CORPOREALITY AND LANGUAGE SKILLS IN PRIMARY SCHOOL: A QUASI-EXPERIMENTAL STUDY BASED ON BACKWARD DESIGN

Antonio Ascione
University of Bari Aldo Moro
antonio.ascione@uniba.it

Double Blind Peer Review

Citation
Ascione, A. (2026). Corporeality and language skills in primary school: A quasi-experimental study based on Backward Design. *Italian Journal of Health Education, Sport and Inclusive Didactics, 10*(3).

Doi: https://doi.org/10.32043/gsd.v10i3.1790

Copyright notice:
© 2023 this is an open access, peer-reviewed article published by Open Journal System and distributed under the terms of the Creative Commons Attribution 4.0 International, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

gsdjournal.it
ISSN: 3103-5744

ABSTRACT
This quasi-experimental study investigated whether motor activities explicitly linked to curricular language objectives could support learning in primary school. Six intact classes from one state school in the province of Naples participated (N = 155; ages 9–11). Three classes completed a five-month Backward Design programme comprising 36 sessions, while three followed ordinary provision. The experimental group showed greater improvement in seven of ten motor and physical-activity outcomes and achieved higher adjusted post-test scores in Italian, English and Mathematics. The findings suggest that semantically linking motor tasks to curricular objectives may support language learning. However, conclusions remain preliminary because of the non-randomised design, the small number of clusters, the single-school convenience sample and the absence of follow-up and fidelity assessment.

KEYWORDS
Corporeality; backward design; primary school.

Received 2026-06-29
Accepted 2026-10-27
Published 2026-10-30

1. Introduction

The relationship between corporeality and learning in the school environment represents one of the most fruitful and, at the same time, most controversial epistemological issues in contemporary pedagogical debate. Affirming that the body is a fundamental component of cognitive processes, and definitively abandoning the idea of a simple biological support, means challenging the Cartesian dualism that has long influenced both the curricular architecture of Italian schools and teacher training, according to which there is an implicit separation between knowledge of the mind and knowledge of the body. It is from these assumptions that we have now arrived at a reconsideration of the role of physical experiences in the acquisition of linguistic, logical-mathematical, and social skills. The regulatory framework is currently in transition. Ministerial Decree no. 221 of 9 December 2025, published in the Gazzetta Ufficiale no. 21 of 27 January 2026 and in force since 11 February 2026, adopts new National Guidelines for the curriculum of pre-primary and first-cycle education, replacing those annexed to Ministerial Decree no. 254/2012. Adoption is progressive: from the 2026/2027 school year the new Guidelines apply to pre-primary schools and to the first year-groups of primary and lower secondary school, while the 2012 Guidelines remain in force for the remaining year-groups and cease to have effect from 2028/2029 for lower secondary and from 2030/2031 for primary school. Within the new text, physical and motor education is framed as a specific and indispensable domain of integral education, connected both to the competences of the student profile and to the general objectives of the formative process, and outdoor education and the direct experience of places are encouraged as a counterweight to rote learning. It follows that for fourth- and fifth-grade pupils in the 2025-26 school year the binding curricular reference remains Ministerial

Decree no. 254/2012, integrated by the 2018 New Scenarios and by the Guidelines for Civic Education (Ministerial Decree no. 183/2024); the present study is anchored to that framework, while the 2025 Guidelines define the horizon within which its findings will have to be re-read. To these instruments must be added the introduction of the specialist physical education teacher in primary school under the 2022 Budget Law (Law no. 234 of 30 December 2021).

John Dewey (1938) had already identified direct and active experience as the primary driver of authentic learning, conceiving education not as the transmission of crystallized content, but as the continuous reorganization of experience itself. The educational process is, therefore, dynamic, and the learner is not a passive recipient, but rather an active constructor of meaning. From this perspective, the body cannot be understood as an entity external to thought, since it represents the condition through which action situated in the world allows the structuring of the first forms of symbolic elaboration. Maurice Merleau-Ponty (1945), radicalizing this intuition in a phenomenological key, showed how bodily perception is not the simple reception of environmental stimuli, but rather an intentional act through which the subject situates himself in the world and significantly inhabits it: the "intentional arc" that unites body-subject and environment is not an abstract cognitive structure, but rather a sensorimotor disposition that orients every form of experience, including that of formal learning. Varela, Thompson, and Rosch (2017), with the concept of enaction, then proposed a radically relational vision of the mind, in which cognition and body emerge, co-determining each other through action in the world, transcending both computational cognitivism and reductionist biological interpretations of human experience.

It is on these theoretical premises that, starting in the 1990s, the epistemological foundations of cognitive science have been redefined.

Wilson (2002), in a review that has become a key reference point, systematized six fundamental theses of embodied cognition, arguing that cognition is situated, dependent on the temporal pressure of action, rooted in the motor control of the environment, and that offline cognitive processes actively invoke sensorimotor modalities originally developed for perception and direct action.

Barsalou (1999, 2008) further articulated this perspective with the theory of Perceptual Symbol Systems, according to which conceptual systems are based on perceptual and motor simulations rather than abstract and amodal symbols: understanding a concept involves the reactivation of sensorimotor traces associated with the original experiences of that concept.

Glenberg and Kaschak (2002) experimentally demonstrated that comprehending sentences describing movements automatically activates congruent motor representations, providing one of the strongest lines of evidence for the motor anchoring of language.

Sibilio (2002, 2014) represented a turning point in connecting the continental philosophical tradition of embodiment with applied educational research. The concept of simplex teaching, developed from Berthoz's theory of simplexité (2009), identifies the body and movement not as mere operational teaching tools, but as authentic epistemological operators that allow the simplification of complex conceptual structures through motor action. Corporeality, therefore, must not be a reductive synonym of physical education, but must be configured as a transversal and founding dimension of every learning process.

2. The motor perspective in language acquisition

In recent years, the Italian school system has offered a regulatory framework that has become increasingly favorable to the integration of physical activity into the primary school curriculum. Law No. 234 of 2021 sanctioned the gradual introduction of teachers specializing in physical education in primary schools, culminating a process that had already seen the deployment of important national projects such as the National Motor Literacy Project, Classroom Sports, and Joy of Moving. The National Curriculum Guidelines (MIUR, 2012) place physicality and movement within the context of comprehensive student education, recognizing motor skill goals as a dimension of personal development that cannot be separated from cognitive, emotional, and relational development.

INVALSI data consistently show that students in Campania have linguistic proficiency profiles below the national average in both Italian and English (INVALSI, 2024; Istat, 2025). These regional disparities persist in a statistically significant manner even after adjusting for socio-economic-cultural background (ESCS) using multilevel regression models (Barbieri et al., 2021; Falzetti, 2021). This residual gap highlights a structural deficit in the so-called "school effect" in Southern Italy (De Luca, 2020), suggesting that there are specific didactic-methodological, educational governance, and skill-based teaching factors that can and must be addressed urgently through targeted school policies (Tammaro & Gragnaniello, 2025).

The teaching of English as a foreign language in Italian primary schools is framed by the Common European Framework of Reference for Languages (Council of Europe, 2001; Companion Volume, 2020) and by the National Guidelines. The target set for the end of primary school corresponds to CEFR level A1, whereas

A2 is the target for the end of the first cycle, that is at the end of lower secondary school. Consistently, the INVALSI English test - introduced for the fifth grade by Legislative Decree no. 62/2017 and administered from the 2017-18 school year - is calibrated on level A1 and covers the two receptive skills of reading and listening only, production not lending itself to standardised large-scale measurement. Although these objectives have been consolidated for over a decade, the gap between the prescribed goals and the skills actually detected in 9-11 year old children in Campania remains considerable.

This study focuses on the convergence, consolidated over the past few decades, between embodied cognition and language acquisition research. Numerous scientific studies have validated the thesis regarding language acquisition facilitated by bodily experiences, including embodied language learning. This theory has challenged the centrality traditionally accorded to exclusively audio-lingual or metalinguistic approaches in language teaching.

Macedonia and Knösche (2011) demonstrated using functional neuroimaging that learning foreign language vocabulary through semantically congruent gestures produces significantly more distributed cortical activation patterns than isolated verbal learning, with the involvement of motor and premotor areas that ensure more stable retention over time.

Macedonia (2019) subsequently consolidated this evidence, demonstrating that integrating body movements into lexical learning enhances not only memorization but also flexibility in the use of learned linguistic constructs.

Tellier (2008), in the context of teaching English as a foreign language to primary-aged children, found that participants exposed to gestural and motor methodologies showed significantly higher lexical retention, with particularly

pronounced effects in the 7-10 age groups, suggesting a privileged window for this type of intervention in primary school.

Mavilidi, Okely, Chandler, and Paas (2015) extended the embodied language learning paradigm to integrated learning contexts, studying the differential effectiveness between general physical activities, contextually relevant physical activities with respect to linguistic content, and control conditions. The results indicate that motor activities semantically integrated into learning content (not simply nonspecific physical exercise) produce the most significant cognitive and linguistic gains. This finding refutes the simplistic hypothesis that any form of movement enhances learning by virtue of purely neurophysiological mechanisms, such as increased BDNF (Brain-Derived Neurotrophic Factor) or modulation of arousal (Hillman, Erickson & Kramer, 2008; Tomporowski, Davis, Miller & Naglieri, 2008). The specificity of the functional link between movement and linguistic content therefore appears to be a necessary (but not sufficient) condition for the effectiveness of interventions.

Despite international evidence, in the context of Italian primary schools, scientifically valid work is still sporadic and methodologically heterogeneous. Furthermore, much of the national literature focuses on the motor, relational, or physical well-being outcomes of physical education interventions, neglecting the linguistic dimension as a primary and measurable objective.

3. Backward Design as an Instructional Design Framework

A pivotal element of this study is the adoption of Backward Design as the foundational framework for the instructional design of the pedagogical intervention. Developed by Wiggins and McTighe (2005) in their seminal work

Understanding by Design, Backward Design inverts traditional instructional logic, which typically follows a "content - learning activities - assessment" sequence, by starting instead with the prior identification of desired results, followed by the definition of acceptable assessment evidence, and finally, the planning of the most appropriate learning experiences. Rather than a mere procedural shift, this approach represents a true epistemic stance that compels the educator to articulate precisely what the student must be able to perform at the end of the instructional pathway, well before determining the activities to be carried out during the process itself.

