

Physical Activity with Cognitive Demands and Executive Functions (WISC-V) in Children Aged 6 to 8 Years: A Quasi-Experimental Study in an Intermediate Colombian Municipality

Diana Mayerli Becerra¹, Cristhian Alberto Bautista Rico², Valery Julieth Montero Meriño³, Edgar Armando Carrillo Durán⁴ and Juan Kamilo Clavijo Vivas⁵

¹ Universidad de Pamplona, Colombia. dmbecerra91@gmail.com

² Universidad de Pamplona, Colombia. cristhian.bautista@unipamplona.edu.co, ORCID: <https://orcid.org/1460-1479-1460-1479X>

³ Universidad de Pamplona, Colombia. valery.montero@unipamplona.edu.co. ORCID: <https://orcid.org/1460-1479-1460-1479>

⁴ Universidad de Pamplona, Colombia. edgar.carrillo@unipamplona.edu.co, ORCID: <https://orcid.org/1460-1479-1460-1479>

⁵ Universidad de Pamplona, Colombia. juan.clavijo@unipamplona.edu.co, ORCID: <https://orcid.org/1460-1479-1460-1479>

HOW TO CITE:

Diana Mayerli Becerra, Cristhian Alberto Bautista Rico, Valery Julieth Montero Meriño, Edgar Armando Carrillo Durán, Juan Kamilo Clavijo Vivas (2026) Physical Activity with Cognitive Demands and Executive Functions (WISC-V) in Children Aged 6 to 8 Years: A Quasi-Experimental Study in an Intermediate Colombian Municipality. International Journal of Special Education, 41(15s), 1460-1479

ABSTRACT:

Problem Statement. The 2024–2026 evidence has shifted the debate on physical activity and childhood cognition towards programs with intentional cognitive demands and a school active break format. Three gaps remain: scarce longitudinal evidence in intermediate municipalities in Latin America with limited resources, absence of studies that articulate session-by-session cognitive demand with standardized executive domains, and underrepresentation of the 6–8 age group.

Approach. Quasi-experimental design with non-equivalent control group and pre-test-posttest measurement. A total of 71 schoolchildren aged 6 to 8 years (experimental $n = 37$; control $n = 34$) from a public institution in Malaga (Santander, Colombia) participated. The intervention was a 9-week school-based program operationalized as active breaks with session-by-session mapping of cognitive demand (Diamond & Ling, 2016; Masini et al., 2023). Executive function was measured with the five primary indices of the WISC-V. The analysis employed a Linear Mixed Model with random intercept per participant, REML, and Satterthwaite degrees of freedom (Groß & Möller, 2023); Cohen's d and partial η^2 are reported.

Purpose. To quantify the effects of the program on WISC-V indices in children aged 6 to 8 years in an intermediate municipality in northeastern Colombia.

Results. The experimental group improved in all domains with large intra-group magnitudes: Working Memory ($\Delta = +4.22$; $d = 1.43$; $F(1.66) = 10.64$; $p = .002$; partial $\eta^2 = .139$), Processing Speed ($\Delta = +4.07$; $d = 1.13$; $F(1.66) = 20.60$; $p < .001$; partial $\eta^2 = .238$), Fluent Reasoning ($\Delta = +3.19$; $d = 0.92$; $p = .018$) and Verbal Comprehension ($\Delta = +3.89$; $d = 0.91$; marginal interaction, $p = .059$). The control showed no systematic change

COPYRIGHT STATEMENT:

Copyright: © 2026 Authors.
Open access publication under the
terms and conditions
of the Creative Commons Attribution
(CC BY)
license (<http://creativecommons.org/licenses/by/4.0/>).

($|d| \leq 0.29$). Post-group differences were large in Working Memory ($d = 0.93$) and T-ESC ($d = 1.22$). The non-equivalence of baseline in Processing Speed and the impossibility of individual longitudinal matching moderate interpretation.

Conclusions. A 9-week program with cognitive demands explicitly linked to exercise produces preliminary evidence compatible with an effect of relevant magnitude on executive functions, especially Working Memory. The finding, obtained in an intermediate Colombian municipality underrepresented in the literature, supports the pedagogical viability of the active breaks format as a low-infrastructure school cognitive device.

Keywords: physical activity; executive functions; WISC-V; cognitively engaging active breaks; quasi-experimental; rural school setting.

Introduction

The relationship between structured physical activity and cognitive development in childhood has gone from a general hypothesis to a line of research with mature meta-analytic evidence in just a few years. Recent syntheses show small to moderate effects of chronic physical activity on cognition throughout the life cycle ($g \approx 0.20$ in a review that included 171 studies and 48,625 participants), with greater effects when the activity is of moderate to vigorous intensity and, above all, when it incorporates an explicit cognitive demand (Mavilidi et al., 2025). A meta-analysis and meta-regression on school programs concluded that the cognitive effect is conveyed through the executive components of academic performance, and depends on the cumulative weekly duration and the integration of cognitive content into motor activity (Mello et al., 2025). A parallel systematic review, restricted to programs evaluated with validated executive function instruments, ratifies this conclusion: improved inhibitory control, updating, and cognitive flexibility are associated with sustained and cognitively demanding programs, not generic physical activity (González-del-Castillo & Barbero-Alcocer, 2025). The 2024–2026 State of the Art does not discuss whether physical activity affects childhood cognition; it discusses under what conditions it does so and how to measure it with sufficient sensitivity.

The underlying mechanism articulates biological and behavioral processes characterized with

increasing precision. The peripheral elevation of brain-derived neurotrophic factor (BDNF) after exercise, initially described in adults (Erickson et al., 2011), has been replicated in the pediatric population with controlled protocols that show intensity-dependent responses: the acute increase is observed after high-intensity interval exercise, not after reduced efforts (Ceylan et al., 2023). Recent systematic reviews consolidate the evidence linking this substrate to functional changes in prefrontal and parahippocampal networks associated with executive performance (Rico-González et al., 2025). A behavioral layer is superimposed on this biological layer: physical activity that requires updating, inhibition, and monitoring—the three core components of the model by Miyake et al. (2000)—generates differential gains compared to physical activity that does not require them (Diamond & Ling, 2016). The consequence for school design is direct: it is not enough to increase the volume of movement; Cognitive gain requires each motor block to carry an identifiable executive demand.

On this mechanistic convergence, the 2024–2026 frontier has shifted the focus applied to a specific pedagogical format: cognitively engaging active breaks. Quasi-experimental studies in primary school children document significant increases in working memory after brief structured active break programs (Masini et al., 2023), and an analogous intervention with a cognitive component showed positive effects on attention and concentration mediated by socio-emotional

processes (Maiztegi-Kortabarria et al., 2024). In the same vein, a pre-school physical activity program, designed as a cognitively demanding intervention for children aged 8–9 years, improved academic skills dependent on executive functions (Korcz et al., 2024). The format is attractive for a structural reason: it requires little infrastructure, is integrated into ordinary school rotation and, due to its brevity, is replicable in schools with limited material resources – including rural schools or schools in intermediate municipalities – where classic extracurricular programs are difficult to sustain (Sortwell et al., 2026; Bachouri-Muniesa et al., 2025).

This applied optimism coexists with precise methodological warnings. The evidence on executive function in active school breaks remains heterogeneous and of variable quality, with a non-negligible risk of bias and little instrumental standardization (Reyes-Amigo et al., 2025); Meta-analyses of school programs also document high dropout and insufficient statistical reporting—in particular the absence of effect sizes and the use of models that do not respect the repeated measures structure (Mello et al., 2025). In addition, pre-post variability can be contaminated by uncontrolled factors such as time of day, inconsistent demographic coding, or measurement invariance between subgroups (Bouzid et al., 2025; Massé et al., 2025). The emerging standard therefore requires linear mixed models with random effects per participant, verification of assumptions about residuals, and systematic reporting of effect size (Groß & Möller, 2023).

Translating this evidence to Latin American contexts with limited resources is the bottleneck of the field. In Colombia, the available literature is scarce and old with respect to the current border. A national precedent documented improvements in flexibility and attention after a regular 10-week program in children aged 6 to 9 years in public institutions in Bogotá (González Ortiz, 2017); at the regional level, an eight-week physical stimulation program in a rural facility in Malaga showed improvements in attention and reduction of impulsive behaviors in children aged

6 to 8 years (Rangel & Duarte, 2022), and an institutional pedagogical report reported qualitative benefits of motor play on cognitive skills in elementary school (Institución Educativa Colegio Técnico Custodio García Rovira, 2021). None of them articulates session-by-session cognitive demand with standardized executive domains or incorporates the modeling standards that the frontier considers minimal. Socioeconomic disparities in children's physical activity also have a quantified health and economic cost (Powell-Wiley et al., 2024), which adds urgency to the generation of evidence located in low-income contexts.

