0.32$).	Tabata $p=0.79$; CG $p=0.32$; between-group not separately reported.	Null	Non-significant
Ill-health avoidance	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales, 6-point Likert scale.	Ill-health avoidance 3.90 ± 1.07 to 3.89 ± 0.97 in Tabata ($p=0.49$); CG 3.90 ± 1.16 to 3.71 ± 1.22 ($p=0.32$).	Tabata $p=0.49$; CG $p=0.32$; between-group not separately reported.	Null	Non-significant
Positive health	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales, 6-point Likert scale.	Positive health 4.23 ± 0.94 to 4.11 ± 0.87 in Tabata ($p=0.55$); CG 4.23 ± 1.21 to 4.28 ± 0.90 ($p=0.88$).	Tabata $p=0.55$; CG $p=0.88$; between-group not separately reported.	Null	Non-significant
Weight management	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales, 6-point Likert scale.	Weight management 3.28 ± 0.80 to 3.42 ± 0.83 in Tabata ($p=0.85$); CG 3.70 ± 1.10 to 3.73 ± 1.04 ($p=0.94$).	Tabata $p=0.85$; CG $p=0.94$; between-group not separately reported.	Null	Non-significant
Appearance	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales,	Appearance 3.53 ± 1.10 to 3.70 ± 0.97 in Tabata ($p=0.19$); CG $3.68 \pm$	Tabata $p=0.19$; CG $p=0.95$; between-group not	Null	Non-significant

		6-point Likert scale.	1.23 to 3.73 ± 1.00 (p=0.95).	separately reported.		
	Strength and endurance motivation	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales, 6-point Likert scale.	Strength and endurance 4.38 ± 0.78 to 4.25 ± 0.72 in TABATA (p=0.20); CG 4.22 ± 1.21 to 4.33 ± 0.71 (p=1.00).	Tabata p=0.20; CG p=1.00; between-group not separately reported.	Null	Non-significant
	Nimbleness	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales, 6-point Likert scale.	Nimbleness 4.00 ± 0.88 to 4.03 ± 0.93 in Tabata (p=0.49); CG 4.02 ± 1.12 to 4.04 ± 0.90 (p=0.97).	Tabata p=0.49; CG p=0.97; between-group not separately reported.	Null	Non-significant
	Physical activity enjoyment	Physical Activity Enjoyment Scale (PACES), 18 items, 7-point bipolar scale; total score 18-126.	As measured by PACES scores, enjoyment levels revealed no significant difference between the Tabata group (95.72 ± 15.41) and the control group (91.73 ± 17.81; p = 0.35).	Tabata p=0.42; CG p=0.49; between-group not separately reported.	Null	Non-significant
Mitić, 2024	Physical activity enjoyment	Physical Activity Enjoyment Scale (PACES), 18 items, 7-point bipolar scale; total score 18-126.	As measured by PACES scores, enjoyment levels revealed no significant difference between the Tabata group (95.72 ± 15.41) and the control group (91.73 ± 17.81; p = 0.35).	p=0.35	Null	Non-significant
	Physical well-being / quality of life	KIDSCREEN-27	The results revealed changes in physical well-being following the Tabata intervention (p = 0.01) compared to the control group (p = 0.42).	Tabata p=0.01; CG p=0.42; between-group p<0.05 per Table 2 footnote	Positive	Significant
	Psychological well-	KIDSCREEN-27	Psychological Well-Being showed no	Tabata p=0.14; CG p=0.93.	Null	Non-significant

	being / quality of life		statistically significant changes ($p > 0.05$) within either the Tabata or control groups from pre-test to post-test.			
	Autonomy and parents / quality of life	KIDSCREEN-27	Autonomy and Parents showed no statistically significant changes ($p > 0.05$) within either the Tabata or control groups from pre-test to post-test.	Tabata $p=0.71$; CG $p=0.80$.	Null	Non-significant
	Peers and social support / quality of life	KIDSCREEN-27	Peers and Social Support showed no statistically significant changes ($p > 0.05$) within either the Tabata or control groups from pre-test to post-test.	Tabata $p=0.84$; CG $p=0.83$.	Null	Non-significant
	School environment / quality of life	KIDSCREEN-27	School Environment showed no statistically significant changes ($p > 0.05$) within either the Tabata or control groups from pre-test to post-test.	Tabata $p=0.45$; CG $p=0.44$.	Null	Non-significant
Kim, 2021	Stage of physical activity / intention to participate in regular physical activity	Stage of change scale for physical activity, Korean version.	Experimental group shifted from PC/CO/PR at baseline to 22 action and 1 maintenance at week 12; control group mostly remained PC/CO/PR.	$p < 0.001$	Positive	Significant
	Perceived benefits toward physical	Decisional Balance Scale for Exercise, Korean version.	EG: 2.90 ± 0.32 baseline, 3.97 ± 0.46 week 12; CG: 2.74 ± 0.58 baseline,	$p < 0.001$	Positive	Significant

	activity (pros)		3.00 ± 0.37 week 12.			
	Perceived barriers toward physical activity (cons)	Decisional Balance Scale for Exercise, Korean version.	EG: 2.48 ± 0.56 baseline, 1.83 ± 0.31 week 12; CG: 2.40 ± 0.82 baseline, 2.54 ± 0.41 week 12.	time F(1,58)=3.77, p<0.05; group F(1,58)=1.33, not marked significant	Positive	Mixed
	Exercise self-efficacy / confidence for doing physical activity	Exercise self-efficacy questionnaire, Korean version; 5-point Likert scale.	EG: 2.87 ± 0.56 baseline, 3.76 ± 0.24 week 12; CG: 2.56 ± 0.72 baseline, 2.23 ± 0.37 week 12.	p < 0.001	Positive	Significant
	Cognitive processes of change for physical activity	Processes of Change Scale, Korean version; cognitive processes subscale.	EG: 2.88 ± 0.71 baseline, 3.60 ± 0.52 week 12; CG: 2.81 ± 0.90 baseline, 2.87 ± 0.41 week 12.	p < 0.01	Positive	Significant
	Behavioral processes of change for physical activity	Processes of Change Scale, Korean version; behavioral processes subscale.	EG: 2.68 ± 0.65 baseline, 3.22 ± 0.35 week 12; CG: 2.74 ± 0.82 baseline, 2.70 ± 0.60 week 12.	group p < 0.001; time p < 0.01	Positive	Significant
Broad,2023	On-task behavior	Systematic time-sampling classroom behavior observation using a custom smartphone app; behavior definitions from BOSS.	When on-task behavior was averaged across both observation periods, the AM, PM, and BOTH conditions were greater (p < .001, f = 0.778, large) than CTRL.	p < .001	Positive	Significant
	Off-task motor behavior	Systematic time-sampling classroom behavior observation using a custom smartphone app; behavior definitions from BOSS.	There was a condition x time interaction (p < .001, f = 0.453, medium) for off-task motor behavior.	p < .001	Mixed	Significant
	Off-task passive	Systematic time-sampling	There was a condition x time	p = .004	Mixed	Significant