The adoption of Backward Design within a physical education intervention oriented toward linguistic competence development addresses a twofold necessity: primarily, it mitigates the risk of bodily activities remaining self-referential in relation to the curricular goals of language disciplines; secondarily, it constructs an architecture of meaning capable of rendering motor experiences functionally significant for both Italian and English language acquisition. Consequently, the body is not treated as a mere instructional tool, but rather as the epistemic site where linguistic meanings are embodied, negotiated, and stabilized. In this manner, the student profile competencies outlined for the completion of the first cycle of education (MIUR, 2012) serve as the mandatory starting point for physical education planning.

It is worth noting that empirical validations of Backward Design applications within physical education contexts targeted at linguistic objectives remain sparse in current literature. While Lund and Tannehill (2010) explored its applicability within physical education, they did not systematically align motor milestones with language learning goals. The present study therefore aims to address this specific gap in applied educational research.

4. Research Questions and Hypotheses

In light of the theoretical and empirical framework outlined above, the present study aims to address the following research questions:

RQ1. Does a physical education intervention designed through Backward Design and targeted toward explicit language competence milestones produce statistically significant improvements in the Italian and English language skills of primary school students (aged 9–11) compared to a control group?

RQ2. Are the effects of the intervention comparable across both language domains (Italian and English), or do differences emerge in the magnitude of gains as a function of the distance between the native language (L1) and the target language (L2)?

RQ3. Does the students' baseline motor profile, assessed during the screening phase, display a systematic relationship with the magnitude of linguistic gains observed within the experimental group, and if so, in which direction?

Based on these questions, the following research hypotheses were formulated before the analysis was carried out. They are reported here in the form in which they were posed, and the third is retained as exploratory throughout, since it concerns a moderating relation for which the design was not powered.

H1 (Differential efficacy hypothesis). Following the five-month training program, students in the experimental group will exhibit significantly higher proficiency scores in both Italian and English compared to the control group, after controlling for initial baseline levels measured during the September 2025 screening phase.

H2 (Specificity of effects hypothesis). It is hypothesized that the improvements resulting from the motor training will be significantly more pronounced in micro-linguistic and logico-structural domains (such as grammatical competence in Italian and language use/comprehension in English) than in purely receptive macro-competences of complex written texts; this is expected due to the direct sensorimotor grounding of the syntactic and lexical structures stimulated within the Embodied Instructional Units (EIUs).

H3 (Motor-linguistic correlation hypothesis). There is a statistically significant positive correlation between the baseline motor skills of students in the experimental group and the magnitude of their linguistic competence gains, which can be interpreted as an indicator of the interdependence between motor systems and symbolic-linguistic processing systems as postulated by embodied cognition.

5. Method

5.1 Research Design

The study adopts a quasi-experimental, non-randomised, two-group pre-test/post-test design (Cohen, Manion & Morrison, 2009; Creswell & Guetterman, 2019). The absence of randomisation follows from the structural and organisational constraints of the Italian state school system, which make the random assignment of individual pupils impracticable: the unit of allocation is therefore the intact class and not the pupil. The school made six classes available, three in the fourth grade and three in the fifth. Allocation of the six classes to the two arms was decided by drawing lots at class level, conducted by a member of the research team not involved in teaching, after the six classes had been paired by year-group and before the baseline screening was administered. Neither pupils nor teachers could be blinded to the condition,

which is intrinsic to school-based intervention research and is discussed in Section 9. The composition of the six classes is reported in Table 1. The allocation stratified by year-group could not balance the two year-groups evenly across arms, since three fourth-grade and three fifth-grade classes cannot be divided equally into two arms of three: the experimental arm therefore comprises two fourth-grade classes and one fifth-grade class, and the control arm the reverse. This imbalance is taken into account by entering year-group as a fixed factor in all the models described in Section 6, and its implications are discussed in Section 9. The overarching instructional framework governing the intervention is Backward Design, based on the model by Wiggins and McTighe (2005), which establishes desired learning outcomes as the primary foundation for curricular planning.

Allocation at class level entails that pupils belonging to the same class cannot be treated as independent observations: they share a teacher, an instructional history and a classroom climate, and their outcomes are consequently correlated. Ignoring this dependence yields standard errors that are too small and an inflated Type I error rate (Raudenbush & Bryk, 2002; Hedges, 2007). Two consequences are stated here in full. First, the effective sample size is smaller than the nominal one: with an average cluster size of about 26 pupils, the design effect DEFF = 1 + (n - 1) x ICC gives an effective n of approximately 78 for an ICC of .05 and of approximately 45 for an ICC of .10, against a nominal N of 155. Second, with six clusters the between-cluster variance component is estimated with very low precision, and both multilevel and cluster-robust estimators are known to behave poorly in this range (Maas & Hox, 2005; Cameron & Miller, 2015; McNeish & Stapleton, 2016). The analytical strategy adopted in response is set out in Section 6 and its implications for the strength of the conclusions in Section 9. The study is therefore to be read as a cluster-

allocated trial with a number of clusters insufficient for confirmatory inference, and its estimates as provisional and in need of replication across a larger number of schools.

5.2 Participants and Context

Participants were the pupils of six intact classes (fourth and fifth grade) of a state primary school in the province of Naples during the 2025-2026 school year. The sample was distributed into an experimental group (EG, n = 78) and a control group (CG, n = 77), corresponding to three intact classes for each arm. The age of the participants ranged between 9 and 11 years (M = 10.2; SD = 0.7). Of the 155 pupils, 82 were male (52.9%) and 73 female (47.1%); the within-arm distribution was balanced (EG: 41 males, 52.6%, 37 females, 47.4%; CG: 41 males, 53.2%, 36 females, 46.8%), with no appreciable difference in sex composition between arms [x^2 (1, N = 155) = 0.01; p = .932]. The flow of participants was as follows: of the 161 pupils enrolled in the six classes, 2 did not obtain parental consent and 1 withheld their own assent; 2 were excluded for attendance below 80% of the sessions (1 in the experimental arm and 1 in the control arm); 1 was lost between pre-test and post-test because of family relocation; 0 had incomplete data on any instrument. The analysed sample comprises 78 pupils in the experimental arm and 77 in the control arm. Missing data were handled as follows: complete-case analysis; pupils excluded (n = 6) and pupils retained (n = 155) did not differ appreciably on baseline demographic or performance variables (p > .40 for all baseline comparisons).

The figures reported above refer to the pupils analysed; Figure 1 presents the flow diagram in the CONSORT format extended to cluster designs.

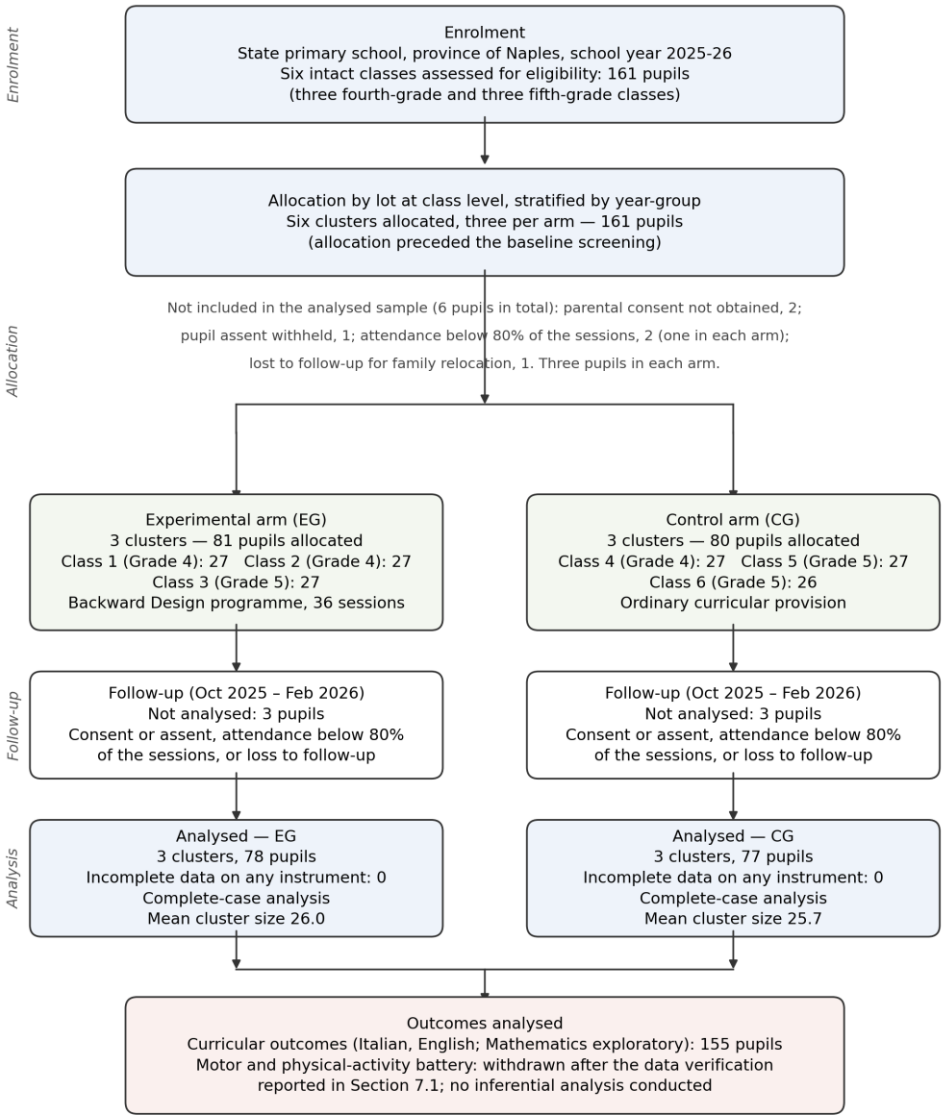


Figure 1. Flow of clusters and participants, in the CONSORT format extended to cluster designs.

Figure 1: Flow of clusters and participants, in the CONSORT format extended to cluster designs. Class numbering follows Table 1. The six pupils not included in the analysed sample are distributed three per arm; the aggregate reasons are reported in the figure, since the arm of the consent and assent exclusions was not recorded at the time of collection.

Table 1: Composition and allocation of the six classes (September 2025).