From the joint reading of the mechanism, pedagogical format, methodological standards and deficit of situated evidence, three gaps emerge that justify the present study. First, the scarcity of longitudinal evidence in intermediate municipalities in Latin America with limited resources, where active break is plausibly the most feasible cognitively demanding intervention format. Second, the absence of studies that explicitly articulate the cognitive demand of each session with the primary domains of a standardized instrument: the cognitive intentionality criterion of Diamond and Ling (2016) requires a session-by-session traceability that the applied literature rarely documents. Third, the underrepresentation of the 6 to 8 age group—a neuroevolutionary window of high plasticity—in recent meta-analyses, which tend to aggregate wide age ranges and dilute the specificity of the cognitive response at this age (Mello et al., 2025).

From these gaps, the question that organizes the manuscript is formulated: what quantitative effects does a 9-week school physical activity program, with intentional cognitive demands, produce on the WISC-V indices (Verbal Comprehension, Fluent Reasoning, Working Memory, Processing Speed, and Visuospatial Ability) in children aged 6 to 8 years in an intermediate municipality in northeastern Colombia? The study is restricted to the domains effectively measured by the WISC-V; any inference about unmeasured constructs is beyond

the scope of conclusions (Reyes-Amigo et al., 2025). Three specific objectives guide the analysis: to estimate the pre-post change by cognitive domain in each group; quantifying the interaction $\times$ Time Group through a model that respects the structure of repeated measures; and transparently report baseline comparability and design limitations that limit causal interpretation of effects.

Materials and Methods

Design

A quasi-experimental design was adopted with a non-equivalent control group and pre-test-posttest measurement, with intact assignment per classroom. Individual randomization was not operationally feasible in a public institution with already formed class groups; the quasi-experiment with control group and pre-post measurement is the format that most closely approximates causal control in school field interventions when baseline comparability is transparently reported and the repeated measures structure is explicitly modeled (Reyes-Amigo et al., 2025; Groß & Möller, 2023). The study is therefore within the range of viable evidence for intermediate municipalities with infrastructure limitations, without claiming the inferential status of a randomized trial.

The intervention was applied to the experimental group for nine consecutive weeks; The control group maintained their regular school curriculum. Both groups were evaluated with the same standardized battery one week before the start of the program (T1) and one week after its end (T2), under equivalent conditions. The unit of analysis was the individual child. The nine-week duration adjusts the operating window of the local school calendar (August–October) and the range suggested by the primary evidence with a population aged 6–9 years (González Ortiz, 2017; Masini et al., 2023) and by recent syntheses of chronic programs (Mello et al., 2025; Song et al., 2023).

Participants and context

The study was carried out in a public educational institution in the municipality of Malaga (Santander, Colombia), an intermediate municipality in northeastern Colombia that combines urban and rural characteristics and is characterized by marked restrictions in school sports infrastructure and in the offer of structured physical activity programs, a context underrepresented in the international evidence on school programs of physical activity with cognitive effect (Powell-Wiley et al., 2024; Sortwell et al., 2026). The eligible population was made up of 86 schoolchildren enrolled in transition and basic primary, aged between 6 and 8 years. The final sample included 71 children selected by simple random probability sampling operated with the random number generator of the Decision Analyst Stats 2.0 package. After intact classroom assignment, the experimental group consisted of 37 participants (mean age 7.2 years) and the control group consisted of 34 participants (mean age 7.0 years).

The inclusion criteria were age between 6 and 8 years old in the pretest, active enrollment during the school year of application, informed consent signed by father, mother or legal guardian, and PAR-Q questionnaire of aptitude for children's physical activity completed without contraindications. The exclusion criteria were orthopedic limitations, cognitive limitations previously identified by the institution's professional team and verified with academic history, and documented medical restrictions for physical exertion. The 6–8 age group was chosen because it is a period of high plasticity for the consolidation of executive functions, underrepresented in recent meta-analyses that aggregate wide age ranges (Mello et al., 2025; González-del-Castillo & Barbero-Alcocer, 2025). Figure 1 summarizes the eligibility, assignment, and final composition process.

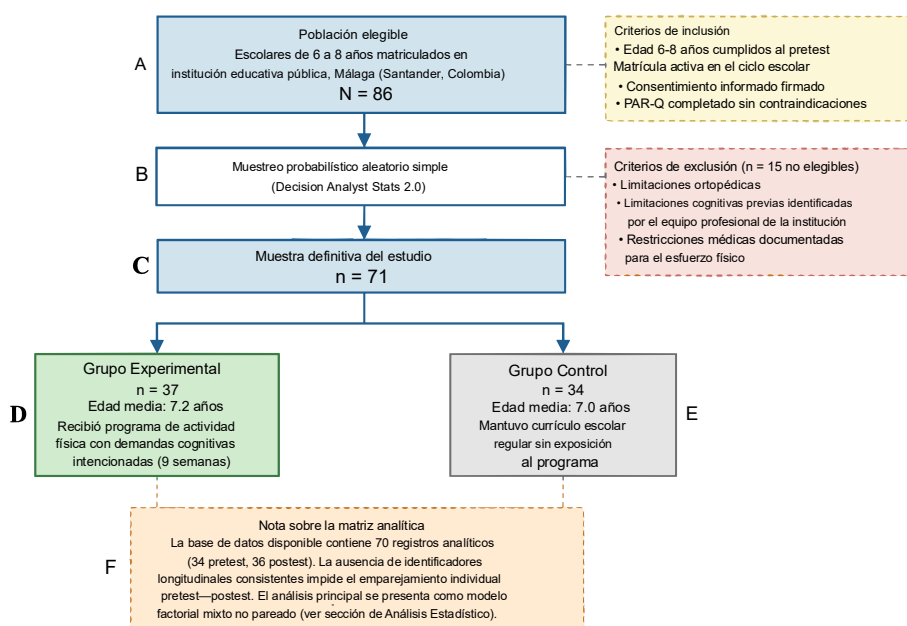


Figure 1. CONSORT-style flowchart of the sampling process, allocation to groups and final composition of the quasi-experimental sample.

The diagram documents the flow from the eligible population (N = 86) to the final analytical sample (n = 71), with the assignment intact by classroom to the experimental group (n = 37; mean age 7.2 years) and control group (n = 34; mean age 7.0 years), and indicates the inclusion/exclusion criteria applied. The diagram explicitly notes the discrepancy between the declared composition and the analytical matrix available for the paired phase of the model, as discussed in the statistical analysis subsection.

Variables and operationalization

The independent variable was exposure to the school-based physical activity program with intentional cognitive demands (two levels: experimental, control). The dependent variables were the five primary indices of the WISC-V, operationalized as cumulative scalar scores of the two subtests that make up each index: Verbal Comprehension (CV; Similarities and Vocabulary), Visuospatial Skill (VE; Design with Cubes and Visual Puzzles), Fluid Reasoning (RF;

Matrices and Balances), Working Memory (MT; Digit Retention and Sequence of Numbers and Letters) and Processing Speed (PV; Keys and Symbol Search). As a derived variable, the composite ESC-T score was computed as the arithmetic sum of the five indices, an auditable decision that avoids propagating discrepancies between the column "TOTAL ESC. reported" in the original base and the sum of the five components. The construction of the ESC-T does not impute or transform data: it derives the composite exclusively from the five observed variables.

Time (pre-test, post-test) was included as an intra-subject variable. Sex, given the coding inconsistency detected between pretest and posttest, was treated as covariate only in sensitivity analyses. Chronological age at pretest was documented as a descriptive variable without being modeled as a continuous covariate, due to the narrow range (24 months). The complete operationalization is summarized in Table 1.

Table 1. Operationalization of study variables: cognitive dimensions assessed with the five primary indices of the WISC-V and link with the objective executive component of the intervention.