	behavior	classroom behavior observation using a custom smartphone app; behavior definitions from BOSS.	interaction ($p = .004$, $f = 0.380$, medium) for off-task passive behavior.			
	Off-task verbal behavior	Systematic time-sampling classroom behavior observation using a custom smartphone app; behavior definitions from BOSS.	There was no condition x time interaction ($p = .432$, $f = 0.167$, small); however, there was a main effect of condition ($p < .001$, $f = 0.671$, large) where the AM and BOTH conditions were lower ($p < .001$, $d > 0.82$, large) than CTRL and PM.	$p = .432$ interaction; $p < .001$ condition main effect	Mixed	Mixed
Robinson, 2022	Inhibition	NIH Toolbox Flanker Inhibitory Control and Attention Test, Ages 12+	No significant group $\times$ time effect was observed for inhibition, and none of the adjusted between-group comparisons was statistically significant.	All adjusted between-group comparisons $p > 0.05$.	Null	Non-significant
	Cognitive flexibility	NIH Toolbox Dimensional Change Card Sort Test, Ages 12+	No significant group $\times$ time effect was observed for cognitive flexibility, and none of the adjusted between-group comparisons was statistically significant.	All adjusted between-group comparisons $p > 0.05$.	Null	Non-significant
	Episodic memory	NIH Toolbox Picture Sequence Memory Test	Episodic memory improved significantly in the resistance-training-without-cognitive-training group (RTNC) compared with the	CON-RTNC: $p = 0.014$; all remaining comparisons $p > 0.05$.	Mixed	Significant for CON-RTNC only

		control group (CON). The adjusted CON-RTNC difference was -9.87 units (95% CI -17.71 to -2.03; d=0.72). The remaining pairwise comparisons were non-significant.			
On-task behavior	Momentary time-sampling observations adapted from Behavior Observation of Students in Schools and Applied Behavior Analysis for Teachers.	Significant group-by-time effects were observed for participants' on-task behavior between the CON and SECT groups (mean change = 28, 95% CI: 0.48 to 56.03, p = 0.046, d = 0.84), between CON and RTCT (mean change = -45.26, 95% CI: -71.50 to -19.03, p = 0.001, d = 1.23), between SECT and RTNC (mean change = -36.44, 95% CI: -65.26 to -7.61, p = 0.015, d = 1.48), between SECT and RTCT (mean change = -73.51, 95% CI: -101.29 to -45.74, p = 0.001, d = 2.52), and RTNC and RTCT (mean change = -37.09, 95% CI: -64.42 to -9.75, p = 0.009, d = 1.34).	p=0.046; p=0.548 for CON-RTNC in Table 3; p=0.001; p=0.015; p<0.001 in Table 3; p=0.009	Mixed	Mixed
Off-task behavior	Momentary time-sampling observations adapted from Behavior Observation of Students in Schools and Applied	Table 3 reported adjusted differences for off-task behavior: CON-SECT - 30.18 (p=0.034), CON-RTNC 8.17 (p=0.548), CON-RTCT 45.26	p=0.034; p=0.548; p=0.001; p=0.010; p<0.001; p=0.009	Mixed	Mixed

Behavior Analysis for Teachers. (p=0.001), SECT-RTNC 38.35
(p=0.010), SECT-RTCT 75.44
(p<0.001), and RTNC-RTCT 37.09
(p=0.009).

Note. Tabata refers to the Tabata or Tabata-style exercise group unless otherwise specified. CG, control group; EG, experimental group; EMI-2, Exercise Motivation Inventory-2; PACES, Physical Activity Enjoyment Scale; TTM, Transtheoretical Model; BOSS, Behavior Observation of Students in Schools; CAB, classroom activity break; AM, classroom activity break performed in the morning; PM, classroom activity break performed in the afternoon; BOTH, classroom activity breaks performed in both the morning and afternoon; CTRL, no activity-break control condition; CON, sedentary control condition without cognitive training; SECT, sedentary condition with cognitive training; RTNC, resistance training without cognitive training; RTCT, resistance training with cognitive training; NIH, National Institutes of Health; CI, confidence interval; PC, precontemplation; CO, contemplation; PR, preparation; AC, action; MA, maintenance. Positive indicates a favorable effect of Tabata or Tabata-style exercise; Null indicates no significant or meaningful effect; Mixed indicates inconsistent findings across subscales, time points, or comparisons.

Supplementary Table S4. Quantitative data and between-group/within-group results for psychological, psychosocial, cognitive, and behavioral outcomes

Study	Outcome	Measurement instrument	Quantitative data reported	Between-group result	Within-group result
Mitić 2025	Stress management	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales, 6-point Likert scale.	Tabata: 3.68 ± 0.93 pre, 3.76 ± 0.91 post; CG: 3.70 ± 1.10 pre, 3.87 ± 0.88 post.	No significant differences were found within or between groups ($p > 0.05$).	HIIT $p=0.42$; CG $p=0.49$; between-group not separately reported.
	Revitalization	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales, 6-point Likert scale.	Tabata: 4.07 ± 0.92 pre, 3.89 ± 0.96 post; CG: 3.98 ± 1.01 pre, 3.84 ± 0.98 post.	No significant differences were found within or between groups ($p > 0.05$).	HIIT $p=0.57$; CG $p=0.28$; between-group not separately reported.
	Enjoyment (exercise motivation subscale)	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales, 6-point Likert scale.	Tabata: 4.04 ± 0.97 pre, 3.94 ± 0.98 post; CG: 4.06 ± 1.12 pre, 4.13 ± 0.91 post.	No significant differences were found within or between groups ($p > 0.05$).	HIIT $p=0.80$; CG $p=0.83$; between-group not separately reported.
	Challenge	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales, 6-point Likert scale.	Tabata: 3.92 ± 0.83 pre, 4.08 ± 0.90 post; CG: 4.11 ± 1.12 pre, 4.22 ± 0.81 post.	No significant differences were found within or between groups ($p > 0.05$).	HIIT $p=0.12$; CG $p=0.76$; between-group not separately reported.
	Social recognition	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales, 6-point Likert scale.	Tabata: 3.17 ± 1.08 pre, 3.57 ± 0.89 post, $p=0.02$; CG: 3.53 ± 1.19 pre, 3.43 ± 1.07 post, $p=0.97$.	Significant time x group interaction, $p=0.04$; partial eta squared $\eta^2 p=0.079$.	HIIT increased by 11.86%, $p=0.02$; CG change 2.87%, $p=0.92/0.97$ as reported.
	Affiliation	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales, 6-point Likert scale.	Tabata: 3.58 ± 1.17 pre, 3.99 ± 0.80 post, $p=0.02$; CG: 3.62 ± 1.14 pre, 3.75 ± 1.06 post, $p=0.35$.	No significant time $\times$ group interaction marker reported for this subscale; significant time effect only.	HIIT improved, $p=0.02$; CG non-significant, $p=0.35$.
	Competition	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales,	Tabata: 3.48 ± 1.19 pre, 3.93 ± 0.87 post, $p=0.02$; CG: 3.72 ± 1.18 pre,	No significant time $\times$ group interaction marker reported	HIIT improved, $p=0.02$; CG non-significant, $p=0.74$.