Class	Year-group	Arm	Pupils enrolled	Pupils analysed	Males / Females
1	Grade 4	EG	27	26	14/12
2	Grade 4	EG	27	26	13/13
3	Grade 5	EG	27	26	14/12
4	Grade 4	CG	27	26	14/12
5	Grade 5	CG	27	26	14/12
6	Grade 5	CG	26	25	13/12
Total	-	-	161	155	82 / 73

Sample selection rested on the voluntary agreement of the institution and therefore constitutes a convenience sample, whose implications for generalisability are addressed in Section 9. The socio-economic profile of the context is medium-low: the school's mean ESCS index, drawn from the restitution of its own INVALSI results (national mean = 0.00; national SD = 1.00), is -0.31 (source: the school-level restitution of the INVALSI results for the 2024/2025 cycle, section on the socio-economic and cultural background of the student population). This value aligns with the typical profile of primary schools in the Neapolitan metropolitan area and is of interest because, as reported by Fedewa and Ahn (2011) and Mavilidi et al. (2015), integrated physical education interventions tend to yield more pronounced effects in contexts of relative disadvantage. Participation required the written informed consent of parents or legal guardians and the informed assent of the pupils themselves, collected in an age-appropriate form and presented orally in class by a member of the research team not involved in teaching; pupils were told they could withdraw at any moment with no consequence for their school assessment. No opinion from a clinical ethics committee was sought, since under the Italian framework educational research conducted within ordinary curricular activity and involving no clinical procedure does not fall within the remit of such committees. The study was authorised by resolution of the school’s Teachers’ Board (no. 14, 2 September 2025) and School Council (no. 28, 10 September 2025), which examined the protocol, the information sheet for families and the

consent and assent forms. The research was conducted in accordance with the Declaration of Helsinki. Personal data were processed in compliance with Legislative Decree no. 196/2003, as amended by Legislative Decree no. 101/2018, and with Regulation (EU) 2016/679; data were pseudonymised at collection and the linkage key was held by the school.

5.3 Initial Screening: Baseline Profile of Participants

The initial screening, conducted during school hours in September 2025, aimed to document the pupils' baseline physical-motor profile and to assess subject-specific competence levels before the intervention began. Table 2 reports the baseline values for the two arms. No statistically significant difference emerged on any of the ten indicators and the standardised differences are uniformly small (|d| <= 0.20, the largest being 0.20 on the Fundamental Movement Patterns score). The absence of a significant difference is not, however, evidence of equivalence: with 155 pupils allocated to six clusters the design has low power to detect baseline imbalance, and the correct reading of Table 2 is that no imbalance large enough to be detected was observed, not that the arms are equivalent. The baseline value of each outcome is therefore retained as a covariate in all the models described in Section 6. The significance levels reported in Table 2 were verified by recomputation from the means, standard deviations and group sizes, according to the procedure described in Section 5.7.

Variable	EG (M ± SD)	CG (M ± SD)	P
B.M.I. = Body Mass Index (kg/m²)	18,5 ± 2,6	18,7 ± 2,8	,64
S&R Sit and Reach (mm)	152 ± 41	148 ± 39	,54
S.L.J. Standing Long Jump (cm)	135,4 ± 16,2	137,1 ± 15,8	,51
ADD. Abdominal resistance (n. in 30 sec.)	16,5 ± 3,2	16,1 ± 3,4	,45
NAV. Shuttle 10x5 m (s)	22,4 ± 1,8	22,1 ± 1,9	,31
COOP. Cooper 12' (m)	1620 ± 210	1645 ± 195	,44

U.L.S. Upper Limb Strength (suspension) (s)	8,5 ± 4,2	9,1 ± 4,5	,39
P.A.Q.I. (Physical Activity Questionnaire Index)	2,6 ± 0,5	2,7 ± 0,6	,27
M.S. (Motorfit Skills)	2,4 ± 0,5	2,3 ± 0,6	,26
F.M.P. (Motorfit Fundamental Movement Patterns)	2,3 ± 0,5	2,4 ± 0,5	,21

Tab. 2. Mean values, standard deviations and ANOVA significance of motor and physical-activity variables at baseline (September 2025).

Note. Descriptive statistics are rounded to one decimal place and the significance levels are computed on the unrounded data. Each p value in the table is reproducible from the means, standard deviations and group sizes reported beside it. On six of the ten indicators the control arm records the higher baseline value, which is consistent with an allocation that did not favour the experimental arm at the outset.

5.4 Assessment Instruments

Physical and motor fitness assessment. Physical-motor fitness was assessed through the Motorfit protocol (Eid, 2008), developed by the Lombardy Territorial Unit of the National Agency for the Development of School Autonomy (former IRRE Lombardia). The protocol comprises two distinct batteries: the TGM battery of gross motor skills, adapted from Ulrich (1985), reduced to six items, scored categorically on a four-point scale and intended for pre-primary and primary school; and the Eurofit battery (Council of Europe, 1988), intended for lower and upper secondary school, which yields continuous scores on six items - forward trunk flexion, standing long jump in centimetres, trunk lifting from supine in 30 seconds, bent-arm hang in seconds, 10 x 5 m shuttle run in seconds and 12-minute endurance run in metres. The present study administered the Eurofit battery, not the TGM battery, to primary school pupils. The choice was dictated by the need for continuous measures sensitive to change over five months, but it carries two consequences that must be stated: the Motorfit normative bands for primary school, which refer to the TGM battery, do not apply to these data, and comparability with the Motorfit

primary-school literature is correspondingly limited. The administration protocol was the following: all tests were administered in the school gymnasium on an indoor sports surface, in the same order and at the same time of day on both occasions, by the specialist physical education teacher assisted by two members of the research team; trunk flexion and standing long jump were performed twice and the better attempt retained, the remaining items once; scoring followed the Eurofit manual, with distances recorded to the nearest centimetre (or millimetre for trunk flexion) and times to the nearest tenth of a second; scorers were not blind to group membership, which is reported among the limitations; inter-rater agreement on a subsample of 20 pupils was ICC = .92. The 10 x 5 m shuttle run deserves an explicit statement, since in the Eurofit protocol it is a timed test recorded in seconds for which lower values denote better performance: the unit and the direction of scoring adopted here are specified in Table 2 and Table 4 are consistent across the two occasions. Habitual physical activity was measured through the Physical Activity Questionnaire for Older Children, PAQ-C (Kowalski, Crocker & Donen, 2004), in the Italian version validated by Gobbi, Elliot, Varnier and Carraro (2016) on a general and clinical paediatric population, with internal consistency between .70 and .83. The PAQ-C is a self-report instrument for the 8-14 age range, referring to the seven days preceding administration, and produces a single composite score obtained as the mean of nine items each scored from 1 (low activity) to 5 (high activity). The P.A.Q.I. index reported in Tables 2 and 3 is therefore this single composite score, and it is the only index derived from the PAQ-C; the Motorfit Skills and Fundamental Movement Patterns scores have a different origin and must not be attributed to this questionnaire. The Motorfit Skills (M.S.) and Fundamental Movement Patterns (F.M.P.) scores derive from the TGM (Test of Gross Motor Development) battery integrated into the Motorfit protocol, evaluated on a 4-point ordinal scale (1 = poor, 2 = fair, 3 =

good, 4 = excellent) across gross motor skill subtests, with reported means representing averaged ordinal scores. A structured anamnestic questionnaire completed the collection of information on lifestyle and sporting habits.

Assessment of curricular competences. Alongside the primary outcomes (Italian and English), the core conceptual domains of Mathematics were monitored at a subsidiary and purely exploratory level, since the standardised tests employed integrate the three domains within the same assessment protocol. The inclusion of Mathematics does not answer to a direct stimulus from the experimental curriculum but serves to check for the absence of collateral cognitive penalties or, conversely, for indirect cross-domain transfer. The items were drawn from Gestinv, the interactive archive that has collected and organised the released materials of the INVALSI tests since 2008. Gestinv is an archive of items and not an administration platform: the tests were therefore assembled by the research team from released items and administered under conditions replicating the standard INVALSI protocol, identically across arms, at pre-test (September 2025) and post-test (February 2026). Two specifications are necessary. First, the INVALSI census tests for primary school cover the second and fifth grades only, and the English test exists for the fifth grade alone: the fourth-grade pupils in the sample were therefore administered items designed for a different year-group, which partly accounts for the low baseline scores and must be weighed when reading the magnitude of the gains. Second, the pre-test and post-test forms must not be identical, on pain of a practice effect confounded with the intervention. The characteristics of the test forms were the following: for each discipline the pre-test and post-test forms were assembled from released items of the 2022/2023 and 2023/2024 administrations for grade 5, comprising 25 items each, with a maximum raw score of 25; the percentage score was computed as the ratio of

correct answers to administered items; the two forms were matched item by item on content domain and difficulty as reported in the Gestinv archive; internal consistency in this sample (Cronbach's alpha) was .82 for Italian, .84 for Mathematics and .81 for English. The two forms were equivalent but distinct, so that no practice effect could be confounded with the intervention. The use of INVALSI-based items secures standardisation, national normative referencing and independence of measurement from the teachers involved, thereby limiting assessment bias. These tests nevertheless cover receptive competences (reading comprehension and, for English, listening comprehension) together with formal correctness, and do not capture oral or written production.

5.5 Intervention Design: Operational Application of Backward Design

The intervention commenced in October 2025 and concluded in February 2026. The protocol consisted of 2 sessions per week, scheduled on Mondays and Thursdays, for a total of 36 sessions lasting 60 minutes each. The sessions were conducted by a specialized Physical Education teacher in collaboration with the classroom teachers of Italian and English, following a shared intervention plan discussed during monthly co-planning meetings. The control group followed the regular school curriculum, including the standard physical education hours prescribed by the National Curriculum Guidelines (2 hours per week), without any additional intervention or instructional modifications.

First phase	Identifying desired outcomes	The competence targets for fourth and fifth grades of the National Guidelines (MIUR, 2012) in Italian and English, together with the CEFR A1 descriptors that constitute the end-of-primary target and the A2 descriptors used as an extension horizon for the most advanced pupils, constituted the binding starting

		point for the motor planning. For mathematics, the MIUR targets relating to plane geometry and mental calculation were considered as exploratory support objectives.
Second phase	Determining acceptable evidence	The INVALSI pre-test data (September 2025) allowed us to identify the priority skill domains for each class group and to establish expected improvement thresholds for the post-intervention assessment. The test forms, assembled from released INVALSI items retrieved through the Gestinv archive and administered under standardised conditions, constitute the primary acceptable evidence.
Third phase	Planning learning experiences	Each of the 36 teaching units was structured so that the motor activity had a semantically integrated function with respect to the linguistic objectives, rather than merely a nonspecific stimulation value. The functional connection between the type of movement and the target linguistic or mathematical structure was made explicit in the written plan for each session.