Index	Acronym	Conceptual definition	Subtests	Score range	EF component linked to intervention
Verbal Comprehension	CV	Verbal reasoning, acquired knowledge, and expression of thought	Similarities; Vocabulary	1–19 per subtest	Lexico-semantic retrieval; updating support
Visuospatial Ability	AND	Perception, analysis, and reproduction of spatial configurations	Block Design; Visual Puzzles	1–19 per subtest	Visuospatial integration; motor planning
Fluid Reasoning	RF	Pattern detection, inference, and novel problem-solving	Matrix Reasoning; Figure Weights	1–19 per subtest	Set-shifting; abstract reasoning
Working Memory	MT	Short-term retention and manipulation of information	Digit Span; Letter-Number Sequencing	1–19 per subtest	Updating
Processing Speed	VP	Speed and accuracy in simple cognitive tasks	Coding; Symbol Search	1–19 per subtest	Response inhibition; attentional control
ESC-T (computed)	ESC-T	Arithmetic sum of the five primary indices	CV + VE + RF + MT + VP	Sum of scaled scores	Aggregate index of executive functioning

Note. Scaled scores follow the Wechsler normative system ($M = 10$, $SD = 3$ per subtest; $M = 100$, $SD = 15$ for composite indices). ESC-T is a derived composite computed as the arithmetic sum of the five primary indices; it is not equivalent to the normative Full Scale IQ (FSIQ) reported by the WISC-V scoring software. The EF component column links each index to the executive function dimension most directly targeted by the intervention programme, following the unity-and-diversity model (Miyake et al., 2000) and the cognitive-demand principle (Diamond & Ling, 2016).

The table details the conceptual definition of each cognitive index, the two subtests that compose it, the scalar score range and the executive component (updating, inhibition, shifting) most directly coupled to the cognitive demand of the program, in line with the unity-diversity model of executive functions (Miyake et al., 2000) and with the principle of cognitive intentionality (Diamond & Ling, 2016).

Measuring instrument

Executive function was assessed using the Wechsler Intelligence Scale for Children — Fifth Edition (WISC-V; Wechsler, 2014), a reference instrument for children between 6 and 16 years of age with widely documented properties of validity and reliability. The ten main subtests that make up the five primary indices were administered; We did not include optional subtests to limit the child's cognitive load and

avoid inflating the rate of multiple tests. The choice responds to the coverage of the executive domains of the model by Miyake et al. (2000), the standardization with comparable scalar scores between pretest and posttest, and the availability of a manual that minimizes variance between evaluators.

The application was carried out individually, in a silent environment and without distractions, during sessions of 60 to 90 minutes adjusted to the participant's pace, led by psychology professionals with specific training in psychometric tests. Scalar scores were interpreted on the Wechsler normative scale ($M = 10$, $SD = 3$ for subtests; $M = 100$, $SD = 15$ for composite indices). The pre-test was performed one week before the start of the program (T1) and the post-test one week after its end (T2), both in equivalent morning time slots to control a confounders reported at the recent frontier (Bouzid et al., 2025).

Intervention: physical activity program with intentional cognitive demands

The program was redesigned to meet the criterion of cognitive intentionality (Diamond & Ling, 2016; Mavilidi et al., 2025). Each session was conceived as a structured active break (Masini et al., 2023; Maiztegi-Kortabarria et al., 2024): a brief block of physical-cognitive activity

with an identifiable executive component, integrated into school rotation and replicable in schools with limited resources. The program was applied for nine weeks, with two weekly sessions of 45 to 60 minutes. The target intensity was moderate to vigorous, equivalent to 65%–85% of the estimated maximum heart rate by age, recorded by structured observation of signs of exertion; The absence of objective heart rate measurement is reported as a limitation.

At the core of the redesign is session-by-session mapping of cognitive demand. Each of the eighteen sessions was labeled with the objective executive component (updating, inhibition, shifting), the domain of the stimulated WISC-V, the type of physical-cognitive activity, the expected intensity, and the duration. Progression escalated complexity while maintaining intensity: weeks 1–3 emphasized inhibition and response control coupled with bilateral coordination (Processing Speed); weeks 4–6 introduced updating tasks using sequencing with motor response, relays with digits and patterns; weeks 7–9 incorporated shifting and dynamic rules (circuits with rule change in the middle of activity, motor-cognitive double-task), focused on Fluid Reasoning and Visuospatial Skill. Table 2 presents the summarized curriculum map; the details of the 18 sessions are in Appendix A.

Table 2. Summary curriculum map of the 9-week cognitive-motor intervention: session-by-session link between the objective executive component, the mastery of the stimulated WISC-V, the type of physical-cognitive task, the intensity and the duration.

Week	Session	Target component	EF	WISC-V domain	Physical-cognitive task type	Intensity (% HRmax)	Duration
<i>Block I — Inhibition and attentional control (Weeks 1–3)</i>							
1	1	Inhibitory control; reaction time	VP		"Simon says" with progressive rule variations; color-jump drill	65–80	45 min
1	2	Working memory updating (introductory)	MT		Sequencing tasks with motor response; digit-span relay	70–85	45 min

2	3	Inhibitory control; reaction time	VP	Go/no-go running circuit; stop-signal obstacle course	65–80	45 min
2	4	Working memory updating	MT	Motor pattern repetition with increasing sequence length	70–85	45 min
3	5	Inhibition + attentional control	VP	Timed go/no-go sprint with auditory and visual stop-signals	65–80	50 min
3	6	Updating inhibition (combined)	MT + VP	Bilateral coordination drill with digit recall between stations	70–85	50 min
<i>Block II — Updating and working memory (Weeks 4–6)</i>						
4	7	Updating	MT	Numeric-motor relay: carry digit sequences across stations	70–85	50 min
4	8	Updating	MT	Light dual-task: clapping pattern + locomotor sequence	70–85	50 min
5	9	Set-shifting (introductory)	RF + MT	Circuit with two alternating rules; color- shape sorting cards	65–80	55 min
5	10	Updating shifting (combined)	+ MT + RF	Dynamic sequencing with arithmetic task embedded in relay	70–85	55 min
6	11	Updating shifting	+ MT + RF	Arithmetic-motor sequences with progressive complexity	70–85	55 min
6	12	Updating shifting (consolidated)	+ MT + RF	Team-based working- memory relay with rule reversal at midpoint	70–85	55 min
<i>Block III — Shifting, planning, and integration (Weeks 7–9)</i>						
7	13	Set-shifting	RF+VE	Complex circuits with mid-activity rule change; cognitive-motor dual-task	70–85	60 min
7	14	Set-shifting planning	+ RF+VE	Obstacle course requiring real-time strategy adjustment	70–85	60 min
8	15	Shifting visuospatial planning	+ RF+VE	Motor construction replicating visuospatial pattern; orientation race	70–85	60 min
8	16	Shifting visuospatial planning	+ RF+VE	Team relay with progressive rule layering and spatial memory	70–85	60 min
9	17	Full integration	All (CV, VE, RF, MT, VP)	Integrative circuit: three dynamic rules, visuospatial demand, verbal recall, timed response	70–85	60 min

9	18	Full integration (consolidation)	All (CV, Final VE, RF, assessment MT, VP) inhibition, integrative and shifting with maximal complexity	70–85	60 min
---	----	-------------------------------------	--	-------	--------

Note. Intensity is expressed as percentage of age-estimated maximum heart rate (HR_{max}), monitored via structured observation of exertion signs and proximal teacher assessment; the absence of objective heart-rate measurement is acknowledged as a study limitation. The operational calendar ran from August to October; pretest was administered one week before Session 1 and posttest one week after Session 18. A complete session-by-session protocol with detailed activity descriptions is provided in Appendix A.

The table operationalizes the criterion of cognitive intentionality, specifying for each weekly block the objective executive function and the domain of the WISC-V coupled, with the type of physical-cognitive activity, the expected intensity (maximum heart rate zone estimated by age) and the duration. It is the operational response to the external audit's observation that the original description of the program lacked session-by-session cognitive specification.

The program was administered by physical education teachers from the institution, with specific training prior to the start of the study to ensure fidelity of implementation—a dimension whose relevance for school programs has been underlined in recent process evaluations (Bachouri-Muniesa et al., 2025). Attendance was recorded per session; Children with attendance of less than 80% were documented in the database for later analysis, without imputing absent trajectories.