		6-point Likert scale.	3.77 ± 1.02 post, $p=0.74$.	for this subscale; significant time effect only.	
	Health pressures	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales, 6-point Likert scale.	Tabata: 3.12 ± 0.96 pre, 3.17 ± 0.99 post; CG: 3.19 ± 1.23 pre, 3.41 ± 1.06 post.	No significant differences were found within or between groups ($p > 0.05$).	HIIT $p=0.79$; CG $p=0.32$; between-group not separately reported.
	Ill-health avoidance	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales, 6-point Likert scale.	Tabata: 3.90 ± 1.07 pre, 3.89 ± 0.97 post; CG: 3.90 ± 1.16 pre, 3.71 ± 1.22 post.	No significant differences were found within or between groups ($p > 0.05$).	HIIT $p=0.49$; CG $p=0.32$; between-group not separately reported.
	Positive health	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales, 6-point Likert scale.	Tabata: 4.23 ± 0.94 pre, 4.11 ± 0.87 post; CG: 4.23 ± 1.21 pre, 4.28 ± 0.90 post.	No significant differences were found within or between groups ($p > 0.05$).	HIIT $p=0.55$; CG $p=0.88$; between-group not separately reported.
	Weight management	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales, 6-point Likert scale.	Tabata: 3.28 ± 0.80 pre, 3.42 ± 0.83 post; CG: 3.70 ± 1.10 pre, 3.73 ± 1.04 post.	No significant differences were found within or between groups ($p > 0.05$).	HIIT $p=0.85$; CG $p=0.94$; between-group not separately reported.
	Appearance	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales, 6-point Likert scale.	Tabata: 3.53 ± 1.10 pre, 3.70 ± 0.97 post; CG: 3.68 ± 1.23 pre, 3.73 ± 1.00 post.	No significant differences were found within or between groups ($p > 0.05$).	HIIT $p=0.19$; CG $p=0.95$; between-group not separately reported.
	Strength and endurance motivation	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales, 6-point Likert scale.	Tabata: 4.38 ± 0.78 pre, 4.25 ± 0.72 post; CG: 4.22 ± 1.21 pre, 4.33 ± 0.71 post.	No significant differences were found within or between groups ($p > 0.05$).	HIIT $p=0.20$; CG $p=1.00$; between-group not separately reported.
	Nimbleness	Exercise Motivation Inventory-2 (EMI-2), 51 items, 14 subscales, 6-point Likert scale.	Tabata: 4.00 ± 0.88 pre, 4.03 ± 0.93 post; CG: 4.02 ± 1.12 pre, 4.04 ± 0.90 post.	No significant differences were found within or between groups ($p > 0.05$).	HIIT $p=0.49$; CG $p=0.97$; between-group not separately reported.
Mitić, 2024	Physical activity enjoyment	Physical Activity Enjoyment Scale (PACES), 18 items, 7-	HIIT: 95.72 ± 15.41 ; CG: 91.73 ± 17.81 .	Between-group $p=0.35$.	Not applicable/not reported as pre-post change.

		point bipolar scale; total score 18-126.			
	Physical well-being / quality of life	KIDSCREEN-27	HIIT: 18.23 ± 3.40 pre, 19.17 ± 2.94 post; CG: 19.53 ± 2.76 pre, 19.87 ± 2.34 post.	Significant difference between HIIT and CG, p < 0.05.	HIIT p=0.01; CG p=0.42.
	Psychological well-being / quality of life	KIDSCREEN-27	HIIT: 26.10 ± 5.38 pre, 26.97 ± 5.99 post, p=0.14; CG: 27.73 ± 4.41 pre, 27.77 ± 3.96 post, p=0.93.	No significant between-group difference reported/marked.	HIIT p=0.14; CG p=0.93.
	Autonomy and parents / quality of life	KIDSCREEN-27	HIIT: 30.87 ± 3.39 pre, 31.10 ± 3.85 post, p=0.71; CG: 30.53 ± 2.80 pre, 30.47 ± 3.50 post, p=0.80.	No significant between-group difference reported/marked.	HIIT p=0.71; CG p=0.80.
	Peers and social support / quality of life	KIDSCREEN-27	HIIT: 17.90 ± 3.17 pre, 17.93 ± 3.26 post, p=0.84; CG: 17.37 ± 2.34 pre, 17.13 ± 2.80 post, p=0.83.	No significant between-group difference reported/marked.	HIIT p=0.84; CG p=0.83.
	School environment / quality of life	KIDSCREEN-27	HIIT: 12.77 ± 3.51 pre, 12.50 ± 3.52 post, p=0.45; CG: 14.17 ± 2.63 pre, 13.77 ± 3.64 post, p=0.44.	No significant between-group difference reported/marked.	HIIT p=0.45; CG p=0.44.
Kim, 2021	Stage of physical activity / intention to participate in regular physical activity	Stage of change scale for physical activity, Korean version.	Experimental group shifted from PC/CO/PR at baseline to 22 action and 1 maintenance at week 12; control group mostly remained PC/CO/PR.	Stage distribution significantly changed over 12 weeks; between-group baseline difference not reported as significant.	Experimental group stage significantly increased; control group had no statistical significance.
	Perceived benefits toward physical activity (pros)	Decisional Balance Scale for Exercise, Korean version.	EG: 2.90 ± 0.32 baseline, 3.97 ± 0.46 week 12; CG: 2.74 ± 0.58 baseline, 3.00 ± 0.37 week 12.	Group F(1,58)=63.14, p<0.001.	Time F(1,58)=51.47, p<0.001.