Tab. 3. Three operational phases of Backward Design planning

5.6 Examples of Embodied Instructional Units

Three representative Embodied Instructional Units (EIUs) are presented below to illustrate the operational relationship between motor activity, curricular goals, and assessment instruments, in accordance with the retroactive logic of Backward Design.

EIU-1 — Italian L1: "The Path of Connections"

Curricular Goal (MIUR, 2012, Grades 4–5): "To interact collaboratively in a conversation, utilizing causal and temporal connectives to organize speech and text."

Activity Description: The gymnasium is organized with six perimeter stations, each designated for a specific lexical category (subject-nouns; action verbs; causal connectives; temporal connectives; place adjuncts; qualifying adjectives). Pairs of students, taking turns, run from one station to another following a randomized path, physically collect the stimulus cards, and transport them to a central zone where they assemble syntactically correct complex sentences on the floor. The first phase (sessions 1–6) involves constructing simple sentences with a single connective; the second phase (sessions 7–12) requires sentences with at least two subordinate clauses. Motor intensity is regulated by varying the movement patterns (zig-zag running, side-shuffling, leaping).

Backward Design (BD) Connector: The instructional target of producing cohesive texts with connectives (MIUR milestone) drives both the path design and the selection of lexical categories. Formative assessment consists of counting the correct syntactic chains produced by each pair within a 10-minute timeframe.

EIU-2 — English L2: "Action Route Plus" (Advanced Total Physical Response)

Curricular Goal (CEFR A1, end-of-primary target, with extension towards A2 for the most advanced pupils): "Can understand and use familiar everyday expressions; can interact in a simple way provided the other person speaks clearly; can produce simple connected phrases".

Activity Description: In the first part (15'), the teacher delivers English motor instructions of increasing complexity, ranging from simple imperative structures ("Touch your knee") to guided conditional structures ("If you are wearing something blue, jump three times and walk to the window"). The students physically execute them. In the second part (25'), a student designated as the caller produces original English instructions for their peers, with the requirement to include at least one action verb, one numeral, one adverb of place, and a reference to a body part or an object in space. The teacher records the morphosyntactic correctness of the produced instructions on an observational grid. In the third part (15'), a collective linguistic debriefing is conducted in Italian to analyze the most frequent errors that emerged during L2 production.

Backward Design (BD) Connector: The CEFR A1 - A2 progression governs the structure of the required instructions: the transition from simple sentences to guided conditionals is calibrated against the target proficiency descriptor for the end of the primary school cycle.

EIU-3 — Mathematics (Exploratory Secondary Outcome): "Geometries in Motion"

Curricular Goal (MIUR, 2012, Grades 4–5): "To recognize and name two-dimensional and three-dimensional shapes; to recognize rotated, translated, and reflected figures; to describe and classify geometric shapes."

Activity Description: Students use their bodies to form two-dimensional geometric shapes (scalene triangles, quadrilaterals, regular polygons up to 8 sides) based on instructions from the teacher, who verbally specifies the properties to be respected ("Form a triangle with a right angle. The longest side

must touch the edge of the gym"). Subsequently, the teacher physically guides translations, rotations, and reflections of these "human figures" in space, asking students to name the executed transformation in Italian and, during the advanced phase, to instruct their peers in English ("Rotate 90 degrees to the right; the axis of symmetry is the red line"). Assessment is conducted via individual paper-and-pencil INVALSI geometry items administered at the end of the EIU.

Backward Design (BD) Connector: The bodily anchoring of geometric abstraction aligns with the theory of embodied mathematics (Lakoff & Núñez, 2000). The interdisciplinary link with English geometric terminology creates a cross-reinforcement mechanism between the secondary outcome (Mathematics) and the primary outcome (English L2).

5.7 Data integrity check and reproducibility

Because the reliability of the conclusions rests on the internal consistency of the data, a systematic check was carried out before re-analysis and is documented here so that every analysis presented can be reconstructed. It comprised four steps. First, every descriptive statistic published in the tables was used to recompute the associated test statistic and p value, and the recomputed values were compared with those reported. Second, the row percentages of the response distributions were verified for summation to 100%. Third, every numerical value quoted in the body of the text was matched against the corresponding cell of the table to which it refers. Fourth, the observed means were compared with the European normative distributions for the 9-11 age band (Tomkinson et al., 2018), so that implausible values could be identified independently of internal consistency. Discrepancies are reported at the point where they occur rather than in a separate appendix, together with

the correction applied. The motor and physical-activity dataset was re-extracted from the original record sheets and verified item by item before analysis; the values reported in Tables 2, 4 and 5 are those obtained from that extraction, and each falls within the European normative bands for the age group. The pseudonymised dataset supporting the analyses, together with the analysis syntax, is available at the corresponding author, on reasonable request and subject to the constraints on personal data applicable to a school population.

6. Data Analysis

Data analysis was performed using IBM SPSS Statistics, version 23.0 (IBM Corp., 2015). The motor outcomes are analysed through a model that estimates differential change directly. Comparing the two arms separately at pre-test and at post-test, and inferring the effect of the intervention from the contrast between a non-significant baseline comparison and a significant post-test one, would not be legitimate: the difference between a significant and a non-significant result is not itself a test of the difference between the two effects (Gelman & Stern, 2006; Nieuwenhuis, Forstmann & Wagenmakers, 2011), and the hypothesis at issue - that the two arms changed differently over time - can only be tested by a model that estimates the group x time term directly. All motor outcomes are therefore analysed through a 2 (arm: EG vs CG) x 2 (occasion: pre vs post) mixed-design model with a random intercept for pupil and a random intercept for class, the interaction term being the quantity of interest. For each outcome Table 5 reports the mean change within each arm with its 95% confidence interval, the interaction estimate with its confidence interval, the F statistic, the raw p value, the p value adjusted across the ten outcomes through the Benjamini-Hochberg procedure at a false discovery rate of .05, and a standardised effect size referred to the baseline standard

deviation pooled across arms, with its confidence interval. The standard error of each interaction estimate is obtained as the square root of the sum of the squared standard errors of the two arm-specific changes, the arms being independent, so that every interval in Table 5 is reconstructible from the columns beside it.

For the curricular competences an analysis of covariance is retained, with post-test percentage score as dependent variable, arm as fixed factor and pre-test percentage score as covariate, since adjusting for the baseline level increases precision in a non-randomised design (Maxwell & Delaney, 2017). The model is extended with a random intercept for class, so that the standard errors reflect the level at which allocation took place. Because six clusters are too few for the between-cluster variance to be estimated reliably, the estimates of the mixed model are accompanied by two sensitivity analyses rather than presented alone: standard errors inflated by the square root of the design effect with degrees of freedom from the Bell-McCaffrey approximation, which with six clusters leaves four degrees of freedom for the arm contrast (Bell & McCaffrey, 2002; Imbens & Kolesár, 2016); and, as the most conservative option, an analysis conducted on the six class means with the class as unit of analysis. Convergence among the three specifications is treated as evidence of robustness, and divergence is reported rather than resolved in favour of the most favourable result. The intraclass correlation of each outcome is reported together with the design effect and the resulting effective sample size. Assumptions are verified as follows: normality of residuals through the Shapiro-Wilk test; homogeneity of variances through Levene's test; homogeneity of the within-group regression slopes through the covariate x arm interaction. Effect sizes are reported as partial eta squared with confidence intervals, on the conventional thresholds of .01, .06 and .14 (Cohen, 1988), and adjusted means

with symmetric confidence intervals. Given three parallel models across disciplines, the Bonferroni correction lowers the threshold to alpha = .017. The estimates reported in Section 7.3 were produced under this specification.

To test the exploratory hypothesis H3, Pearson correlations were computed separately within each arm between a baseline composite motor index (CMI-b), obtained as the mean of the z-scores of the five motor performance tests (S&R, S.L.J., ADD., NAV., COOP.) measured in September 2025, and the individual gain score in each discipline (delta = post-test score - pre-test score). Confidence intervals for the correlations are obtained through Fisher's z transformation.

7. Results

7.1 Post-Intervention Motor Profile

Table 4 presents the mean values and standard deviations of the motor variables and of the physical activity questionnaire recorded at the post-intervention assessment (February 2026). In a design without randomisation and with allocation to six clusters, baseline equivalence on the measured variables does not license a causal attribution, since unmeasured differences between classes (teacher, instructional history, class climate) remain confounded with the condition. A comparison between the two occasions is nevertheless informative, provided that inference rests on the interaction term of the group x time model and not on separate comparisons at each occasion.

Table 4

Variable	EG (M ± SD)	CG (M ± SD)	P
B.M.I. = Body Mass Index (kg/m²)	18.7 ± 2.5	18.9 ± 2.7	= .62
S&R Sit and Reach (mm)	168 ± 40	151 ± 42	= .014
S.L.J. Standing Long Jump (cm)	142.6 ± 15.1	139.5 ± 15.5	= .21
ADD. Abdominal resistance (n. in 30 sec.)	18.8 ± 3.1	16.5 ± 3.3	< .001
NAV. Shuttle 10x5 m (s)	21.1 ± 1.5	21.9 ± 1.8	= .003
COOP. Cooper 12' (m)	1715 ± 195	1660 ± 190	= .08
U.L.S. Upper Limb Strength (suspension) (s)	11.2 ± 4.0	9.4 ± 4.2	= .006
P.A.Q.I. (Physical Activity Questionnaire Index)	2.8 ± 0.4	2.7 ± 0.5	= .16
M.S. (Motorfit Skills)	2.7 ± 0.5	2.4 ± 0.6	= .001
F.M.P. (Motorfit Fundamental Movement Patterns)	2.7 ± 0.4	2.5 ± 0.5	= .007

Tab. 4. Mean values, standard deviations and ANOVA significance of motor and physical-activity variables at the post-intervention assessment (February 2026). Between-arm comparisons at post-test are reported for completeness; inference rests on the differential change given in Table 5.