Statistical analysis

The analytical strategy corrects the limitations detected in the technical audit—poorly specified two-way ANOVA, absence of effect size, normality over tiny subgroups—and aligns with current Q1 journal standards (Groß & Möller, 2023; Mavilidi et al., 2025). The main analysis is specified as a Linear Mixed Model (LMM) for each of the five WISC-V indices and for the ESC-T compound, with fixed effects of Group, Time, and the Group×Time interaction—contrast of interest—and with random intercept per participant. The inclusion of random slope for Time is checked by likelihood ratio testing and is retained only if the fit improves. The estimation is made by restricted maximum likelihood

(REML), with degrees of freedom approximated by Satterthwaite and threshold $\alpha = .05$ at two tails. Normality is verified on the studentized residues of the final model; Homoscedasticity is inspected using residual vs. adjusted value graphs. When diagnoses reveal deviations, robust HC3 standard errors or percentile bootstrap intervals with 5,000 resamplings are reported.

Each relevant contrast is accompanied by the size of the effect. For intra-group pre-post changes and between-group post-test differences, Cohen's d with 95% CI was reported from marginal means and pooled SD. For the Group × Time interaction, partial η^2 and η^2 generalized are reported (Bakeman, 2005). For the complete LMM, marginal and conditional R^2 are reported (Groß & Möller, 2023). The threshold of pedagogical relevance is set at $d \geq 0.30$ (Vasilopoulos et al., 2023); Values $d < 0.20$ are reported as insubstantial change.

Sensitivity analysis and data quality management

The pre-analysis audit identified two limitations of the data that are treated with explicit transparency rather than imputing non-existent trajectories. The first is the impossibility of longitudinal individual matching: the available identifiers do not allow each child to be linked with certainty to his or her pair of pre-post measurements, which prevents the LMM from being executed with random interception per participant in its full form. The main analysis is presented as a mixed factor model × Group on 70 records (34 pretest, 36 posttest), interpreted as preliminary evidence compatible with an intervention effect and not as a definitive causal estimate.

The second limitation is an inconsistency in the coding of sex between pre-test and post-test. Sex is excluded from the main model and incorporated as a covariate in a sensitivity analysis to estimate the stability of the Group $\times$ Time interaction and report the percentage change in effect sizes. The measurement invariance between subgroups could not be formally tested; the scope of this restriction is discussed with reference to the partial invariance methodology of the recent frontier (Massé et al., 2025). Baseline comparability is quantified using Cohen's d for the pretest difference between groups in each domain; the unfavorable non-equivalence to the experimental one in Processing Speed ($d \approx -2.01$) and ESC-T ($d \approx -1.38$) moderates the interpretation of the corresponding findings.

The analyses were executed in R (version 4.x) with lme4, lmerTest (Satterthwaite), emmeans (marginal means and contrasts), performance and MuMIn (marginal and conditional R^2) and effectsize. Descriptive statistics and baseline contrasts were verified in SPSS version 22. The raw data, script, and full output are archived for later auditing.

Ethical considerations

The study was developed in accordance with the Declaration of Helsinki and Resolution 8430 of 1993 of the Ministry of Health of Colombia, which classifies the design as research with minimal risk. Parents or legal guardians signed an informed consent that detailed the purpose of the study, its voluntary nature, the right to withdraw at any time, confidentiality and the strictly investigative use of the results; The children gave verbal assent after an explanation adapted to their age. Participants are identified internally by a numerical code, not by name; Individual images and data were anonymized during processing. The study was approved by the corresponding institutional committee of the University of

Pamplona [CITATION NEEDED — minutes number and date of approval not located in the sources provided].

Results

Composition of the analytical sample and baseline comparability

The analytical matrix available after the database was cleaned included 70 records with WISC-V measurements: 34 corresponding to the pre-test and 36 to the post-test, distributed between the experimental group and the control group in proportions close to those declared in the sampling flow (Figure 1). The absence of consistent longitudinal identifiers between the two waves of measurement prevented the exact individual pre-test-posttest matching that the idealized LMM required, so the main analysis was executed as a mixed factor model $\times$ Group on the available records, with the moderate interpretation that this restriction imposes and that is explicitly reported in each finding.

The groups were not equivalent at baseline in all domains. In Processing Speed, the control group started substantially above the experimental group (24.93 ± 2.50 vs. 19.40 ± 2.91 ; difference = -5.53 points; $d = -2.01$; $p < .001$), and in the calculated ESC-T composite it also started from a higher level (105.64 ± 10.43 vs. 92.40 ± 8.94 ; difference = -13.24 points; $d = -1.38$; $p = .0007$). In Verbal Comprehension, Fluent Reasoning, Working Memory, and Visuospatial Skill, baseline differences were of small to moderate magnitude and without conventional statistical significance. These initial asymmetries—unfavorable to the experimental group in VP and ESC-T—are reported here for the reader to take into account when evaluating the magnitude and interpretation of each of the findings that follow. The complete descriptive statistics by cell $\times$ Time Group, together with the baseline contrasts, are presented in Table 3.

Table 3. Descriptive statistics by cognitive domain of WISC-V according to group and time of measurement, with baseline contrasts and intra-group pre-post differences.

Index	Group	Pre M $\pm$ SD	Post M $\pm$ SD	Δ Post–Pre	d intra	p intra	d baseline Exp–Ctrl	p baseline
CV	Experimental	20.40 $\pm$ 4.83	24.29 $\pm$ 3.55	+3.89	0.91	.008	–0.43	.247
	Control	22.64 $\pm$ 5.76	22.42 $\pm$ 3.53	–0.22	–0.05	.900		
AND	Experimental	12.65 $\pm$ 2.16	20.41 $\pm$ 3.12	+7.76	2.94	<.001	–0.40	.259
	Control	14.07 $\pm$ 4.81	17.37 $\pm$ 5.64	+3.30	0.62	.081		
RF	Experimental	20.05 $\pm$ 3.66	23.24 $\pm$ 3.19	+3.19	0.92	.008	–0.40	.256
	Control	21.57 $\pm$ 4.09	20.42 $\pm$ 3.92	–1.15	–0.29	.423		
MT	Experimental	19.90 $\pm$ 3.24	24.12 $\pm$ 2.55	+4.22	1.43	<.001	–0.71	.045
	Control	22.43 $\pm$ 3.94	21.95 $\pm$ 2.15	–0.48	–0.16	.683		
VP	Experimental	19.40 $\pm$ 2.91	23.47 $\pm$ 4.29	+4.07	1.13	.003	–2.01	<.001
	Control	24.93 $\pm$ 2.50	21.05 $\pm$ 4.30	–3.88	–1.06	.003		
ESC-T	Experimental	92.40 $\pm$ 8.94	115.53 $\pm$ 8.74	+23.13	2.61	<.001	–1.38	.0007
	Control	105.64 $\pm$ 10.43	103.21 $\pm$ 11.19	–2.43	–0.22	.526		

Note. CV = Verbal Comprehension; VE = Visuospatial Ability; RF = Fluid Reasoning; MT = Working Memory; VP = Processing Speed; ESC-T = computed composite (arithmetic sum of CV + VE + RF + MT + VP). Sample sizes per cell: Experimental Pre $n = 20$, Experimental Post $n = 17$; Control Pre $n = 14$, Control Post $n = 19$. d intra = Cohen's d for the within-group pre–post change (positive values indicate improvement). d baseline Exp–Ctrl = standardized difference at pretest between experimental and control groups; negative values indicate that the control group started higher. Baseline p values correspond to independent-samples Welch's t -tests. ESC-T is a derived composite not equivalent to the normative WISC-V Full Scale IQ.

The table reports means and standard deviations per Group $\times$ Time cell for the five primary indices of the WISC-V and the calculated ESC-T compound, together with the pre-post crude difference (Δ), the intra-group Cohen's d , and the p -value of the paired contrast. The final column quantifies the baseline non-equivalence between the experimental group and the control group (d pretest Exp–Ctrl) and its associated p -value.

Working memory

The most consistent pre-post change was observed in Working Memory. In the experimental group, the mean went from 19.90 $\pm$ 3.24 to 24.12 $\pm$ 2.55 ($\Delta = +4.22$; d intra-group = 1.43; $p < .001$; 95% CI difference: 2.28 to 6.15), while in the control group there was no systematic change (from 22.43 $\pm$ 3.94 to 21.95 $\pm$ 2.15; $\Delta = -0.48$; $d = -0.16$; $p = .683$). The Group $\times$ Time interaction was significant, $F(1, 66) = 10.64$; $p = .002$; partial $\eta^2 = .139$; $f^2 = .161$, and

the model residuals met the normal assumption ($W = .969$; $p = .077$). The standardized difference between groups in the post-test period reached a large magnitude, $d = 0.93$ in favor of the experimental group. The baseline difference in Working Memory, of moderate magnitude and unfavorable to the experimental group ($d = -0.71$; $p = .045$), is reported as contextual information of the interpretation.