	Perceived barriers toward physical activity (cons)	Decisional Balance Scale for Exercise, Korean version.	EG: 2.48 ± 0.56 baseline, 1.83 ± 0.31 week 12; CG: 2.40 ± 0.82 baseline, 2.54 ± 0.41 week 12.	Group $F(1,58)=1.33$, not significant.	Time $F(1,58)=3.77$, $p<0.05$; EG decreased from 2.48 to 1.83.
	Exercise self-efficacy / confidence for doing physical activity	Exercise self-efficacy questionnaire, Korean version; 5-point Likert scale.	EG: 2.87 ± 0.56 baseline, 3.76 ± 0.24 week 12; CG: 2.56 ± 0.72 baseline, 2.23 ± 0.37 week 12.	Group $F(1,58)=79.45$, $p<0.001$.	Time $F(1,58)=49.31$, $p<0.001$.
	Cognitive processes of change for physical activity	Processes of Change Scale, Korean version; cognitive processes subscale.	EG: 2.88 ± 0.71 baseline, 3.60 ± 0.52 week 12; CG: 2.81 ± 0.90 baseline, 2.87 ± 0.41 week 12.	Group $F(1,58)=17.17$, $p<0.01$.	Time $F(1,58)=16.60$, $p<0.01$.
	Behavioral processes of change for physical activity	Processes of Change Scale, Korean version; behavioral processes subscale.	EG: 2.68 ± 0.65 baseline, 3.22 ± 0.35 week 12; CG: 2.74 ± 0.82 baseline, 2.70 ± 0.60 week 12.	Group $F(1,58)=29.97$, $p<0.001$.	Time $F(1,58)=12.32$, $p<0.01$.
Broad,2023	On-task behavior	Systematic time-sampling classroom behavior observation using a custom smartphone app; behavior definitions from BOSS.	Compared to control, CAB improved overall on-task behavior by 10.4%, 10.5%, and 14% in AM, PM, and BOTH.	AM, PM, BOTH > CTRL, $p<.001$; condition x time interaction $p<.001$.	Not a longitudinal within-group pre-post design; condition comparisons across experimental days.
	Off-task motor behavior	Systematic time-sampling classroom behavior observation using a custom smartphone app; behavior definitions from BOSS.	Morning: AM and BOTH lower than CTRL and PM ($p<.014$); afternoon: PM lower than CTRL and AM ($p<.024$), BOTH lower than AM ($p=.036$).	Condition x time interaction $p<.001$.	Not a longitudinal within-group pre-post design; condition comparisons across experimental days.
	Off-task passive behavior	Systematic time-sampling classroom behavior observation	Morning: AM lower than CTRL ($p=.003$); afternoon: AM, PM, BOTH	Condition x time interaction $p=.004$.	Not a longitudinal within-group pre-post design; condition comparisons

		using a custom smartphone app; behavior definitions from BOSS.	lower than CTRL (p<.028), PM and BOTH lower than AM (p<.014).		across experimental days.
	Off-task verbal behavior	Systematic time-sampling classroom behavior observation using a custom smartphone app; behavior definitions from BOSS.	Condition x time p=.432; main effect condition p<.001; AM/BOTH lower than CTRL/PM p<.001.	No condition x time interaction; significant main effect of condition.	Not a longitudinal within-group pre-post design; condition comparisons across experimental days.
Robinson, 2022	Inhibition	NIH Toolbox Flanker Inhibitory Control and Attention Test, Ages 12+	CON: 103.18 (99.73, 106.63) to 104.15 (100.58, 107.72); SECT: 106.55 (102.98, 110.11) to 109.44 (105.73, 113.15); RTNC: 105.79 (102.74, 108.85) to 108.07 (105.06, 111.07); RTCT: 107.90 (104.28, 111.52) to 112.00 (108.25, 115.66). Values are reported as mean (95% CI).	No significant group × time effect; all adjusted between-group comparisons were non-significant (all p > 0.05).	Not separately reported.
	Cognitive flexibility	NIH Toolbox Dimensional Change Card Sort Test, Ages 12+	CON: 103.91 (98.85, 108.97) to 103.65 (98.46, 108.84); SECT: 104.09 (98.87, 109.31) to 106.21 (100.68, 111.73); RTNC: 111.36 (106.90, 115.82) to 110.21 (105.80, 114.61); RTCT: 111.20 (105.90, 116.50) to 113.99 (108.59, 119.39). Values are reported as mean (95% CI).	No significant group × time effect; all adjusted between-group comparisons were non-significant (all p > 0.05).	Not separately reported.

Episodic memory	NIH Toolbox Picture Sequence Memory Test	CON: 106.41 (100.67, 112.15) to 99.41 (93.46, 105.35); SECT: 108.93 (102.95, 114.91) to 105.13 (98.77, 111.49); RTNC: 108.82 (103.69, 113.95) to 111.69 (106.69, 116.69); RTCT: 114.35 (108.33, 120.37) to 113.94 (107.68, 120.20). Values are reported as mean (95% CI).	RTNC improved significantly compared with CON: adjusted CON–RTNC difference –9.87 units (95% CI –17.71 to –2.03; p=0.014; d=0.72). All remaining pairwise comparisons were non-significant.	Not separately reported.
On-task behavior	Momentary time-sampling observations adapted from Behavior Observation of Students in Schools and Applied Behavior Analysis for Teachers.	Adjusted differences: CON-SECT 28.25 (95% CI 0.48 to 56.03), p=0.046, d=0.84; CON-RTNC -8.18, p=0.548; CON-RTCT -45.26, p=0.001, d=1.23; SECT-RTNC -36.44, p=0.015, d=1.48; SECT-RTCT -73.51, p<0.001, d=2.52; RTNC-RTCT -37.09, p=0.009, d=1.34.	Mixed pairwise results; RTCT showed favorable significant effects versus control/sedentary and RTNC, while RTNC was not significantly different from CON.	RTCT increased from 50.42 to 79.97; RTNC 65.15 to 57.61; CON 61.67 to 45.95; SECT 63.63 to 19.66.
Off-task behavior	Momentary time-sampling observations adapted from Behavior Observation of Students in Schools and Applied Behavior Analysis for Teachers.	Adjusted differences: CON-SECT -30.18, p=0.034, d=0.84; CON-RTNC 8.17, p=0.548; CON-RTCT 45.26, p=0.001, d=1.23; SECT-RTNC 38.35, p=0.010, d=1.48; SECT-RTCT 75.44, p<0.001, d=2.53; RTNC-RTCT 37.09, p=0.009, d=1.34.	Mixed pairwise results; RTCT showed favorable significant effects for reduced off-task behavior, while RTNC was not significantly different from CON.	RTCT decreased from 49.58 to 20.04; RTNC 34.83 to 42.38; CON 38.33 to 54.05; SECT 36.37 to 80.34.

Note. Tabata refers to the Tabata or Tabata-style exercise group unless otherwise specified. CG, control group; EG, experimental group; HIIT, high-intensity interval training; EMI-2,

Exercise Motivation Inventory-2; PACES, Physical Activity Enjoyment Scale; TTM, Transtheoretical Model; BOSS, Behavior Observation of Students in Schools; CAB, classroom activity break; AM, classroom activity break performed in the morning; PM, classroom activity break performed in the afternoon; BOTH, classroom activity breaks performed in both the morning and afternoon; CTRL, no activity-break control condition; CON, sedentary control condition without cognitive training; SECT, sedentary condition with cognitive training; RTNC, resistance training without cognitive training; RTCT, resistance training with cognitive training; NIH, National Institutes of Health; CI, confidence interval; PC, precontemplation; CO, contemplation; PR, preparation; AC, action; MA, maintenance. Between-group results refer to comparisons between intervention and comparator groups or conditions; within-group results refer to changes from baseline or pre-test to post-test within the same group.