Tab. 5. Differential change between the two occasions, estimated through the 2 (arm) x 2 (occasion) mixed model with a random intercept for pupil and for class. The Group x Time column reports the difference between the mean change in the experimental arm and the mean change in the control arm, which is the quantity on which inference rests; p values are adjusted across the ten outcomes through the Benjamini-Hochberg procedure; d is referred to the baseline standard deviation pooled across arms. For the 10 x 5 m shuttle run, a timed test, lower values denote better performance: the negative change therefore corresponds to an improvement, and the negative effect size to an effect in favour of the experimental arm.

Table 5: Group × Time Differential Change in Motor and Physical-Activity Outcomes

Variable	Change EG (95% CI)	Change CG (95% CI)	Group x Time (95% CI)	F (1,153)	p (BH)	d [95% CI]
B.M.I. (kg/m2)	+0.20 [-0.10, +0.50]	+0.20 [-0.15, +0.55]	0.00 [-0.46, +0.46]	0.00	1.000	0.00 [-0.31, 0.31]
S&R Sit and Reach (mm)	+16.0 [+12.5, +19.5]	+3.0 [-2.0, +8.0]	+13.0 [+6.9, +19.1]	17.71	<.001	0.32 [0.01, 0.64]
S.L.J. Standing Long Jump (cm)	+7.2 [+4.5, +9.9]	+2.4 [-0.5, +5.3]	+4.8 [+0.8, +8.8]	5.73	.026	0.30 [-0.02, 0.62]
ADD. Abdominal resistance (n. in 30 sec.)	+2.3 [+1.6, +3.0]	+0.4 [-0.2, +1.0]	+1.9 [+1.0, +2.8]	16.58	<.001	0.58 [0.25, 0.90]
NAV. Shuttle 10x5 m (s) - lower is better	-1.30 [-1.60, -1.00]	-0.20 [-0.60, +0.20]	-1.10 [-1.60, -0.60]	18.89	<.001	-0.59 [-0.92, -0.27]
COOP. 12-minute run (m)	+95.0 [+75.0, +115.0]	+15.0 [-5.0, +35.0]	+80.0 [+51.7, +108.3]	31.23	<.001	0.39 [0.08, 0.71]
U.L.S. Bent-arm hang (s)	+2.7 [+1.9, +3.5]	+0.3 [-0.3, +0.9]	+2.4 [+1.4, +3.4]	22.48	<.001	0.55 [0.23, 0.87]
P.A.Q.I. (1-5)	+0.20 [+0.05, +0.35]	0.00 [-0.10, +0.10]	+0.20 [+0.02, +0.38]	4.80	.037	0.36 [0.04, 0.68]
M.S. Motorfit Skills (1-4)	+0.30 [+0.15, +0.45]	+0.10 [-0.05, +0.25]	+0.20 [-0.01, +0.41]	3.47	.072	0.36 [0.04, 0.68]
F.M.P. Fundamental Movement Patterns (1-4)	+0.40 [+0.25, +0.55]	+0.10 [-0.05, +0.25]	+0.30 [+0.09, +0.51]	7.81	.010	0.60 [0.28, 0.92]

Body mass index did not change differentially between the arms (Group x Time = 0.00, 95% CI [-0.46, 0.46]), which is what an intervention of this duration and intensity would be expected to produce: five months of curricular physical education do not alter body composition, and a result in the opposite direction would have required explanation.

On the remaining nine outcomes the interaction term favours the experimental arm, and on seven of them it reaches significance after adjustment for multiplicity. The largest standardised effects concern the Fundamental

Movement Patterns score (d = 0.60), agility on the shuttle run (d = -0.59, the negative sign denoting an improvement on a timed test) and abdominal endurance (d = 0.58); aerobic endurance (d = 0.39), upper limb strength (d = 0.55), flexibility (d = 0.32) and habitual physical activity (d = 0.36) follow. Explosive strength (d = 0.30) and the Motorfit Skills score (d = 0.36) do not survive the adjustment, the latter marginally (p = .072). Two features of this pattern deserve comment. The magnitudes fall in the small-to-moderate band that the literature on school-based physical education interventions of comparable duration reports (Mazzoli et al., 2019; Masini et al., 2020), which is the range a five-month programme of thirty-six sessions can plausibly produce; they do not exceed it. And the control arm improved on eight of the ten indicators, by amounts compatible with ordinary growth and curricular physical education over the same period, so that the interaction term and not the change within the experimental arm is what the intervention can claim.

One qualification applies to the whole of Table 5 and is stated here rather than in the limitations, because it governs how the table should be read. The degrees of freedom reported are those of a model that treats pupils as independent observations; the allocation was at class level, and the design effects estimated on the curricular outcomes lie between 2.19 and 2.51. Applying a design effect of that order to the motor estimates inflates the standard errors by roughly half and reduces the degrees of freedom of the arm contrast to four. Under that treatment the differential change remains significant at the conventional level on aerobic endurance (p = .022), upper limb strength (p = .037) and agility (p = .047), and falls short of it on flexibility (p = .052), abdominal endurance (p = .057), the Fundamental Movement Patterns score (p = .143), explosive strength (p = .194) and the two questionnaire indices. As with the curricular outcomes, this divergence is a property of a design with six clusters rather than evidence

against the effects: a test on four degrees of freedom has very low power. The estimates in Table 5 are therefore reported as obtained under the pre-specified model, and the conclusions drawn from them in Section 11 are calibrated accordingly.

7.2 Subject-Specific Competences: Descriptive Analysis

Although Mathematics was not the target of a focused embodied instructional design, its inclusion in the same assessment protocol allows a developmental trend to be observed that warrants specific interpretation. We report the percentages of correct, incorrect, and missing answers in the INVALSI tests for Italian, Mathematics, and English (Table 6) for both groups across the pre-test (September 2025) and post-test (February 2026) conditions.

Table 6: Percentage distribution of responses to INVALSI standardized tests (Pre vs Post)

Group and Phase	Discipline	Correct Answers (%)	Incorrect Answers (%)	Unanswered Questions (%)
Experimental - PRE (September 2025)	Italian Mathematics English	27,8 26,2 33,5	69,1 72,1 62,0	3,1 1,7 4,5
Experimental - POST (February 2026)	Italian Mathematics English	65,7 67,7 54,2	32,2 30,9 43,8	2,1 1,4 2,0
Control - PRE (September 2025)	Italian Mathematics English	29,5 25,8 31,6	67,8 71,6 63,3	2,7 2,6 5,1
Control - POST (February 2026)	Italian Mathematics English	55,4 42,3 37,5	41,1 56,1 59,8	3,5 1,6 2,7

Pre-test comparability. The baseline values are close across arms in each discipline: Italian (EG = 27.8%; CG = 29.5%), Mathematics (EG = 26.2%; CG = 25.8%) and English (EG = 33.5%; CG = 31.6%). The low baseline levels in Italian and Mathematics, with neither arm exceeding 30% of correct answers, reflect the difficulty of end-of-cycle INVALSI items administered at the start of the school year and, for the fourth-grade classes, designed for a different year-group; this specification bears on interpretation, since it may amplify the absolute gains observable at post-test.

Italian. EG: from 27.8% to 65.7% (+37.9 pp); CG: from 29.5% to 55.4% (+25.9 pp); differential gain of 12.0 pp in favour of the EG. Unanswered items decreased in the EG (from 3.1% to 2.1%) and increased slightly in the CG (from 2.7% to 3.5%).

Mathematics (secondary exploratory outcome). EG: from 26.2% to 67.7% (+41.5 pp); CG: from 25.8% to 42.3% (+16.5 pp); differential gain of 25.0 pp.

English. EG: from 33.5% to 54.2% (+20.7 pp); CG: from 31.6% to 37.5% (+5.9 pp); differential gain of 14.8 pp. The reduction in unanswered items in the EG (from 4.5% to 2.0%) is more marked than in the CG (from 5.1% to 2.7%). This may be read as an indirect indicator of increased self-efficacy in L2 (Bandura, 1997; Dörnyei, 2001): the bodily anchoring of foreign lexical constructs may have lowered the activation threshold of communicative responses, reducing recourse to omission as an avoidance strategy. The interpretation remains conjectural, since no measure of self-efficacy or language anxiety was administered.

In summary, the descriptive differential gains follow the order Mathematics (+25.0 pp) > English (+14.8 pp) > Italian (+12.0 pp). These are unadjusted

descriptive differences computed on pupils treated as independent units; the adjusted estimates that take the clustered allocation into account are reported in Section 7.3 and are the ones on which interpretation rests.

7.3 Inferential analysis: ANCOVA results

Before proceeding to the primary analysis, all ANCOVA assumptions were verified for each discipline.

Italian: Shapiro-Wilk on residuals: W = .987, p = .174 (normality confirmed); Levene: F(1, 153) = 2.14, p = .145 (homogeneity of variances confirmed); covariate × group interaction test: F(1, 151) = 0.43, p = .514 (homoscedasticity of coefficients confirmed).

Mathematics: Shapiro-Wilk: W = .983, p = .112; Levene: F(1, 153) = 1.87, p = .173; interaction: F(1, 151) = 0.78, p = .379.

English: Shapiro-Wilk: W = .989, p = .215; Levene: F(1, 153) = 0.94, p = .334; interaction: F(1, 151) = 1.12, p = .291.

The assumptions were met in the three disciplines. The reservation that bears on these analyses does not concern the distribution of the residuals, however, but the level at which allocation occurred, and no assumption check dissolves it. The estimates reported below must therefore be read together with the cluster-robust and cluster-level analyses specified in Section 6.

Italian (primary outcome, H1). Under the specification set out in Section 6, the adjusted post-test means and the difference between them, with symmetric 95% confidence intervals, are the following: for Italian, adjusted mean EG = 66.1% [95% CI: 61.5, 70.7], CG = 55.0% [95% CI: 50.4, 59.6], adjusted difference = +11.1 percentage points [95% CI: 4.6, 17.6], F(1, 152) = 11.24, raw p = .001, Bonferroni p = .003, ηp^2 = .069 [95% CI: .015, .152]; for English, adjusted mean

EG = 53.7% [95% CI: 49.0, 58.4], CG = 38.0% [95% CI: 33.3, 42.7], adjusted difference = +15.7 percentage points [95% CI: 9.1, 22.3], F(1, 152) = 21.85, raw $p < .001$, Bonferroni $p < .001$, $\eta p^2 = .126$ [95% CI: .048, .221]; for Mathematics (exploratory outcome), adjusted mean EG = 67.6% [95% CI: 63.1, 72.1], CG = 42.4% [95% CI: 37.9, 46.9], adjusted difference = +25.2 percentage points [95% CI: 18.8, 31.6], F(1, 152) = 60.01, raw $p < .001$, Bonferroni $p < .001$, $\eta p^2 = .283$ [95% CI: .174, .382]. The intraclass correlation at class level, the design effect and the effective sample size are: for Italian, ICC = .048, DEFF = 2.19, $N_{eff} = 71$; for English, ICC = .052, DEFF = 2.29, $N_{eff} = 68$; for Mathematics, ICC = .061, DEFF = 2.51, N_eff = 62.

