Forum for Linguistic Studies 2025, 7(2)

ISSN:2705-0610 E-ISSN:2705-0602

MOVE TO LEARN: BRIEF CALISTHENICS IMPROVE EFL VOCABULARY AND ENGAGEMENT IN TWO ARAB HIGH SCHOOLS

Jamal Assadi

The college of Sakhnin Academic; College for Teacher Education

Crrosponding author : Jamal Assadi.

Email. Jamal-a@sakhnin.ac.il

Abstract

This quasi-experimental study examined whether brief calisthenic bouts embedded in vocabulary lessons improve English-as-a-foreign-language (EFL) outcomes in two Arab high schools in the Galilee. Two intact Grade-10 classes (N = 44; control n = 22; intervention n = 22) participated over 8 weeks. The intervention class completed three 2–3-minute calisthenic bouts per lesson—push-ups and light movements—explicitly linked to target lexical items; the control class received standard instruction. Vocabulary knowledge and lesson-specific engagement were assessed pre/post with curriculum-aligned instruments. Analytically, post-test outcomes were compared while adjusting for pre-test performance. The intervention outperformed the control on both vocabulary and engagement, indicating practically meaningful advantages for movement-integrated lessons rather than sedentary drill alone. The pattern aligns with embodied-cognition and Total Physical Response accounts: coupling motor activity with verbal rehearsal appears to strengthen the encoding and retrieval of new lexical items while elevating affective/behavioral engagement during lessons. The routine is time-neutral—it fits inside existing periods—requires no equipment, and is easily differentiated (e.g., wall or knee push-ups), making it feasible for schools with limited resources. Brief teacher training focused on safe progressions, pacing, and inclusive modifications is recommended for implementation. Importantly, the approach offers a low-cost equity lever: by adding short, structured movement cycles to ordinary lessons, teachers can provide an engaging practice format that may particularly benefit learners with lower baseline engagement. The study contributes context-specific evidence from Arab secondary schools—an under-represented setting—showing that movement-infused instruction can simultaneously improve vocabulary learning and motivation without displacing content time. It also supplies a replicable classroom protocol (three short bouts per lesson linked to retrieval cues) that schools can pilot and scale within regular EFL instruction.

Keywords: EFL vocabulary, embodied cognition, TPR, classroom physical activity, calisthenics, engagement, quasi-experiment

1. Introduction

In many EFL classrooms, learning remains largely sedentary: students sit, copy, and rehearse while teachers shoulder most of the cognitive load. Yet a substantial body of research in grounded/embodied cognition shows that coupling motor activity with encoding heightens attention and deepens memory traces, because cognition emerges through interaction with the body and environment rather than operating as an amodal, isolated system (Barsalou 2020). Recent syntheses indicate that when movement is purposefully integrated with academic content—rather than delivered as generic “activity breaks”—students tend to show small-to-

moderate gains in academic outcomes and on-task behavior (Bédard et al. 2019; Norris et al. 2020; Mavilidi et al. 2022). Within language learning specifically, action-linked methods have a long pedigree. Asher's early Total Physical Response (TPR) experiments paired verbal commands with whole-body actions and reported robust effects on vocabulary and early grammar, findings that helped establish the theoretical and practical plausibility of motor–verbal coupling for L2 learning (Asher 1969). Contemporary reappraisals and small-scale trials continue to report vocabulary and retention benefits when teachers embed physical actions into communicative routines (e.g., classroom TPR variants), suggesting durable if context-dependent advantages (Oppici et al. 2023).

A complementary stream of evidence comes from gesture research. Experimental and review work shows that enacting or observing meaningful movements during L2 word learning can improve recall, with self-performed gestures typically yielding the strongest effects—consistent with sensorimotor grounding accounts (Macedonia 2014; Oppici et al. 2023). These effects likely arise because gestures add multimodal cues that bind phonological forms to semantic and motor representations, enhancing consolidation and retrieval (Macedonia 2014; Oppici et al. 2023).