Supplementary Table S5. Operational criteria used for the GRADE certainty assessment

Domain	Operational criteria
Risk of bias	Judged outcome-specifically according to the distribution and likely influence of RoB 2 judgements among studies contributing to each outcome. Downgrading was applied when studies at high risk of bias or outcome-relevant methodological concerns were considered likely to affect the pooled estimate.
Inconsistency	Downgraded by one level when statistical heterogeneity was moderate ($I^2 = 25\%–75\%$) and by two levels when heterogeneity was high ($I^2 > 75\%$).
Indirectness	Downgraded by one or two levels when important differences existed between the included populations, interventions, comparators, or outcomes and those predefined in the PICOS criteria.
Imprecision	Judged according to confidence-interval width, whether confidence intervals crossed the null or included clinically uncertain effects, the number of contributing studies and participants, and information size where available [57].
Publication bias	Downgraded by one level if Egger's test indicated significant funnel-plot asymmetry ($p < 0.05$), if severe publication bias was suspected from funnel plots for outcomes with sufficient data, or if publication bias and small-study effects could not be formally assessed because fewer than 10 study reports or effect-size clusters were available.
Note. GRADE, Grading of Recommendations Assessment, Development, and Evaluation; RoB 2, Risk of Bias 2; PICOS, population, intervention, comparator, outcomes, and study design; I^2 , heterogeneity statistic.	

Supplementary Table S6. Full subgroup analyses for muscular fitness

Moderator	Subgroup	K	Studies	Tabata	CG	Hedges' g (95% CI)	p value	I ²	Moderator p
Gender	Male	6	2	60	60	0.24 (-0.18, 0.65)	0.256	0.0%	0.082
	Female	4	2	50	50	0.75 (0.30, 1.20)	0.002	81.3%	0.082
	Mixed	17	6	187	146	0.16 (-0.09, 0.41)	0.195	47.8%	0.082
Age subgroup	Younger subgroup	12	5	111	80	0.61 (0.36, 0.86)	<0.001	45.0%	0.002
	Older subgroup	15	5	186	176	0.07 (-0.12, 0.26)	0.439	30.2%	0.002
Population type	General Population	22	8	279	239	0.29 (0.03, 0.54)	0.032	61.7%	0.735
	Athletes	5	2	18	17	0.40 (-0.26, 1.07)	0.225	0.0%	0.735
Intervention duration	2 weeks	2	1	20	20	0.55 (-0.10, 1.21)	0.094	81.6%	0.121
	4 weeks	3	1	29	23	-0.28 (-0.83, 0.28)	0.314	0.0%	0.121
	6 weeks	8	3	61	30	0.48 (0.07, 0.89)	0.024	0.0%	0.121
	10 weeks	6	2	97	93	0.09 (-0.27, 0.46)	0.599	14.1%	0.121
	12 weeks	8	3	90	90	0.45 (0.11, 0.79)	0.011	71.0%	0.121
Tabata protocol type	Classic Tabata	8	3	92	56	0.26 (-0.20, 0.71)	0.254	71.3%	0.837
	Modified Tabata	18	6	195	191	0.31 (-0.00, 0.63)	0.052	50.7%	0.837
Physiological component	Power	10	6	214	180	0.37 (0.06, 0.67)	0.020	77.6%	0.713
	Strength	5	3	42	41	0.25 (-0.23, 0.74)	0.290	0.0%	0.713
	Endurance	8	6	229	189	0.20 (-0.12, 0.51)	0.209	39.9%	0.713
	Speed	4	3	123	119	0.39 (-0.01, 0.79)	0.055	70.0%	0.713
Anatomical/functional category	Lower-Body Muscular Fitness	12	9	277	236	0.32 (0.04, 0.60)	0.028	71.4%	0.827
	Upper-Body Muscular Fitness	6	4	110	74	0.31 (-0.09, 0.70)	0.119	21.1%	0.827
	Core Endurance	5	5	186	176	0.19 (-0.15, 0.54)	0.257	35.7%	0.827
	Speed and Agility	4	3	123	119	0.37 (-0.03, 0.77)	0.067	70.0%	0.827

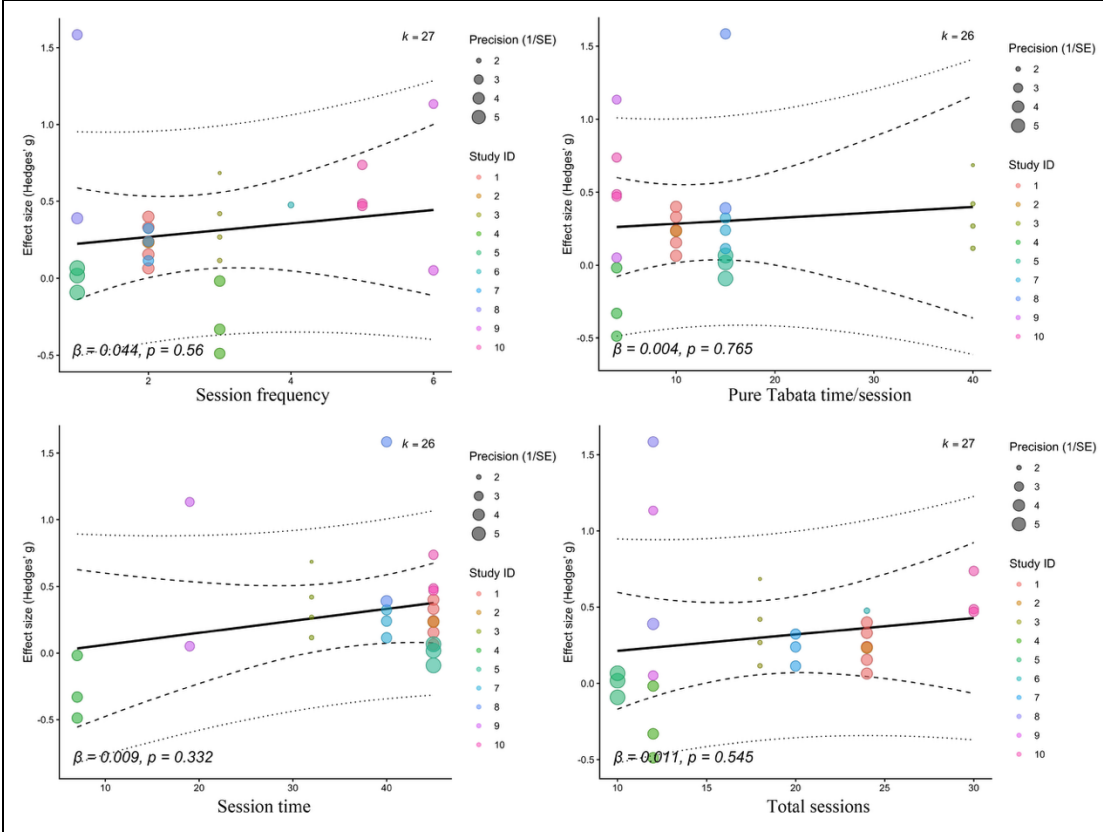
Note: K indicates the number of effect sizes; Studies indicates the number of independent studies. Participant numbers were counted once per study within each subgroup to avoid double-counting.