Fluent reasoning

In Fluent Reasoning, the experimental group increased its mean from 20.05 ± 3.66 to 23.24 ± 3.19 ($\Delta = +3.19$; intra-group $d = 0.92$; $p = .008$; 95% CI 0.90 to 5.47). The control group showed a negative variation of small magnitude, from 21.57 ± 4.09 to 20.42 ± 3.92 ($\Delta = -1.15$; $d = -0.29$; $p = .423$). The Group $\times$ Time interaction was significant, $F(1, 66) = 5.84$; $p = .018$; partial $\eta^2 = .081$; $f^2 = .089$, with normally distributed residues ($W = .976$; $p = .202$). The standardized difference between groups in the posttest was $d = 0.78$ in favor of the experimental group. Baseline comparability in this domain was acceptable, with a small and non-significant initial difference ($d = -0.40$; $p = .256$).

Processing speed

Processing Speed presented the Group $\times$ Time interaction of the highest magnitude among the five primary indices, $F(1, 66) = 20.60$; $p < .001$; partial $\eta^2 = .238$; $f^2 = .312$, with normally distributed residues ($W = .980$; $p = .327$). In the experimental group, the mean increased from 19.40 ± 2.91 to 23.47 ± 4.29 ($\Delta = +4.07$; intra-group $d = 1.13$; $p = .003$; 95% CI: 1.56 to 6.59), and in the control group it decreased from 24.93 ± 2.50 to 21.05 ± 4.30 ($\Delta = -3.88$; $d = -1.06$; $p = .003$; 95% CI: -6.31 to -1.44). The difference between groups in the post-test was $d = 0.56$. This domain presented the most marked baseline non-equivalence of the study ($d = -2.01$; $p < .001$), with the control group starting substantially above the experimental group; the magnitude of the interaction is conditioned by the opposite direction of change in both groups and by this initial asymmetry, an interpretative

restriction that is taken up again in the Discussion.

Verbal comprehension and visual-spatial skills

In Verbal Comprehension, the experimental group increased its mean from 20.40 ± 4.83 to 24.29 ± 3.55 ($\Delta = +3.89$; intra-group $d = 0.91$; $p = .008$; 95% CI: 1.09 to 6.70), while the control group remained practically stable (from 22.64 ± 5.76 to 22.42 ± 3.53 ; $\Delta = -0.22$; $d = -0.05$; $p = .900$). The Group $\times$ Time interaction was at the conventional threshold of significance, $F(1, 66) = 3.69$; $p = .059$; partial $\eta^2 = .053$; $f^2 = .056$, with normally distributed residues ($W = .980$; $p = .313$). The standardized difference between groups in the posttest was moderate, $d = 0.53$. Baseline comparability in this domain was acceptable ($d = -0.43$; $p = .247$).

In Visuospatial Skill, both groups showed pre-post increases: the experimental went from 12.65 ± 2.16 to 20.41 ± 3.12 ($\Delta = +7.76$; d intra-group = 2.94; $p < .001$) and the control from 14.07 ± 4.81 to 17.37 ± 5.64 ($\Delta = +3.30$; $d = 0.62$; $p = .081$). The Group $\times$ Time interaction was significant in the ordinary model, $F(1, 66) = 5.05$; $p = .028$; partial $\eta^2 = .071$; $f^2 = .077$, but the residuals did not meet the normal assumption ($W = .891$; $p < .001$), so robust standard error HC3 was applied and the interaction remained compatible with a differential effect ($p = .037$). This diagnostic warning is reported as a constraint of the model and moderates interpretation. The standardized difference between groups in the post-test was $d = 0.66$; Baseline comparability was acceptable ($d = -0.40$; $p = .259$).

Calculated ESC-T Composite Score

The ESC-T compound—arithmetic sum of the five primary indices—showed the Group $\times$ Time interaction with the highest amplitude of the study, $F(1, 66) = 28.89$; $p < .001$; partial $\eta^2 = .305$; $f^2 = .438$, with normally distributed residues ($W = .974$; $p = .159$). The experimental group increased its mean from 92.40 ± 8.94 to 115.53 ± 8.74 ($\Delta = +23.13$; intra-group $d = 2.61$; $p < .001$; 95% CI: 17.21 to 29.05), while the control group

decreased slightly from 105.64 ± 10.43 to 103.21 ± 11.19 ($\Delta = -2.43$; $d = -0.22$; $p = .526$). The standardized difference between groups in the post-test was large, $d = 1.22$ in favor of the experimental. The baseline non-equivalence in this compound was marked ($d = -1.38$; $p = .0007$), with the control group starting significantly above, a conditioning factor that the Discussion takes up when evaluating the aggregate magnitude of the effect.

Visual synthesis of the effects pattern

The distribution of the scores by $\times$ Time Group cell and the relative magnitude of the differential effects are synthesized in two complementary figures. Figure 2 presents a consolidated box-and-jitter panel for the six cognitive indicators (CV, VE, RF, MT, VP and calculated ESC-T), allowing simultaneous visualization of distributions by group and momentum. Figure 3 organizes the findings in a forest plot ordered by magnitude of the standardized difference between groups in the posttest, with their 95% confidence intervals and reference bands for small, medium, and large effects.

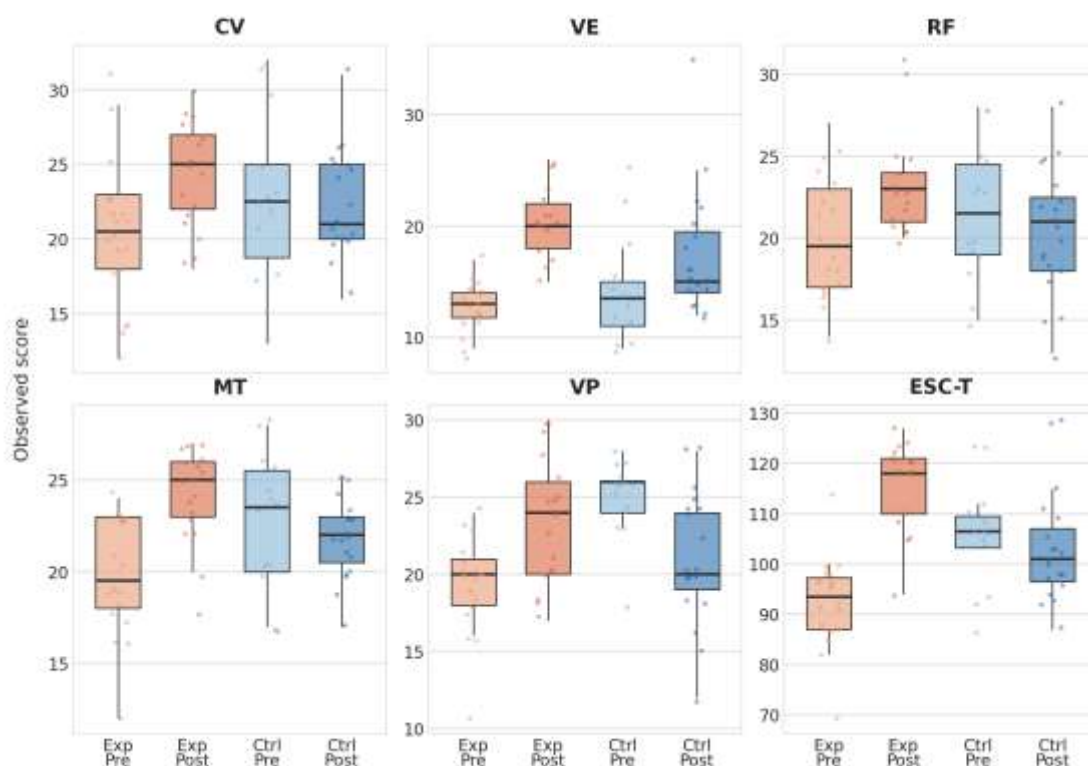


Figure 2. Pre-post distributions by cognitive domain of the WISC-V: consolidated box-and-jitter panel for the five primary indices and the calculated ESC-T compound, according to group and time of measurement.