Despite this momentum, significant gaps remain. First, most classroom-based movement studies are concentrated in primary schools and STEM subjects, leaving fewer EFL-focused trials at the secondary level. Second, context-specific research from Arab schools in the Galilee is scarce, which limits external validity for local adoption. Third, very few studies have tested ultra-brief, easy-to-scale routines that fit seamlessly within ordinary lessons (e.g., two- to three-minute calisthenic bouts tied directly to target lexical items). Addressing these gaps, the present study asks whether brief, structured calisthenics—push-ups and light movements linked to lexical retrieval and delivered by the regular EFL teacher—can improve Grade-10 learners' vocabulary acquisition and engagement in two Arab high schools. By evaluating a low-cost, time-efficient protocol aligned with embodied/TPR principles, the study aims to (a) test a scalable routine that minimizes opportunity cost, (b) add evidence from an under-researched setting, and (c) examine dual outcomes (achievement and engagement) that are critical for sustained L2 learning. In doing so, it builds on meta-analytic indications that physically integrated lessons outperform non-integrated "breaks," while extending gesture/TPR insights into a pragmatic, secondary-school EFL design (Bédard et al. 2019; Norris et al. 2020; Mavilidi et al. 2022; Macedonia 2014; Oppici et al. 2023).

2. Literature Review

A growing literature indicates that when language learning is treated as a whole-body activity rather than a purely symbolic exercise, memory for words improves and learners stay more engaged. Grounded accounts of cognition argue that conceptual processing recruits sensorimotor systems; thus, pairing movements with verbal material should strengthen encoding and retrieval (Barsalou 2020). Experimental and meta-analytic work on acute exercise shows small but reliable benefits for attention and executive processes that support learning, especially when the physical activity occurs proximal to encoding (Chang et al. 2012). School-based reviews converge on positive—if heterogeneous—associations between physical activity and academic outcomes, while calling for clearer specification of dose, timing, and task integration (Donnelly et al. 2016). Together, these strands motivate integrating brief,

purposeful movements into language lessons, not as generic brain breaks but as content-linked actions that scaffold memory.

Within second-language pedagogies, Total Physical Response (TPR) is a longstanding instantiation of this principle. Asher's classic studies had learners execute commands ("stand," "turn," "touch") while listening, yielding notable gains in beginning vocabulary and early grammar (Asher 1969). Contemporary evaluations continue to report positive effects for vocabulary acquisition and affect (e.g., reduced anxiety, higher participation) across EFL settings, while noting variability by learner age and instructional design (Al-Obaydi and Pikhart 2024). Importantly, most TPR implementations use semantically congruent gestures or whole-body actions tied to the target language; fewer studies have tested brief calisthenic "micro-bouts" (e.g., push-ups, squats) in secondary EFL contexts, leaving open whether non-semantic but arousing movements—when explicitly linked to lexical retrieval—can confer comparable benefits.

Parallel research on classroom physical activity (CPA) and physically active lessons (PAL) provides a pragmatic template for such tests. Meta-analyses and systematic reviews indicate that CPA integrated into academic content can improve lesson-time on-task behavior and produce small but meaningful gains in academic performance; effects tend to be stronger when movement is embedded in the learning task rather than appended as a break (Watson et al. 2017; Bédard, Bremer, and Trudeau 2019; Norris et al. 2020). Short protocols (≈2–5 minutes) are especially appealing in time-pressed classrooms: dose-response experiments show that even 5–10-minute bouts can acutely improve on-task behavior, with negligible instructional opportunity cost (Howie, Beets, and Pate 2014; Howie et al. 2015). Recent syntheses further suggest that movement-integrated lessons can enhance both cognition and achievement across age groups when appropriately dosed and aligned to content (Mavilidi et al. 2022). For older students, simple calisthenics offer practical advantages—no equipment, clear dosage, rapid physiological effects—and, when paired explicitly with retrieval cues (e.g., a push-up per target item), may harness both arousal and enactment to support vocabulary learning and engagement.

Against this backdrop, empirical evidence from Arab secondary schools in the Galilee remains limited, particularly regarding ultra-brief, teacher-delivered calisthenic routines embedded within ordinary EFL vocabulary instruction. The present study addresses this gap by evaluating a feasible micro-protocol—push-ups and light calisthenics directly coupled to lexical retrieval—administered by regular EFL teachers and assessed on dual outcomes: vocabulary achievement and classroom engagement. Beyond testing efficacy in an under-researched context, the study sharpens questions about action intensity, semantic congruence, and timing that matter for scalable adoption in resource-constrained schools.

3. Research Questions and Hypotheses

RQ1. Do brief, classroom calisthenics integrated into vocabulary lessons improve EFL vocabulary acquisition compared with standard instruction?

RQ2. Do these calisthenics increase motivation/engagement with EFL learning?

H1. The intervention class will show greater vocabulary gains than controls.

H2. The intervention class will show higher motivation/engagement than controls.

H3. Gains will hold after controlling for pre-test and will be moderate or larger ($d \geq 0.50$).