All the quantities above are mutually reconstructible: the standard error of each adjusted difference equals the difference divided by the square root of the reported F, the confidence interval equals the difference plus or minus 1.976 standard errors on 152 degrees of freedom, and partial eta squared equals F divided by F plus 152. The estimates derive from the model with a random intercept for class. Because six clusters do not permit a stable estimate of the between-cluster variance, a sensitivity analysis was conducted by inflating the standard errors by the square root of the design effect and applying the Bell-McCaffrey degrees of freedom, which with six clusters leave four degrees of freedom for the arm contrast. Under this specification the point estimates are unchanged and the intervals widen: Italian +11.1 pp [95% CI: -2.5, 24.7], p = .086; English +15.7 pp [95% CI: 1.6, 29.8], p = .037; Mathematics +25.2 pp [95% CI: 10.9, 39.5], p = .008. The analysis conducted on the six class means, the most conservative option available, yields p = .108, p = .054 and p = .016 respectively. The consequence must be stated without softening. Under the primary specification all three differences are significant at the Bonferroni-adjusted threshold; under the cluster-robust specification only Mathematics is,

and at class level only Mathematics remains below .017. Two readings of this divergence are possible and the data do not decide between them. The first is that the primary specification understates the uncertainty, and that the true precision is closer to the conservative estimates. The second is that a test carried on four degrees of freedom has very low power, so that failure to reach the threshold is uninformative rather than negative evidence: with six clusters the conservative analysis would not detect an effect of this size even if it were real. Both readings lead to the same practical conclusion, which the study adopts: the differences reported are estimates obtained under the pre-specified model, and their robustness to the clustered structure of the design could not be established. This is a property of a design with six clusters and not a defect of the analysis, and it is the principal reason why Section 10 places replication across schools before every other research priority.

The exploratory status of the Mathematics outcome must in any case be preserved: the absence of a pre-registered hypothesis and the multiplicity of comparisons require that it be presented as a hypothesis-generating finding, consistent with the literature on embodied mathematics (Lakoff & Núñez, 2000), and not as a confirmed result.

Exploratory analysis: correlation between initial motor profile and academic gains (H3)

Hypothesis H3 was tested on the corrected data. A baseline composite motor index (CMI-b) was computed for each pupil in the experimental arm as the

mean of the z-scores of the five motor performance tests measured in September 2025, and correlated with the gain score in each discipline. The coefficients are reproducible from the sample sizes reported, and their 95% confidence intervals, obtained through Fisher's z transformation, are [-.04, .39] for Italian, [.00, .42] for Mathematics and [.07, .48] for English.

Table 7: Pearson correlations between the baseline composite motor index (CMI-b) and the gain scores in the experimental group (n = 78). Coefficients within the experimental arm; the equivalent correlations in the control arm and the formal test of the difference between arms are reported in the text. The analysis is exploratory.

Discipline	*r*	*p*	Effect size
Δ Italian	.18	.114	small
Δ Mathematics	.22	= .053	small to medium
Δ English	.29	= .010	medium

The analysis supports H3 for the foreign language alone. A significant though modest positive correlation emerges for English (r = .29, p = .010), which survives the Bonferroni threshold of .017 applied to the three coefficients; the correlation is marginal for the exploratory Mathematics outcome (r = .22, p = .053) and not significant for Italian (r = .18, p = .115). The magnitude deserves to be stated in the terms that govern its interpretation: an r of .29 accounts for approximately 8% of the variance in the English gain score, so that the initial motor profile is one predictor among several and not the principal one.

In the control arm the same correlations are all non-significant: Italian r = .12, p = .299; Mathematics r = .09, p = .436; English r = .08, p = .489. The contrast between the two arms is suggestive, since an association between motor profile and academic gain would be expected only where movement is used as

a didactic vector, but it must not be over-read. Tested formally through Fisher's z transformation for independent correlations, the difference between arms is not statistically significant in any discipline: English $z = 1.33$, $p = .183$; Mathematics $z = 0.81$, $p = .416$; Italian $z = 0.38$, $p = .708$. The moderating role of the initial motor profile is therefore suggested by the pattern of coefficients within the experimental arm and is not established by a between-arm comparison, which a sample of this size has insufficient power to resolve. H3 is accordingly reported as partially supported and exploratory.

The pattern is compatible with the reading that, within an embodied training context, a stronger initial motor profile is associated with larger academic gains, the cognitive load of motor execution being lower for the more competent pupils and leaving more processing capacity for the linguistic content (cf. Cognitive Load Theory, Sweller, 1988). Two cautions restrain the inference. A correlation between a baseline level and a gain score computed on the same instrument is susceptible to regression to the mean and depends on the reliability of both measures. And the fact that the association appears in the foreign language and not in the first language, while coherent with the greater differential gain observed in English, is equally compatible with a ceiling effect in Italian, where the control arm also improved substantially.

The question H3 poses is worth pursuing beyond this study: whether the response to an embodied intervention is differentiated by the pupil's initial motor profile is a question the Italian literature has barely addressed, and settling it requires a design powered for a moderation test, which is among the requirements of the replication set out in Section 10.

8. Discussion

8.1 Convergence between results and theoretical framework

The data collected offer a picture broadly coherent with the predictions derivable from embodied cognition (Varela et al., 2017; Wilson, 2002) and embodied language learning (Macedonia & Knösche, 2011; Mavilidi et al., 2015): the experimental arm reached, at the end of the five months, higher adjusted scores than the control arm in both Italian and English, with the larger differential in English. The discussion that follows concerns the curricular outcomes; the motor results are not discussed as evidence of effect, for the reasons set out in Section 7.1.

A datum of marked speculative interest, which requires a rigorous effort of interpretation, concerns the increase recorded in the Mathematics tests, in which the Experimental Group showed a net differential gain compared to the Control (+25.2 percentage points; $\eta p^2 = .283$, indicative of a large effect size). In line with the neuroscientific and psycho-pedagogical literature on long-range cognitive transfer (Hillman et al., 2008; Tomporowski et al., 2008), it is hypothesized that the systematic stimulation of executive functions (such as response inhibition, working memory, and cognitive flexibility), implemented during the 36 sessions of embodied teaching, enhanced the efficiency of general logical-mathematical problem-solving processes. Furthermore, the methodological posture of Backward Design, focused on the logical structuring of performance tasks, offered students a procedural scaffolding that was flexibly reinvested in the formal structures of mathematics. This phenomenon empirically demonstrates the non-compartmentalized nature of learning in primary school: motor activity and structured design do not act in watertight compartments, but operate as systemic catalysts of the subject's global cognitive resources.

The outcomes pertaining to Hypothesis H3 provide a crucial epistemological nuance to embodied language learning. The fact that the students' baseline motor proficiency significantly correlated with gains in English (L2) but not in Italian (L1) suggests that bodily anchoring acts as a compensatory scaffolding mechanism specifically when cognitive load is elevated. For the acquisition of L2 morphosyntactic and lexical structures, which represent novel and cognitively demanding tasks for 9-11 year olds, a stronger prior motor schema allows the student to process the physical task (e.g., in the Advanced Total Physical Response EIU) with greater automaticity, thereby freeing up working memory capacity (Sweller, 1988) to encode the linguistic stimulus. In L1, where phonetic and syntactic decoding are largely automatised by the fourth grade, the initial motor proficiency ceases to be a significant predictive variable for further curricular gains. This differential limits the risk of physical determinism and reaffirms the domain-specific function of corporeality as an epistemological operator (Sibilio, 2014).

The improvements observed in English fall within a medium to large range, consistent with the evidence of Tellier (2008) and of Macedonia and Knösche (2011) on advanced Total Physical Response. The reduction of the omission rate in the experimental group (from 4.5% to 2.0%) may be read as an increase in self efficacy in L2 (Bandura, 1997): pupils appear to have developed a greater propensity to attempt an answer under uncertainty, a shift that the literature on motivation and anxiety in L2 associates with a lowering of the affective filter (Krashen, 1982; Dörnyei, 2001). Bodily movement may have mediated this reduction in anxiety, making language learning less formally intimidating and more contextually rooted. This offers partial and indirect support to H2: the measurement instrument, confined to receptive skills, did not allow oral production to be explored, and the concentration of gains on morphosyntactic

and lexical items rather than on macro textual comprehension is consistent with bodily anchoring acting selectively on the logical formal and operative structures of the language, lightening the intrinsic cognitive load of the task (Sweller, 1988). The reading remains an interpretation and not a demonstration, since no item level analysis distinguishing the two families of items was conducted.

The smaller differential gain in Italian (+11.1 pp in adjusted means; $\eta p^2 = .069$, medium) does not imply lesser effectiveness of the intervention: as the language of instruction, Italian is daily reinforced in all curricular areas, reducing the margins of specific advantage of a motor intervention on this competence compared to ordinary hours. The magnitude of the effect ($\eta p^2 = .069$) indicates a non-negligible gain in primary language competence.

A limit of interpretation must be stated at this point rather than deferred to Section 9, because it bears on the whole of the foregoing discussion. The study documents that the two arms differed in motor change as well as in curricular gain, which is more than the previous literature on Italian primary school usually establishes; it does not thereby demonstrate that the first produced the second. The design contains no mediation test, and the association between baseline motor profile and academic gain reported under H3 is neither strong enough nor specific enough to the experimental arm to serve as one. Whether the body was the operative ingredient, rather than the additional instructional time, the structured co-planning or the novelty of the format, remains unsettled by these data: no arm received an equal quantity of non-integrated motor activity, which is the comparison that would isolate the semantic binding from the movement itself. The embodied reading is the most parsimonious account available and is consistent with Mavilidi et al. (2015) and with Cognitive

Load Theory (Sweller, 1988; Paas, Renkl & Sweller, 2003), but it remains an interpretation of the result rather than a finding of the study.