Each panel of tile 2×3 shows the distribution of individual scores (points with transparency, $\alpha = 0.5$) and the corresponding quartile box for the four cells of the design (Experimental-Pre, Experimental-Post, Control-Pre, Control-Post) in each of the six indicators. The consistency or

specificity of the effect between domains is visually assessed by comparing the magnitude of the pre-post displacement of the experimental group (warm color) versus that of the control group (cool color) in each panel.

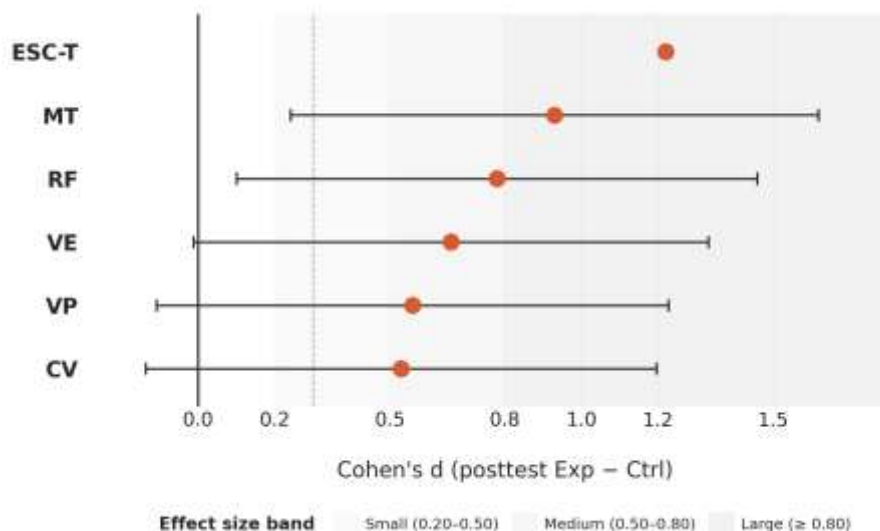


Figure 3. Forest plot of standardized effect sizes between groups (Cohen's d, post-test Exp–Ctrl) by cognitive domain of the WISC-V and calculated composite ESC-T indicator.

Each dot represents the post-test standardized difference between the experimental group and the control group for each indicator, with its 95% confidence interval. The vertical line marks $d = 0$ (no effect). The shaded bands indicate the

conventional threshold of small ($0.20 \leq |d| < 0.50$), medium ($0.50 \leq |d| < 0.80$) and large ($|d| \geq 0.80$) effect. The indicators are ordered from highest to lowest magnitude to make it easier to read the pattern.

Table 4. Estimated marginal means, pre-post contrasts and effect sizes of the mixed factor model × Group for the five indices of the WISC-V and the calculated ESC-T compound.

Index	F(1, 66)	p	h^2p	f^2	Shapiro–Wilk residuals (W; p)	d post Exp–Ctrl
ESC-T	28.89	<.001	.305	.438	W = .974; p = .159	1.22
VP	20.60	<.001	.238	.312	W = .980; p = .327	0.56
MT	10.64	.002	.139	.161	W = .969; p = .077	0.93
RF	5.84	.018	.081	.089	W = .976; p = .202	0.78
AND	5.05	.028	.071	.077	W = .891; p < .001 ^a	0.66
CV	3.69	.059	.053	.056	W = .980; p = .313	0.53

Note. Rows are ordered by descending η^2p (partial eta squared) to reflect the magnitude of the Group × Time interaction. F = interaction statistic from the factorial mixed model with Satterthwaite-approximated degrees of freedom. η^2p = partial eta squared. f^2 = Cohen's f^2 ($\eta^2p / [1 - \eta^2p]$). Shapiro–Wilk tests were applied to the studentized residuals of each model. d post Exp–Ctrl = Cohen's d for the between-group posttest difference calculated with the pooled SD of the two posttest cells. All d values exceed the pre-specified pedagogical relevance threshold of $d \geq 0.30$ (Vasilopoulos et al., 2023).

^a VE residuals did not meet the normality assumption ($W = .891$; $p < .001$). Heteroscedasticity-consistent standard errors (HC3) were applied; the Group × Time interaction remained compatible with a differential effect ($p = .037$ under HC3). This diagnostic caveat moderates interpretation of the VE finding.

The table consolidates the inferential parameters of the mixed factor model for each indicator,

including the F statistic of the Group × Time interaction with its degrees of freedom, the value

of p , partial η^2 , Cohen's f^2 , the result of the Shapiro-Wilk test on the model's residuals, and Cohen's d post-test between groups with their 95% CI. The organization by rows follows the order of presentation of the findings (from highest to lowest magnitude of the differential effect).

Sensitivity analysis

The sensitivity analysis incorporating sex (recorded in the pretest) as an additional covariate in the mixed factor model was applied to the six indicators. The $\times$ Time Group interaction retained its sign and statistical significance in all domains where it had reached the conventional threshold in the main model (MT, RF, VP, VE and calculated T-ESC), and the partial η^2 effect sizes varied by less than 5 % compared to the non-covariate model. In Verbal Comprehension, interaction remained at the marginal threshold ($p \approx .06$). The invariance of measurement between subgroups by sex could not be formally tested due to the coding limitations described in Methods; exploratory stratified analysis, restricted to cases with consistent coding, replicated the direction of effect in both subgroups for MT and RF, without allowing robust inference of three-way interaction. These results are interpreted as tentative and are reported in the sense recommended by the partial invariance methodology of the recent frontier (Massé et al., 2025).

Synthesis of the observed pattern

The set of findings describes a consistent pattern. The experimental group presented large-magnitude pre-post improvements in the six indicators (intra-group d between 0.91 and 2.94), while the control group showed no systematic change in five of the six ($|d| \leq 0.29$) and a significant decrease in Processing Speed ($d = -1.06$). The Group $\times$ Time interactions were significant for five of the six indicators, with magnitudes that in decreasing order corresponded to calculated T-ESC, Processing Speed, Working Memory, Fluid Reasoning and Visuospatial Ability; in Verbal Comprehension,

interaction was at the marginal threshold. The standardized differences between groups in the post-test were moderate to large (d between 0.53 and 1.22), above the pedagogical relevance threshold of $d \geq 0.30$ set a priori. The two critical limitations of the design—the unfavorable baseline non-equivalence to the experimental group in Processing Speed and T-ESC, and the impossibility of individual longitudinal matching—condition the causal magnitude attributable to the intervention, restrictions that the Discussion takes up in a specific way when evaluating each finding.

Discussion

The findings describe a non-homogeneous cognitive pattern: the functions most directly coupled with the sustained cognitive demand of the program —Working Memory and Fluid Reasoning— show the most solid differential improvements; Processing Speed exhibits the interaction of greater magnitude, but conditioned by a severe baseline non-equivalence that forces it to be read with caution; and domains that depend on cumulative exposure or specific training—Verbal Comprehension and Visuospatial Skill—respond partially. This selective profile is the central datum that the rest of this section puts in conversation with the literature, finding by finding, and where each specific limitation is discussed in its rightful place.

Working memory: the most solid finding

The differential improvement in Working Memory is the result of greater robustness of the study: significant Group $\times$ Time interaction with effect size close to the large range (partial $\eta^2 = .139$), superiority of the experimental group in the post-test with $d = 0.93$, and model residuals that meet the assumption of normality. The pattern is theoretically expected. In the formulation of Miyake et al. (2000), working memory is the function where the updating component operates most directly, and the program's curriculum dedicated two entire blocks to updating tasks coupled to motor response —relays with digits, dynamic sequences, arithmetic-motor double task—. Diamond and Ling (2016) argue that only

exercise that intentionally challenges updating, inhibition, or shifting produces consistent executive gain; The domain that responded most in this study is precisely the one that the theory predicts as the most sensitive.

The magnitude observed is in the upper range of recent evidence. The meta-analysis by Vasilopoulos et al. (2023), on 92 RCTs and 25,334 children, identifies working memory as one of the domains where children's physical activity exerts its most consistent effect, with average sizes between $d = 0.30$ and $d = 0.50$ when the intervention lacks explicit cognitive demand. The synthesis by Mavilidi et al. (2025), based on 171 chronic studies, raises this range when activity is accompanied by moderate-vigorous cognitive load. The $d = 0.93$ reported here, within a program with cognitive intentionality mapped session by session, is compatible with the upper pole of that distribution and with the quasi-experimental finding of Masini et al. (2023) on the I-MOVE program.