Supplementary Table S7. Full meta-regression analyses for muscular fitness

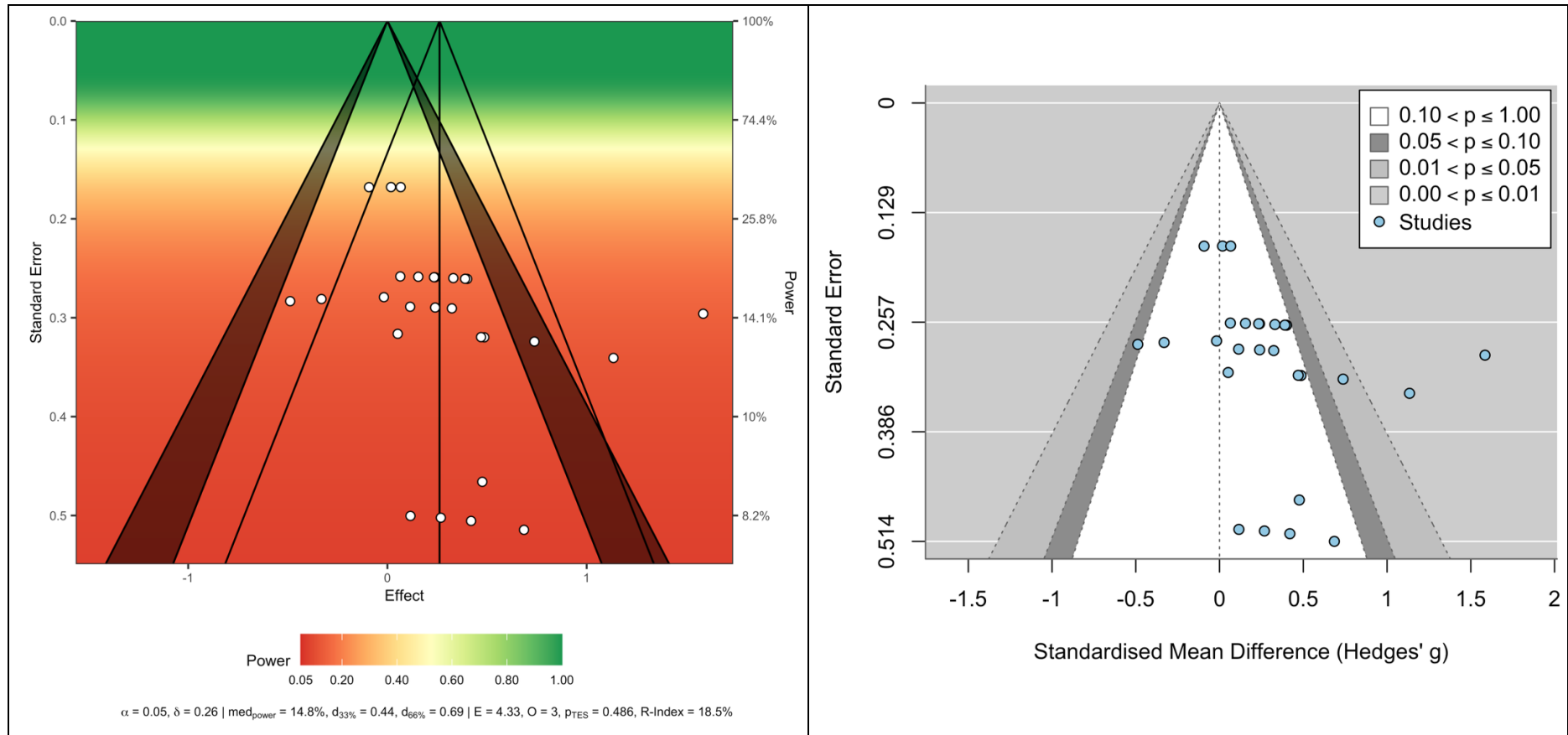
Predictor	K	Studies	β (95% CI)	p value	Residual I ²
Session frequency	27	10	0.044 (-0.109, 0.197)	0.560	55.7%
Session time	26	9	0.009 (-0.010, 0.028)	0.332	55.4%
Pure Tabata time/session	26	9	0.004 (-0.022, 0.030)	0.765	58.6%
Total sessions	27	10	0.011 (-0.025, 0.047)	0.545	55.3%

Pure Tabata time/session refers to the time specifically devoted to Tabata-style exercise within each intervention session. Total sessions were calculated as session frequency multiplied by intervention duration in weeks. CG, control group; CI, confidence interval; I², heterogeneity statistic; β , meta-regression coefficient. p value indicates the test of whether the meta-regression coefficient differs from zero.

Supplementary Figure S5. Full meta-regression plots for muscular fitness

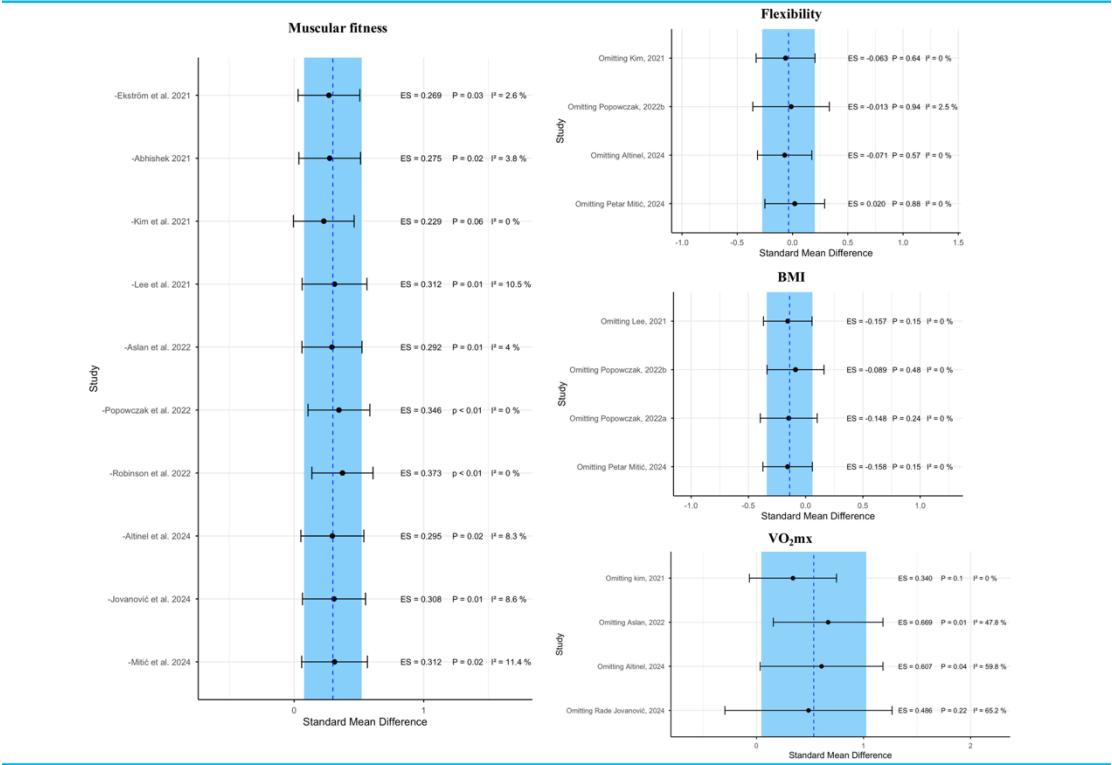


Supplementary Figure S6. Funnel plot and sunset plot for muscular fitness



Supplementary Figure S7. Leave-one-out sensitivity analyses

A Sensitivity Analysis based on Leave-one-out



Supplementary Table S8. Methodological quality assessment [PEDro]