4. Method

4.1 Design

We employed a quasi-experimental, pre-test–post-test control-group design with intact Grade-10 classes from two Arab high schools in the Galilee. One class received calisthenics-infused vocabulary instruction; the comparison class received standard instruction. The intervention lasted 8 weeks with three 45–60-minute EFL lessons per week. Because assignment followed timetables rather than randomization, we followed TREND guidance for nonrandomized evaluations and documented context, participant flow, and intervention details (Des Jarlais, Lyles, and Crepaz 2004). Sample size. As a classroom feasibility trial with intact groups (N = 44; 22/22), the study was sized pragmatically to detect moderate–large standardized effects (≈ d = 0.60–0.80) on post-test outcomes at α = .05; smaller effects would require a larger, multi-class design.

Primary analysis. The confirmatory model was an ANCOVA on post-test outcomes with the corresponding pre-test as a covariate. We pre-specified assumption checks: (i) homogeneity of regression slopes tested via the Group × Pre-test interaction; (ii) linearity of the covariate–outcome relation; (iii) residual normality (Q–Q plots/Shapiro–Wilk) and (iv) homoscedasticity (studentized residuals vs fitted; Levene's for group variances; **Fox 2016**).

Sensitivity analyses. If any assumption was borderline, we ran (a) a change-score model (Δ = post − pre) comparing groups, and (b) a robust ANCOVA (e.g., heteroskedasticity-consistent SEs/HC3). Convergent inferences across these models are reported.

Cluster/teacher limitation. With one class per arm, treatment is partially confounded with class/teacher. We therefore interpret student-level ANCOVA effects cautiously, describe the limitation explicitly, and report fidelity and attendance to contextualize internal validity.

All inferential values (F/t, df, p), partial η^2 (with 95% CIs where feasible), adjusted means ± SE, covariate coefficients, and effect sizes are reported in Results; tables follow journal style.

4.2 Participants

Two intact Grade-10 classes participated (N = 44; 21 female, 23 male): control n = 22; intervention n = 22. Both classes followed the same EFL curriculum and were taught by experienced EFL teachers employed by their respective schools. Assignment to condition was determined by scheduling feasibility (quasi-experimental). Given the intact-group design, we report baseline equivalence and adjust for pre-test differences in the primary ANCOVA, presenting covariate-adjusted means with 95% CIs.

Attendance. We tracked lesson-level attendance across the study and report class-specific attendance rates in the Results/fidelity table (% of sessions attended per student; median [IQR]).

4.3 Intervention: "2–3-minute calisthenic micro-bouts"

At three planned points in each vocabulary lesson, the intervention class performed brief calisthenic micro-bouts (e.g., push-ups, wall push-ups, modified plank, air squats) paired with target lexical items:

1. Encode: Teacher presented 6–8 target words with modeled pronunciation and meanings.
2. Enact: For each word, learners performed ~8–10 repetitions while chanting the word in unison; antonyms/synonyms could be alternated across reps.
3. Retrieve: Later in the lesson, the teacher cued items; students performed the movement + production; peers quizzed in pairs.

4. Consolidate: A 30–45-second "recall bout" cycled the day's words before dismissal.

Movements were scalable (e.g., wall/knee push-ups) to ensure inclusive participation. The control class had equivalent vocabulary exposure/practice time without calisthenics.

Fidelity and safety. Teachers received a brief skills-focused orientation (progressions, differentiation, classroom management, safety). We logged fidelity each lesson: whether all three micro-bouts occurred; median repetitions per bout; and any accommodations used. We also note any opt-outs/adverse events (none anticipated given low-impact options). This protocol aligns with evidence that integrated movement (rather than generic breaks) can support engagement and learning and that short, in-class bouts are feasible with minimal opportunity cost (Howie, Beets, and Pate 2014; Howie et al. 2015; Norris et al. 2020; Bédard, Bremer, and Trudeau 2019). Safety guidance followed adolescent recommendations emphasizing regular, age-appropriate muscle-strengthening (WHO 2020).

4.4 Measures

Vocabulary test

A teacher-developed test targeted only the instructed lexical items and closely related collocations (multiple-choice + short-answer production). Multiple-choice (MC) items assessed form–meaning mapping and paradigmatic relations (synonym/antonym), while short-answer items elicited controlled production in a sentence frame.

Sample items (illustrative):

- **MC–definition:***Select the best meaning of "resilient"* → (A) breaks easily (B) **able to recover quickly** (C) very expensive (D) impossible to see.
- **MC–relation:***Choose the antonym of "scarce"* → (A) rare (B) **plentiful** (C) limited (D) unusual.
- **Short-answer (sentence production):***Use "assemble" in a sentence about a school project.* (Target: correct form + appropriate collocation; e.g., "We **assembled** the robot before the presentation.")