A specific limitation moderates the outcome. The baseline difference in Working Memory was unfavorable to the experimental difference with moderate magnitude ($d = -0.71$; $p = .045$), implying that part of the differential change can be attributed to regression towards the mean. The stability of the control between pretest and posttest ($d = -0.16$) attenuates this concern—if the effect were mostly regression, both groups should have moved closer to the center, which did not occur in the control—but it does not eliminate it. The proper interpretation is that the observed effect is compatible with a real differential benefit, with an additive component of regression that only a randomized design could disaggregate.

Fluid Reasoning: The Theoretically Most Informative Finding

The differential increase in Fluid Reasoning ($F(1, 66) = 5.84$; $p = .018$; partial $\eta^2 = .081$; d posttest = 0.78) is secondary in magnitude compared to Working Memory, but conceptually justifies the architecture of the program. Fluent reasoning is

the least expected domain to be modified in nine weeks: the classic literature of Diamond (2013) and the meta-analysis of Vasilopoulos et al. (2023) document that the effect of exercise on abstract reasoning is notably lower than on working memory, except when the intervention incorporates integrated cognitive demand. The fact that the experimental group outperformed the control with $d = 0.78$ (medium-large range) is the strongest indication that the program operated through a cognitive channel and not only by general physiological improvement.

Two mechanisms compete to explain the result. The first is that the cognitive component of the program (dual-demand tasks, intentional motor planning, circuits with rule change in the middle of the activity) directly exercised shifting operations and abstract reasoning: the blocks of weeks 5 to 9 imposed alternating rules and arithmetic-motor combinations, a type of demand that González-del-Castillo and Barbero-Alcocer (2025) identify as a necessary condition to move the domain in the primary school slot. The second is indirect transfer from improvement in Working Memory: the unity-diversity model of Miyake et al. (2000) predicts that sustained updating functions as an executive infrastructure for fluid reasoning. The data do not allow us to disaggregate both channels—a factorial design that dissociates cognitive demand and physical intensity would be required—but both are consistent with the observed direction, a finding supported by the synthesis of Mello et al. (2025) in recent interventions with children 6–10 years of age.

Baseline comparability in this domain was acceptable ($d = -0.40$; $p = .256$), which reduces, but does not eliminate, the concern about regression to the mean. The specific limitation here is not comparability but duration: nine weeks is a short window to speak of consolidated transfer to reasoning, and the absence of post-intervention follow-up prevents knowing whether the gain is sustained at 8 or 12 weeks, a restriction that recent literature underlines as critical (Reyes-Amigo et al., 2025).

Processing Speed: The Interpretively Delicate Finding

The Group $\times$ Time interaction in Processing Speed is the largest in the study ($F(1, 66) = 20.60$; $p < .001$; partial $\eta^2 = .238$), but its causal interpretation is the most compromised. Three elements coexist in the data. First, the experimental group improved substantially between pre-test and post-test ($\Delta = +4.07$; d intra-group = 1.13). Second, the control group regressed by comparable magnitude ($\Delta = -3.88$; $d = -1.06$; $p = .003$). Third, the control started substantially higher than the experimental one at baseline (24.93 vs. 19.40; baseline difference $d = -2.01$; $p < .001$)—the most severe baseline asymmetry in the study. The magnitude of the interaction is inflated by the opposite direction of change in both groups and by this extreme baseline non-equivalence.

An honest reading of the result requires recognizing two distinct contributions. On the one hand, part of the movement of both cells towards the center of the distribution is compatible with regression to the mean – the control, starting from an atypically high level, tended to approach the mean in a second measurement regardless of the intervention. This does not invalidate the finding, but requires not reading the entire partial .238 η^2 as a causal effect of the intervention. On the other hand, the improvement of the experimental does exceed the pedagogical threshold (intra-group $d = 1.13$) and is consistent with the pattern of the rest of the domains. A parsimonious interpretation is that the intervention produced a genuine gain, possibly of medium magnitude, by indirect transfer—the program did not include specific pure speed training, but inhibition and reaction demands coupled with bilateral coordination—while the control fluctuation responded in part to regression toward the mean. This reading is compatible with the general warning of Vasilopoulos et al. (2023) about quasi-experimental studies with severe baseline non-equivalence, and also constitutes the central limitation of the design: the impossibility of individual randomization allowed intact

assignment per classroom to generate an asymmetry that a controlled trial would avoid by construction.

Verbal Comprehension, Visuospatial Ability, and Composite ESC-T: The Specificity of the Effect

The remaining three findings define the specificity of the program by contrast. In Verbal Comprehension, the intra-group improvement of the experimental was substantial ($d = 0.91$; $p = .008$), but the Group $\times$ Time interaction was at the conventional threshold ($p = .059$; partial $\eta^2 = .053$) and the difference between groups in the post-test was moderate ($d = 0.53$). The pattern is theoretically consistent: verbal comprehension depends substantially on cumulative language exposure, and nine weeks of physical activity, even with cognitive demand, does not constitute a direct channel for such exposure (Diamond, 2013). The fact that the experimental has improved descriptively and control is not consistent with an indirect effect mediated by the improvement in attention and updating during the administration of the WISC-V, prediction of the unit-diversity model when an improvement in updating facilitates but does not directly train lexical-semantic retrieval.

In Visuospatial Skill, both groups improved, but the experimental group improved by substantially greater magnitude (d intra-group = 2.94 vs. 0.62). The interaction was significant in the ordinary model ($F = 5.05$; $p = .028$), although the residuals did not comply with normality ($W = .891$; $p < .001$). With robust standard error HC3, the interaction remained compatible with differential effect ($p = .037$), so the finding is reported as robust to correction but with diagnostic warning. The sheer magnitude is the largest in the study, but some of it can be attributed to familiarity with the Cube Design and Visual Puzzle subtests in a second administration. The intervention did not include specific visuospatial training; The result is compatible with small-moderate transfer, not with targeted training.

The ESC-T composite integrates all five indices and shows the broadest interaction of the study

(partial $\eta^2 = .305$; post-test $d = 1.22$). Three conditions apply. First, ESC-T is an arithmetic sum, not a psychometric metric equivalent to the normative Total IQ; Its interpretation is that of an aggregate indicator of executive functioning, not that of an IQ. Second, the baseline non-equivalence in the compound was marked ($d = -1.38$; $p = .0007$), reflecting the asymmetry in Processing Speed already discussed. Third, the magnitude is consistent with the sum of positive differential effects in five domains, but no single domain reaches, in terms of interpretative robustness, the robustness of Working Memory and Fluent Reasoning. The ESC-T should be read as a numerical synthesis of the pattern—useful for visualizing effect convergence—and not as additional independent evidence.

Contextual contribution: school programs in low-income intermediate municipalities

Beyond the findings by domain, this study provides evidence from a context underrepresented in the international literature. Recent meta-analyses (Mavilidi et al., 2025; Vasilopoulos et al., 2023) and reviews of active break programs (Reyes-Amigo et al., 2025; Maiztegi-Kortabarria et al., 2024) are dominated by studies of urban schools in high-income countries, with consolidated sports infrastructure and intensity monitoring. The application in an intermediate municipality in northeastern Colombia—with marked infrastructure restrictions—expands the empirical base of the theory towards environments where low-cost replicability is an operational condition. The fact that the program can be implemented with minimal material resources and produces a pattern compatible with cognitive demand theory (Diamond & Ling, 2016) is, in itself, a contribution to the ecological generalization of the field (Powell-Wiley et al., 2024; Sortwell et al., 2026).

The specificity for the 1460-1479 age group also contributes to the state of the art. Recent meta-analyses aggregate wide age ranges and dilute sensitivity to periods of plasticity, and the early age of primary schooling—where the consolidation of basic executive functions is at its

highest (Diamond, 2013)—is often underrepresented (Mello et al., 2025; González-del-Castillo & Barbero-Alcocer, 2025). The evidence provided in this age range, with cognitive intentionality mapped session by session, serves as an anchored empirical case for future discussions on dose, duration, and optimal cognitive load.