Author, year	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	Total
Petar Mitić 2025	Y	1	0	1	0	0	0	1	1	1	1	6
Petar Mitić 2024	Y	1	0	1	0	0	0	1	1	1	1	6
Rade Jovanović 2024	Y	1	0	1	0	0	0	1	1	1	1	6
Altinel 2024	Y	1	0	1	0	0	0	0	0	1	1	4
Broad 2023	Y	1	0	1	0	0	0	1	1	1	1	6
Robinson 2022	Y	1	0	1	0	0	0	0	1	1	1	5
Popowczak 2022a	Y	1	0	1	0	0	0	0	0	1	1	4
Popowczak 2022b	Y	1	0	1	0	0	0	0	0	1	1	4
Domaradzki 2022	Y	1	0	1	0	0	0	1	0	0	0	3
Aslan 2022	Y	1	0	1	0	0	0	1	1	1	1	6
Lee 2021	Y	1	0	1	0	0	0	1	1	1	1	6
Kim 2021	Y	1	0	1	0	0	0	1	1	1	1	6
Abhishek Abhinandan 2021	Y	1	0	1	0	0	0	1	1	1	1	6
Ekström 2019	Y	1	0	0	0	0	0	0	0	1	1	3

Note: studies scoring ≥ 6 are considered **high quality**, those scoring 4-5 are considered **moderate quality**, and those scoring ≤ 3 are considered **low quality**.

1. eligibility criteria were specified (not included in the total score)
2. subjects were randomly allocated to groups (in a crossover study, subjects were randomly allocated an order in which treatments were received)
3. allocation was concealed
4. the groups were similar at baseline regarding the most important prognostic indicators
5. there was blinding of all subjects

6. there was blinding of all therapists who administered the therapy
7. there was blinding of all assessors who measured at least one key outcome
8. measures of at least one key outcome were obtained from more than 85% of the subjects initially allocated to groups
9. all subjects for whom outcome measures were available received the treatment or control condition as allocated or, where this was not the case, data for at least one key outcome was analysed by “intention to treat”
10. the results of between-group statistical comparisons are reported for at least one key outcome
11. the study provides both point measures and measures of variability for at least one key outcome

Supplementary Table S9. Risk-of-bias domain-level judgements

Study	D1	D2	D3	D4	D5	Overall
Petar Mitić 2025	Some concerns	Some concerns	Low	Some concerns	Low	Some concerns
Petar Mitić 2024	Some concerns	Some concerns	Low	Some concerns	Low	Some concerns
Rade Jovanović 2024	Some concerns	Some concerns	Low	Low	Low	Some concerns
Altinel 2024	Some concerns	Some concerns	Some concerns	Low	Some concerns	Some concerns
Broad 2023	Some concerns	Some concerns	Low	Some concerns	Low	Some concerns
Robinson 2022	Some concerns	Some concerns	Low	Some concerns	Low	Some concerns
Popowczak 2022a	Some concerns	Some concerns	Some concerns	Low	Low	Some concerns
Popowczak 2022b	Some concerns	Some concerns	High	Low	High	High
Domaradzki 2022	Some concerns	Some concerns	High	Low	Some concerns	High
Aslan 2022	Some concerns	Some concerns	Low	Low	Low	Some concerns
Lee 2021	Some concerns	Some concerns	Low	Low	Low	Some concerns
Kim 2021	Some concerns	Some concerns	Low	Some concerns	Low	Some concerns
Abhishek Abhinandan 2021	Some concerns	Some concerns	Low	Low	Some concerns	Some concerns
Ekström 2019	Some concerns	Some concerns	High	Low	Low	High

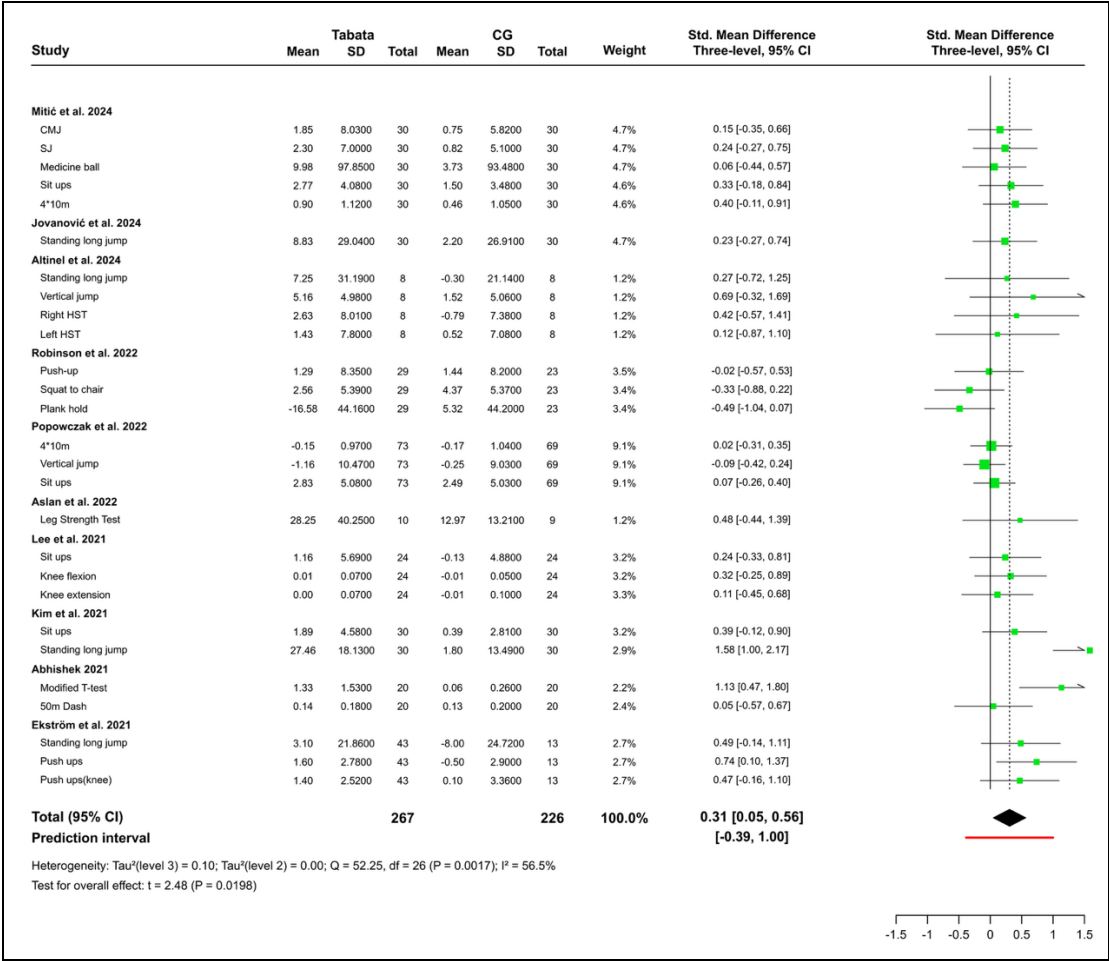
Note. D1, bias arising from the randomization process; D2, bias due to deviations from intended interventions; D3, bias due to missing outcome data; D4, bias in measurement of the outcome; D5, bias in selection of the reported result. Judgements were made using the RoB 2 domains and categorized as low risk, some concerns, or high risk.