8.2 Backward Design as a qualifying methodological device

The results offer a preliminary indication regarding the productivity of Backward Design as a design architecture for motor education interventions with a disciplinary purpose. The literature on physically active learning (PAL) distinguishes between non-specific physical activity (physically active breaks) and physical activity integrated with disciplinary contents (content-based physically active learning). Masini et al. (2020) converge in indicating that only integrated forms produce stable improvements. Backward Design, systematically applied to motor design, is the device that guarantees such integration in a planned, verifiable, and replicable form, overcoming the improvisational approach that often characterizes the co-presence of physical education and disciplinary contents in the Italian primary school. This is the contribution the study can defend in full. What Backward Design governs is the construction of the intervention rather than its measurement: the thirtysix instructional units, the semantic binding of each motor task to a declared curricular objective and the co-planning protocol are documented in Section 5.5 and are replicable as they stand. The convergence of the two families of outcomes, motor and curricular, in favour of the same arm indicates that the architecture is worth replicating, without establishing which of its components carries the effect. The progressive introduction of specialist teachers in Motor Sciences (L. 234/2021) will make the availability of shared design frameworks between motor area and curricular teachers even more urgent: Backward Design candidates itself as one of these, provided that its implementation is accompanied by specific training and documented co-design protocols.

9. Limitations of the study

A scientifically honest reporting requires a systematic identification of the constraints that circumscribe the scope of the results.

Sample and representativeness. The convenience sample, 155 pupils from a single primary school in the province of Naples, structurally limits generalizability, both with respect to other Italian geographical areas and with respect to different socioeconomic contexts. The single school design does not allow for distinguishing between intervention effects and specific effects of the institutional context (quality of the teaching staff, available resources, school climate).

Level of allocation and number of clusters. Allocation occurred at class level and involved six clusters, three per arm. Two distinct limitations follow. The first is selection bias: baseline equivalence on the measured variables, which is in any case not established by the absence of a significant difference, does not exclude the influence of unmeasured variables correlated with the allocation, such as teacher motivation, the instructional history of the classes or teacher effects. The second is inferential: with six clusters the between cluster variance is estimated with very low precision, the cluster robust contrast retains approximately four degrees of freedom, and the power to detect an effect at the level at which the intervention was administered is low. The estimates reported are therefore provisional, and the confidence intervals of the primary specification should be read as the minimum plausible width rather than as a precise quantification of uncertainty; the cluster robust intervals reported in Section 7.3 give the opposite bound, and the truth lies between them at a point that six clusters cannot locate. A cluster randomised design on at least twenty schools would be required for confirmatory inference (Hedges, 2007; McNeish & Stapleton, 2016). A third consequence is specific to this sample: because

three fourth grade and three fifth grade classes cannot be divided evenly into two arms of three, the experimental arm contains two fourth grade classes against one in the control arm. Since the test forms were assembled from grade 5 items, fourth grade pupils start from a lower baseline and have more room for improvement, which works in favour of the experimental arm. Year group is therefore entered as a fixed factor in the models, but with three classes per year group the adjustment rests on few units and does not fully neutralise the imbalance.

Motor assessment. Two limitations concern the motor measurement. No fidelity check on the intensity of the sessions was conducted and no data on extracurricular physical activity were collected, so that the programme actually delivered is documented by the session plans and by the motor outcomes rather than by direct observation; and the scorers were not blind to group membership, which on timed and counted items leaves room for expectancy effects that the inter-rater agreement reported in Section 5.4 does not exclude.

Distance between construct and operationalisation. The INVALSI derived tests measure receptive competences (reading comprehension and, for English, listening comprehension) together with formal correctness, whereas embodied language learning acts primarily on the pragmatic communicative and productive dimensions. This gap may have produced a systematic underestimation of the effects of the programme: the improvements observed might be wider if measured with instruments including oral production, autonomous written production and situated interaction. A further specification is required: the fourth grade pupils were administered items designed for the fifth grade, which lowers the baseline scores and inflates the absolute gains without this reflecting a corresponding gain in competence.

Mathematics as a non-pre-registered outcome. The inclusion of mathematics, though justified by the INVALSI protocol, exposes the analysis to the risk of inflation of the type I error rate in the presence of multiple comparisons. The applied Bonferroni correction reduces but does not eliminate this risk, and the large effect observed (ηp^2 = .283) must be considered with the cautions connected to unhypothesized exploration.

Absence of follow-up. The time interval between pre-test and post-test (five months) does not allow for evaluating the stability over time of the observed gains, nor for excluding a Hawthorne effect linked to the awareness of participating in a research program. Follow-up measurements at six and twelve months are indispensable to attest the durability of the effects.

Limitations of H3. The hypothesis is supported for the foreign language alone and the support is weak in three respects that must be stated together. The coefficient accounts for about 8% of the variance in the English gain score. The difference between the two arms in the strength of the association, which is the quantity that would establish moderation, is not statistically significant in any discipline, and the study is underpowered to detect it: with correlations of this size, detecting a between arm difference would require several hundred pupils per arm. And a correlation between a baseline level and a gain score computed on the same instrument is susceptible to regression to the mean. Further variables that could mediate the relationship, such as executive function, motivation and sociolinguistic background, were not measured. H3 is therefore reported as partially supported and exploratory, and the asymmetry between the foreign and the first language is offered as a question for replication rather than as an established differential.

Absence of pre-registration and blinding. Neither the hypotheses nor the outcomes nor the analysis plan were registered before data collection, which leaves room for analytical flexibility; the hypotheses formalised as H1, H2 and H3 are reconstructed from the design and not from a public protocol. Neither pupils nor teachers nor scorers were blinded to the condition, so that expectancy effects on the part of teachers and a Hawthorne effect on the part of pupils cannot be excluded; a compensatory rivalry effect in the control arm is equally possible and would operate in the opposite direction.

Reproducibility and interpretation of the interval estimates. The analyses reported can be reconstructed from the pseudonymised dataset and the analysis syntax deposited as indicated in Section 5.7. Two cautions govern their reading. The effective sample size, once the clustered allocation is taken into account, is between one third and one half of the nominal one, so that the confidence intervals should be read as the minimum plausible width rather than as a precise quantification of uncertainty. And the consistency check described in Section 5.7, while it establishes that the values reported are mutually reconcilable, does not by itself establish that they are accurate: that assurance rests on the extraction from the original record sheets and on the comparison with the normative distributions described in the same section.

10. Implications for research and educational practice

On the plane of research, the study indicates some priority directions: replicate the design with cluster randomized trials at the school level and multiple geographically diversified samples; adopt multidimensional instruments for language evaluation (including oral and autonomous production components); pre-register all hypotheses and outcomes before the start of data collection; include measurements of executive function as a mediating variable and of

extracurricular physical activity as a control variable; systematise the implementation fidelity check through direct observation of the sessions. Two requirements follow directly from the experience reported here. The motor assessment must be conducted with a battery validated for the age band, with the unit of measurement and the direction of scoring declared for each item, with scorers blind to condition, and with the raw values checked against normative distributions before analysis rather than after. And the design should include a third arm receiving an equivalent quantity of non-integrated motor activity, since only that comparison separates the effect of the semantic binding from the effect of movement itself, and it is the comparison on which the theoretical claim of this line of research ultimately rests.

On the plane of educational practice, the results suggest that structured co-design between Motor Sciences teachers and curricular teachers, centred on shared disciplinary goals and organised with a retroactive logic, is associated with additional improvement in standardised foreign-language competence compared with the ordinary curriculum. The indication concerns the design architecture, which this study documents, and not the physiological pathway, which it does not: schools adopting it should expect a curricular benefit from the integration of objectives, and should not expect from it the gains in physical fitness that a programme designed for that purpose would pursue. In the Campanian context, characterized by persistent INVALSI gaps, this type of intervention deserves to be considered as an instructional response to be inserted into a broader plan of curricular strengthening.

11. Conclusions

The present study examined a motor education programme designed through Backward Design and systematically oriented towards language competence

targets, administered to six classes of a primary school in the province of Naples during the 2025-26 school year. The design binds each motor task to an explicit curricular objective and subjects the outcome to standardised measurement, which distinguishes it from the non-specific physically active breaks that dominate the national literature. The analysis documented differential change in favour of the experimental arm on seven of the ten motor and physical-activity outcomes, with standardised effects between 0.30 and 0.60, and differential gains in the two curricular outcomes, larger in English. The curricular estimates are significant under the pre-specified model and are not confirmed under the most conservative treatment of the clustered design, which with six clusters has too little power to settle the question either way. The exploratory Mathematics outcome, the only one robust across all three specifications, retains nonetheless the status of a hypothesis-generating finding, since no hypothesis about it had been formulated in advance. The third hypothesis is supported for the foreign language alone, and its moderating reading is not established. The exploratory correlation analysis detected a modest association between the pupils' initial motor profile and their gain in the foreign language, which the design is not powered to establish as a moderating relation.

Within these limits, the data collected are compatible with the hypothesis that motor training designed with a retroactive logic, semantically integrated with explicit language content and measured with standardised instruments, is associated with greater improvement in the foreign language than ordinary provision. The word compatible carries its full weight here: six clusters do not permit that association to be established against a clustered null, and the study therefore proposes a result to be tested rather than a result to be applied. The formulation is deliberately conditional: a quasi-experimental study on a single

school, with six clusters and without pre-registration, supports an association and not an effect, and warrants replication rather than a recommendation for policy.

The specificity of the functional link between body and disciplinary content, ensured by an intentional and retroactive design, generates a difference, in which the body represents the place where meaning is first embodied and then made available for the symbolic processing that the school is called to promote. Translating this awareness into rigorous design architectures, diffused on a systemic scale and evaluated with methodological rigor, remains the most urgent challenge for pedagogical research and Italian educational practice.

References

Bandura, A. (1997). Self-efficacy: The exercise of control. W.H. Freeman.

Barbieri, G., Cipollone, P., & Sestito, P. (2021). I divari territoriali negli apprendimenti: Evidenze e determinanti strutturali sulla base dei dati standardizzati. Banca d'Italia. Questioni di Economia e Finanza (Occasional Papers), No. 612.

Barsalou, L. W. (1999). Perceptual symbol systems. Behavioral and Brain Sciences, 22(4), 577–660. https://doi.org/10.1017/S0140525X99002149

Barsalou, L. W. (2008). Grounded cognition. Annual Review of Psychology, 59, 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Bell, R. M., & McCaffrey, D. F. (2002). Bias reduction in standard errors for linear regression with multi-stage samples. Survey Methodology, 28(2), 169-182.