Additional limitations

Three residual limitations deserve joint mention because they affect inference transversally. First, the impossibility of longitudinal individual matching forced an unpaired mixed factor analysis on 70 records, instead of the idealized Mixed Linear Model with random intercept per participant; Although the direction and magnitude of the effects are interpretable, the causal attribution remains in the range of the preliminary evidence compatible with the intervention effect and not in that of the definitive causal estimate. Second, the inconsistency of sex coding between pretest and posttest prevented formal testing of measurement invariance between subgroups; sensitivity analysis with sex as a covariate preserved the significance and magnitude of the interactions for the five domains where the main model had reached them, suggesting that the findings are not guided by sex confusion, but the possibility of unmeasured partial invariance remains open and is reported in the sense recommended by Massé et al. (2025). Third, program intensity was monitored by structured observation of signs of exertion, not by objective heart rate measurement; the effective dose delivered may have varied between sessions and between participants in a magnitude that neither Bouzid et al. (2025) nor Ceylan et al. (2023) consider trivial when discussing the modulation of neuroplastic mechanisms such as BDNF. All three constraints are inherent in quasi-experimental design applied to field conditions and are recognized as such, without claiming that a single study will solve what all the literature in the field still negotiates.

Conclusion

A nine-week school physical activity program with intentional cognitive demands mapped session by session produced, in children aged 6 to 8 years from an intermediate municipality in northeastern Colombia, a pattern of differential improvement consistent with the cognitive demand theory: Working Memory responded most robustly (d posttest = 0.93), followed by Fluid Reasoning (d = 0.78). with minor and nuanced gains in Processing Speed, Visuospatial Ability and Verbal Comprehension. The effects

exceeded the pedagogical threshold set a priori ($d \geq 0.30$) in all domains and were maintained when controlling for sex in sensitivity, although their causal attribution is conditioned by the lack of baseline equivalence and the impossibility of individual longitudinal pairing. Active breaks with cognitive intentionality, replicable with minimal resources, can contribute to the strengthening of basic executive functions during the 6 to 8 year window: it is cognitive intentionality, and not physical activity itself, that explains the magnitude and specificity of the observed pattern.

References

- Bachouri-Muniesa, H., Verloigne, M., Zaragoza Casterad, J., & Asún-Dieste, S. (2025). Enabling sustainability through co-adaptation: A process evaluation of a school-based physical activity program. *Health Promotion International*, 40(1), Article daaf224. <https://doi.org/10.1093/heapro/daaf224>
- Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, 37(3), 379–384. <https://doi.org/10.3758/BF03192707>
- Bouزيد, M. S., Melki, H., Racil, G., Hawani, A., Hammoudi, S., Rezgani, Y., & Padulo, J. (2025). The influence of time of day on disruptive behaviours in middle school students during physical education classes. *Frontiers in Sports and Active Living*, 7, Article 1546436. <https://doi.org/10.3389/fspor.2025.1546436>
- Cañete-Massé, C., Krieger, V., Perú-Cebollero, M., Amador-Campos, J. A., & Guàrdia-Olmos, J. (2025). Measurement invariance and cross-linguistic validation of the PSS-4 in university context: Multidimensional analysis and associations with psychological and behavioral outcomes. *Frontiers in Psychology*, 16, Article 1648070. <https://doi.org/10.3389/fpsyg.2025.1648070>
- González Ortiz, J. A. (2017). Influence of physical activity on executive functions in primary school children. *Colombian Journal of Educational Psychology*, 9(1), 33–47.
- Groß, J., & Möller, A. (2025). Effect size estimation in linear mixed models. *METRION*, 83, 353–363. <https://doi.org/10.1007/s41460-1479-00295-w>
- Educational Institution Colegio Técnico Custodio García Rovira. (2021). *[Insert Title of the Institutional Pedagogical Report]* [Unpublished Institutional Pedagogical Report]. Málaga, Santander, Colombia.
- Korcz, A., Krzysztozek, J., Bojkowski, Ł., Koszałka-Silska, A., Khorkova, M., Gomołysek, A., & Bronikowski, M. (2024). The effects of the “active before school” programme on the academic skills of 8–9-year-old children: A physically and cognitively engaging intervention. *Frontiers in Public Health*, 12, Article 1402901. <https://doi.org/10.3389/fpubh.2024.1402901>
- Maiztegi-Kortabarria, J., Arribas-Galarraga, S., Luis-de Cos, I., Espoz-Lazo, S., & Valdivia-Moral, P. (2024). Effect of an active break intervention on attention, concentration, academic performance, and self-concept in compulsory secondary education. *European Journal of Investigation in Health, Psychology and Education*, 14(3), 447–462. <https://doi.org/10.3390/ejihpe14030030>
- Masini, A., Marini, S., Ceciliani, A., Barone, G., Lanari, M., Gori, D., Bragonzoni, L., Toselli, S., Stagni, R., Bisi, M., Sansavini, A., Tessari, A., & Dallolio, L. (2023). The effects of an active breaks intervention on physical and cognitive performance: Results from the I-MOVE study. *Journal of Public Health*. <https://doi.org/10.1093/pubmed/fdad102>
- Mavilidi, M. F., Vazou, S., Lubans, D. R., Robinson, K., Woods, A. J., Benzing, V., Anzeneder, S., Owen, K. B., Álvarez-Bueno, C., Wade, L., Burley, J., Thomas, G., Okely, A. D., & Pesce, C. (2025). How physical activity context relates to cognition across the lifespan: A systematic review and meta-analysis. *Psychological Bulletin*, 151(5), 544–579. <https://doi.org/10.1037/bul0000478>

- Mello, J. B., Costa, R. R., da Silva, F. F., & Cristi-Montero, C. (2025). School ACTIVE, brain active: A meta-analysis and meta-regression on chronic school physical activity effects on cognitive performance in children and adolescents. *Educational Research Review*, 42, Article 100658. <https://doi.org/10.1016/j.edurev.2024.100658>
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49–100. <https://doi.org/10.1006/cogp.1999.0734>
- Powell-Wiley, T. M., Martinez, M. F., Heneghan, J., & Lee, B. Y. (2024). Health and economic value of eliminating socioeconomic disparities in US youth physical activity. *JAMA Health Forum*, 5(4), Article e240088. <https://doi.org/10.1001/jamahealthforum.2024.0088>
- Rangel, M., & Duarte, J. (2022). Implementation of a physical stimulation program to improve care for primary school children in rural contexts. *Rural Education Journal*, 6(1), 19–34.
- Reyes-Amigo, T., Salinas-Gallardo, G., Mendoza, E., Ovalle-Fernández, C., Ibarra-Mora, J., Gómez-Álvarez, N., Carrasco-Beltrán, H., Páez-Herrera, J., Hurtado-Almonacid, J., Yañez-Sepúlveda, R., Zapata-Lamana, R., Sepúlveda-Figueroa, F., Olivares-Arancibia, J., & Mota, J. (2025). Effectiveness of school-based active breaks on classroom behavior, executive functions and physical fitness in children and adolescent: A systematic review. *Frontiers in Public Health*, 13, Article 1469998. <https://doi.org/10.3389/fpubh.2025.1469998>
- Rico-González, M., González-Devesa, D., Gómez-Carmona, C. D., & Moreno-Villanueva, A. (2025). Exercise as modulator of brain-derived neurotrophic factor (BDNF) in children: A systematic review of randomized controlled trials. *Life*, 15(7), Article 1147. <https://doi.org/10.3390/life15071147>
- Song, T. F., Chi, L., Chu, C. H., Chen, F. T., Zhou, C., & Chang, Y. K. (2023). Do acute and chronic physical activity interventions affect the cognitive function of preschool children? A systematic review and meta-analysis. *Psychology of Sport and Exercise*, 67, Article 102419. <https://doi.org/10.1016/j.psychsport.2023.102419>
- Sortwell, A., Diezmann, C. M., Ramirez-Campillo, R., & Murphy, A. J. (2026). Tackling paediatric dynapenia: AI-guided neuromuscular active break model for early-year primary school students. *Applied Sciences*, 16(8), Article 3654. <https://doi.org/10.3390/app16083654>
- Vasilopoulos, F., Jeffrey, H., Wu, Y., & Dumontheil, I. (2023). Multi-level meta-analysis of physical activity interventions during childhood: Effects on cognitive and academic outcomes. *Educational Psychology Review*, 35(2), Article 47. <https://doi.org/10.1007/s11460-1479-01460-1479>
- Wechsler, D. (2014). *WISC-V: Technical and interpretive manual*. Pearson.