Supplementary Table S10. Sensitivity analyses for overlapping reports and cohorts

Outcome	Analysis	Handling of overlapping reports/cohorts	Effect sizes / studies	Participants	Hedges' g (95% CI)	p value	I ²	Prediction interval	Interpretation
BMI	Primary analysis	Primary BMI report selected	3 studies	250	-0.15 (-0.40, 0.10)	0.243	0.0%	-0.69 to 0.40	No significant effect
BMI	Sensitivity analysis	Alternative BMI report substituted	3 studies	249	-0.09 (-0.34, 0.16)	0.485	0.0%	-0.63 to 0.46	Conclusion unchanged
Muscular fitness	Primary analysis	Report-level clustering	27 effect sizes / 10 reports	553	0.30 (0.07, 0.53)	0.014	53.4%	-0.35 to 0.95	Positive effect
Muscular fitness	Sensitivity analysis	Cohort-level clustering	27 effect sizes / 9 independent cohorts across 10 reports	493	0.31 (0.05, 0.56)	0.020	56.5%	-0.39 to 1.00	Conclusion unchanged

Note. BMI, body mass index; CG, control group; CI, confidence interval; I², heterogeneity statistic. For BMI, Popowczak 2022a and Popowczak 2022b were identified as overlapping reports from the same cohort. The primary analysis used the BMI estimate from Popowczak 2022b, and the sensitivity analysis substituted the BMI estimate from Popowczak 2022a to assess robustness. For muscular fitness, the primary analysis treated Mitić et al. 2024 and Jovanović et al. 2024 as separate study reports, whereas the sensitivity analysis assigned them to the same cohort-level cluster. "Conclusion unchanged" indicates that the direction and statistical significance of the pooled result were unchanged compared with the primary analysis.

Supplementary Figure S8. Cohort-level sensitivity forest plot for muscular fitness



Supplementary Table S11. Cohort-level sensitivity subgroup analyses for muscular fitness

Moderator	Subgroup	K	Studies	Tabata	CG	Hedges' g (95% CI)	p value	I ²	Moderator p
Gender	Male	6	1	30	30	0.24 (-0.33, 0.80)	0.395	0.0%	0.117
	Female	4	2	50	50	0.74 (0.26, 1.23)	0.004	81.3%	0.117
	Mixed	17	6	187	146	0.17 (-0.10, 0.44)	0.213	47.8%	0.117
Age subgroup	Younger subgroup	12	5	111	80	0.61 (0.35, 0.87)	<0.001	45.0%	0.002
	Older subgroup	15	4	156	146	0.06 (-0.15, 0.26)	0.575	35.9%	0.002
Population type	General Population	22	7	249	209	0.29 (0.00, 0.58)	0.046	65.4%	0.761
	Athletes	5	2	18	17	0.41 (-0.29, 1.10)	0.240	0.0%	0.761
Intervention duration	2 weeks	2	1	20	20	0.55 (-0.14, 1.25)	0.113	81.6%	0.153
	4 weeks	3	1	29	23	-0.28 (-0.88, 0.33)	0.353	0.0%	0.153
	6 weeks	8	3	61	30	0.48 (0.04, 0.92)	0.033	0.0%	0.153
	10 weeks	6	2	97	93	0.10 (-0.31, 0.50)	0.622	14.1%	0.153
	12 weeks	8	2	60	60	0.52 (0.10, 0.94)	0.017	78.3%	0.153
Tabata protocol type	Classic Tabata	8	3	92	56	0.26 (-0.22, 0.74)	0.277	71.3%	0.818
	Modified Tabata	18	5	165	161	0.33 (-0.03, 0.69)	0.074	56.4%	0.818
Physiological component	Power	10	5	184	150	0.38 (0.05, 0.71)	0.025	81.2%	0.721
	Strength	5	3	42	41	0.26 (-0.23, 0.75)	0.289	0.0%	0.721
	Endurance	8	6	229	189	0.21 (-0.11, 0.54)	0.183	39.9%	0.721
	Speed	4	3	123	119	0.40 (-0.00, 0.81)	0.051	70.0%	0.721
Anatomical/functional category	Lower-Body Muscular Fitness	12	8	247	206	0.33 (0.03, 0.63)	0.034	74.8%	0.827
	Upper-Body Muscular Fitness	6	4	110	74	0.32 (-0.08, 0.72)	0.115	21.1%	0.827
	Core Endurance	5	5	186	176	0.21 (-0.15, 0.56)	0.237	35.7%	0.827
	Speed and Agility	4	3	123	119	0.38 (-0.02, 0.78)	0.064	70.0%	0.827

Note. K, number of effect sizes; Studies, number of independent cohort-level clusters; Moderator p, test of subgroup differences. Participant totals count each cohort once per subgroup. Mitić et al. (2024) and Jovanović et al. (2024) were treated as one cohort-level cluster.

Supplementary Table S12. Cohort-level sensitivity meta-regression analyses for muscular fitness

Predictor	K	Studies	β (95% CI)	p value	Residual I^2
Session frequency	27	9	0.043 (-0.120, 0.205)	0.595	59.3%
Session time	26	8	0.010 (-0.011, 0.030)	0.331	59.1%
Pure Tabata time/session	26	8	0.004 (-0.024, 0.031)	0.789	62.4%
Total sessions	27	9	0.012 (-0.027, 0.052)	0.529	58.7%

Note. β , meta-regression coefficient; CI, confidence interval; I^2 , heterogeneity statistic. Studies indicates the number of independent cohort-level clusters. Pure Tabata time/session refers to the time specifically devoted to Tabata-style exercise within each intervention session. Total sessions were calculated as session frequency multiplied by intervention duration in weeks.