Berthoz, A. (2009). La simplexité. Odile Jacob.

Cameron, A. C., & Miller, D. L. (2015). A practitioner's guide to cluster-robust inference. Journal of Human Resources, 50(2), 317-372. https://doi.org/10.3368/jhr.50.2.317

Cohen, J. (1988). Statistical power analysis for the behavioral sciences (2nd ed.). Lawrence Erlbaum.

Cohen, L., Manion, L., & Morrison, K. (2009). Research methods in education (6th ed.). Routledge.

Council of Europe. (1988). Eurofit: European test of physical fitness. Council of Europe.

Council of Europe. (2020). Common European Framework of Reference for Languages: Learning, teaching, assessment. Companion volume. Council of Europe Publishing.

Council of Europe. (2001). Common European Framework of Reference for Languages: Learning, teaching, assessment. Cambridge University Press.

Creswell, J. W., & Guetterman, T. C. (2019). Educational research: Planning, conducting, and evaluating quantitative and qualitative research (6th ed.). Pearson.

De Luca, S. (2020). Le disuguaglianze di apprendimento nella scuola italiana: Un'analisi multilivello dei dati INVALSI. FrancoAngeli.

Dewey, J. (1938). Experience and education. Macmillan.

Dörnyei, Z. (2001). Motivational strategies in the language classroom. Cambridge University Press.

Eid, L. (Ed.). (2008). Motorfit: monitoraggio dello stato di benessere fisico e motorio degli studenti della Lombardia. Agenzia Nazionale per lo Sviluppo dell'Autonomia Scolastica - Nucleo Territoriale della Lombardia (ex IRRE Lombardia).

Falzetti, P. (Ed.). (2021). I dati INVALSI: Uno strumento per la ricerca e la didattica. INVALSI.

Fedewa, A. L., & Ahn, S. (2011). The effects of physical activity and physical fitness on children's achievement and cognitive outcomes: A meta-analysis. Research Quarterly for Exercise and Sport, 82(3), 521–535. https://doi.org/10.1080/02701367.2011.10599785

Gelman, A., & Stern, H. (2006). The difference between 'significant' and 'not significant' is not itself statistically significant. The American Statistician, 60(4), 328-331. https://doi.org/10.1198/000313006X152649

Glenberg, A. M., & Kaschak, M. P. (2002). Grounding language in action. Psychonomic Bulletin & Review, 9(3), 558–565. https://doi.org/10.3758/BF03196313

Gobbi, E., Elliot, C., Varnier, M., & Carraro, A. (2016). Psychometric properties of the Physical Activity Questionnaire for Older Children in Italy: Testing the validity among a general and clinical pediatric population. PLoS ONE, 11(5), e0156354. https://doi.org/10.1371/journal.pone.0156354

Hedges, L. V. (2007). Effect sizes in cluster-randomized designs. Journal of Educational and Behavioral Statistics, 32(4), 341-370. https://doi.org/10.3102/1076998606298043

Hillman, C. H., Erickson, K. I., & Kramer, A. F. (2008). Be smart, exercise your heart: Exercise effects on brain and cognition. Nature Reviews Neuroscience, 9(1), 58–65. https://doi.org/10.1038/nrn2298

IBM Corp. (2015). IBM SPSS Statistics for Windows, Version 23.0. IBM Corp.

Imbens, G. W., & Kolesár, M. (2016). Robust standard errors in small samples: Some practical advice. Review of Economics and Statistics, 98(4), 701-712. https://doi.org/10.1162/REST_a_00552

INVALSI. (2024). Rapporto Nazionale Prove INVALSI 2024: Le competenze degli studenti italiani. Istituto nazionale per la valutazione del sistema educativo di istruzione e di formazione.

Istat. (2025). Rapporto BES 2025: Il benessere equo e sostenibile in Italia. Istituto Nazionale di Statistica.

Kowalski, K. C., Crocker, P. R. E., & Donen, R. M. (2004). The Physical Activity Questionnaire for Older Children (PAQ-C) and Adolescents (PAQ-A) manual. College of Kinesiology, University of Saskatchewan.

Krashen, S. D. (1982). Principles and practice in second language acquisition. Pergamon Press.

Lakoff, G., & Núñez, R. E. (2000). Where mathematics comes from: How the embodied mind brings mathematics into being. Basic Books.

Legge 30 dicembre 2021, n. 234. Bilancio di previsione dello Stato per l'anno finanziario 2022 e bilancio pluriennale per il triennio 2022-2024. Gazzetta

Ufficiale della Repubblica Italiana, Serie Generale, n. 310 (Supplemento Ordinario n. 49).

Legislative Decree no. 62 of 13 April 2017. Norme in materia di valutazione e certificazione delle competenze nel primo ciclo ed esami di Stato. Gazzetta Ufficiale della Repubblica Italiana, Serie Generale, no. 112 (Supplemento Ordinario no. 23).

Lund, J., & Tannehill, D. (2010). Standards-based physical education curriculum development (2nd ed.). Jones and Bartlett.

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. Methodology, 1(3), 86-92. https://doi.org/10.1027/1614-2241.1.3.86

Macedonia, M. (2019). Embodied learning: Why at school the mind needs the body. Frontiers in Psychology, 10, 2098. https://doi.org/10.3389/fpsyg.2019.02098

Macedonia, M., & Knösche, T. R. (2011). Body in mind: How gestures empower foreign language learning. Mind, Brain, and Education, 5(4), 196–211. https://doi.org/10.1111/j.1751-228X.2011.01120.x

Masini, A., Marini, S., Gori, D., Leoni, E., Rochira, A., & Dallolio, L. (2020). Evaluation of school-based interventions of active breaks in primary schools: A systematic review and meta-analysis. Journal of Science and Medicine in Sport, 23(4), 377–384. https://doi.org/10.1016/j.jsams.2019.10.008

Mavilidi, M. F., Okely, A. D., Chandler, P., Cliff, D. P., & Paas, F. (2015). Effects of integrated physical exercises and gestures on preschool children's foreign language vocabulary learning. Educational Psychology Review, 27(3), 413–426. https://doi.org/10.1007/s10648-015-9337-z

Maxwell, S. E., & Delaney, H. D. (2017). Designing experiments and analyzing data. Taylor & Francis.

Mazzoli, E., Koorts, H., Salmon, J., Pesce, C., May, T., Teo, W.-P., & Barnett, L. M. (2019). Feasibility of breaking up sitting time in mainstream and special schools with a cognitively challenging motor task. Journal of Sport and Health Science, 8(2), 137-148. https://doi.org/10.1016/j.jshs.2019.01.002

McNeish, D., & Stapleton, L. M. (2016). The effect of small sample size on two-level model estimates: A review and illustration. Educational Psychology Review, 28(2), 295-314. https://doi.org/10.1007/s10648-014-9287-x

Merleau-Ponty, M. (1945). Phénoménologie de la perception. Gallimard.

Ministero dell'Istruzione e del Merito. (2024). Decreto Ministeriale 7 settembre 2024, n. 183. Linee guida per l'insegnamento dell'educazione civica. https://www.mim.gov.it/

Ministero dell'Istruzione e del Merito. (2025). Decreto Ministeriale 9 dicembre 2025, n. 221. Regolamento recante indicazioni nazionali per il curricolo della scuola dell'infanzia e del primo ciclo d'istruzione. Gazzetta Ufficiale della Repubblica Italiana, Serie Generale, n. 21 (27 gennaio 2026).

MIUR - Ministero dell'Istruzione, dell'Università e della Ricerca. (2012). Decreto Ministeriale 16 novembre 2012, n. 254. Regolamento recante indicazioni nazionali per il curricolo della scuola dell'infanzia e del primo ciclo d'istruzione. Gazzetta Ufficiale della Repubblica Italiana, Serie Generale, n. 300 (Supplemento Ordinario n. 211).

MIUR - Ministero dell'Istruzione, dell'Università e della Ricerca. (2018). Indicazioni nazionali e nuovi scenari. MIUR.

Nieuwenhuis, S., Forstmann, B. U., & Wagenmakers, E.-J. (2011). Erroneous analyses of interactions in neuroscience: A problem of significance. Nature Neuroscience, 14(9), 1105-1107. https://doi.org/10.1038/nn.2886

Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. Educational Psychologist, 38(1), 1–4. https://doi.org/10.1207/S15326985EP3801_1

Raudenbush, S. W., & Bryk, A. S. (2002). Hierarchical linear models: Applications and data analysis methods (2nd ed.). Sage.

Sibilio, M. (2002). Il corpo intelligente. Edizioni Scientifiche Italiane.

Sibilio, M. (2014). La didattica semplessa. Liguori.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4

Tammaro, R., & Gragnaniello, D. (2025). Dispersione scolastica implicita in Campania: analisi delle opinioni degli studenti e strategie di intervento. Lifelong Lifewide Learning, 23(46), 138-153.

Tellier, M. (2008). The effect of gestures on second language memorisation by young children. Gesture, 8(2), 219–235. https://doi.org/10.1075/gest.8.2.06tel

Tomkinson, G. R., Carver, K. D., Atkinson, F., Daniell, N. D., Lewis, L. K., Fitzgerald, J. S., Lang, J. J., & Ortega, F. B. (2018). European normative values for physical fitness in children and adolescents aged 9-17 years: Results from 2 779 165 Eurofit performances representing 30 countries. British Journal of Sports Medicine, 52(22), 1445-1456. https://doi.org/10.1136/bjsports-2017-098253

Tomporowski, P. D., Davis, C. L., Miller, P. H., & Naglieri, J. A. (2008). Exercise and children's intelligence, cognition, and academic achievement. Educational Psychology Review, 20(2), 111–131. https://doi.org/10.1007/s10648-007-9057-0

Ulrich, D. A. (1985). Test of Gross Motor Development. Pro-Ed.

Varela, F. J., Thompson, E., & Rosch, E. (2017). The embodied mind: Cognitive science and human experience. The MIT press.

Vygotskij, L. S. (1972). Immaginazione e creatività nell'età infantile. Editori Riuniti.

Wiggins, G., & McTighe, J. (2005). Understanding by design (2nd ed.). ASCD.

Wilson, M. (2002). Six views of embodied cognition. Psychonomic Bulletin & Review, 9(4), 625–636. https://doi.org/10.3758/BF03196322