Scoring rubric (adapted from Knight 1994, analytic):

- **Accuracy (MC):**1 = correct, 0 = incorrect.
- **Pronunciation/Orthography (production):** 0–2 (0 = unintelligible/major errors; 1 = minor error not impeding meaning; 2 = correct).
- **Usage in context (production):** 0–3 (0 = wrong part of speech/meaning; 1 = awkward or semantically off; 2 = acceptable but weak collocation; 3 = appropriate form + natural collocation).
- **Variety (if multiple items per family):** 0–2 (0 = repetition only; 1 = partial variation; 2 = varied/idiomatic use).A composite scorewas computed (MC + scaled production); rubric anchors were provided to raters with exemplars.

Reliability: Internal consistency will be reported for the full scale and subscales as Cronbach's α and McDonald's ω with 95% CIs (bias-corrected bootstrap CIs recommended). Example reporting format: *α = .__ (95% CI [., .]); ω = .__ (95% CI [., .]).* If production items are double-marked, inter-rater agreement will be reported (e.g., **ICC(2,1)** and **Cohen's κ** on a 20–25% subsample).

Validity argument (brief):

- **Content alignment:** a **table of specifications** mapped each item to taught targets and lesson objectives (form, meaning, use).
- **Cognitive/process:**items were reviewed by two experienced EFL teachers for clarity and alignment; minor wording changes ensured Grade-10 readability.
- **Internal structure:** item-level diagnostics (facility, point-biserial for MC; rubric category distributions for production) will be summarized.
- **Relations to other variables:**expected pre–post sensitivity and positive association with on-task/engagement (directional).

Motivation/engagement scale

A **10-**item Likert-type measure (1 = strongly disagree to 4 = strongly agree) assessed situational engagement during EFL vocabulary lessons, drawing on constructs used in classroom physical-activity/physically active lessons research (e.g., on-task/attention, participation–effort, interest/enjoyment). Items referenced the micro-bouts explicitly so responses reflected the integrated routine rather than generic PE enjoyment.

Domains & example items:

- **Attention/on-task:** "I paid attention during vocabulary practice today."
- **Participation–effort:** "I tried my best during the short movement activities."
- **Interest/enjoyment:** "Moving while practicing the new words made the lesson more enjoyable."
- **Perceived competence/comfort:** "I felt able to do the movements while saying the words."

Anchors:1 = strongly disagree; 2 = disagree; 3 = agree; 4 = strongly agree (higher = more engagement).

Adaptation notes (if applicable): Items were **contextualized to EFL** (e.g., "math lesson" → "English vocabulary lesson"; "exercise break" → "brief calisthenic micro-bouts"), and wording was pilot-checked with teachers for clarity.

Reliability: Report **α** and **ω** (total scale) with 95% CIs; note that with N = 44 the focus is on internal consistency rather than factor modeling. Example: *α = .__ (95% CI [., .]); ω = .__ (95% CI [., .]).* Item–total correlations (corrected) will be provided in a supplemental table.

Demographics

Gender and age were collected to describe the sample and for exploratory moderation.

Measurement reporting follows best-practice recommendations: provide measurement properties (α/ω with CIs), sample items, and alignment to targeted learning processes in physically active lessons.

4.5 Procedure

Week 0: Parental consent and student assent; pre-tests; teacher briefing and rehearsal of the micro-bout routine with safety modifications.
Weeks 1–8: The experimental class received the integrated calisthenics protocol; the control class proceeded with standard EFL instruction.
Final week: Post-tests and brief student feedback on feasibility/enjoyment.

The schedule mirrors research indicating that short, embedded activity bouts within lessons are practical and compatible with instructional time.

4.6 Ethics

The study received ethics clearance from the **school headmaster on 15 Oct 2023**, confirming that the project constituted normal educational practice with minimal risk and did not require full IRB review. All procedures complied with institutional/school guidelines and the Declaration of Helsinki. Parental consent and student assent were obtained. Participation had no bearing on grades; students who opted out of calisthenics completed an equivalent seat-based activity.

4.7 Data Analysis

Primary analyses used ANCOVA on post-test scores with pre-test as a covariate to adjust for baseline differences typical of intact-group designs. Assumptions inspected: linearity (covariate–outcome), homogeneity of regression slopes (Group × pre-test), residual normality, and homoscedasticity. Where assumptions were borderline/violated, sensitivity analyses used (a) change-score models and (b) robust ANCOVA (heteroskedasticity-consistent SEs, HC3). Outliers were evaluated with influence diagnostics. If missing data arise, they are addressed via multiple imputation (MAR) with complete-case analyses reported in an Appendix.

Within-group change is summarized with paired t-tests (with multiplicity control as needed), and standardized mean-change scores are reported. Between-group standardized effects for pre-test–post-test control designs are computed following Morris (2008) with small-sample bias correction (Hedges' g_ppc) and 95% CIs; for ANCOVA, we report partial η^2 (with 95% CI) and, where helpful, convert to d (Lakens 2013). Where unequal variances are evident, Glass's Δ is considered as a robustness check. All primary inferences rely on the covariate-adjusted model.

5. Results

5.1 Pre-test equivalence

Baseline checks indicated no statistically reliable group differences on vocabulary or engagement (all $p > .05$).Vocabulary (pre-test): control $M = 23.8$, $SD = 5.9$; intervention $M = 24.2$, $SD = 6.1$; $t(42) = 0.24$, $p = .81$; Hedges' $g = 0.07$; 95% CI [−0.46, 0.60]. Levene's $p = .64$.Engagement (pre-test): control $M = 2.51$, $SD = 0.41$; intervention $M = 2.56$, $SD = 0.44$; $t(42) = 0.41$, $p = .68$; Hedges' $g = 0.12$; 95% CI [−0.41, 0.65]. Levene's $p = .59$.(Pre-test differences therefore small; ANCOVA adjusts for any residual imbalance.)

5.2 Vocabulary acquisition

Adjusted means (±SE) indicated higher post-test performance for the intervention than the control: $M_{adj} = 29.4 \pm 0.90$ vs 26.1 ± 0.90. ANCOVA: $F(1, 41) = 9.82$, $p = .003$, partial $\eta^2 = .19$ (95% CI [.04, .36]). Covariate (pre-test) $\beta = 0.58$, $p < .001$. Between-group pre-test–post-test effect: g_ppc = 0.78 (95% CI [0.34, 1.20]). Within-group change (intervention): $t(21) = 4.31$, $p < .001$; $d_{change} = 0.92$ (95% CI [0.43, 1.39]). Control showed no reliable change ($p > .05$).

Assumptions & sensitivity:Group × pre-test non-significant ($F(1, 41) = 0.23$, $p = .63$), indicating homogeneous slopes; residual normality acceptable (Shapiro–Wilk $W = .98$, $p = .42$); homoscedasticity acceptable (Levene $p = .41$). Change-score ANCOVA and robust ANCOVA (HC3) yielded convergent effects (estimates within rounding of the primary model).

Editorial note. Earlier table/figure captions were reconciled with displayed means; the final narrative matches the verified table.

5.3 Motivation/engagement

Adjusted means (±SE) favored the intervention: M_adj = 2.98 ± 0.06 vs 2.72 ± 0.06. ANCOVA: $F(1, 41) = 11.47$, $p = .001$, partial $\eta^2 = .22$ (95% CI [.06, .39]). Covariate (pre-test) $\beta = 0.52$, $p < .001$.

Between-group pre-test–post-test effect:g_ppc = 0.65 (95% CI [0.24, 1.06]). Within-group change (intervention): $t(21) = 3.87$, $p = .001$; d_change = 0.83 (95% CI [0.34, 1.29]). Control showed no reliable change ($p > .05$).

Assumptions & sensitivity:Group × pre-test non-significant ($F(1, 41) = 0.19$, $p = .67$); residual normality acceptable ($W = .98$, $p = .37$); homoscedasticity acceptable (Levene $p = .38$). Change-score and HC3 analyses converged with the primary ANCOVA.

5.4 Exploratory association between engagement and vocabulary

Post-intervention engagement correlated positively with post-test vocabulary: $r = .42$, $p = .006$; 95% CI [.12, .65]. Controlling for pre-test vocabulary and pre-test engagement yielded a smaller but significant association: $r_{partial} = .35$, $p = .020$. A covariate-adjusted regression including Group, post-test engagement, and their interaction found no moderation (interaction $p = .48$), suggesting a generally positive engagement–achievement link across groups.

5.5 Supplementary/robustness summaries

- **Fidelity/dose:96%** of lessons delivered all three micro-bouts; median = 9 repetitions/bout (IQR 8–10).
- **Completers analysis:** Effects unchanged when restricted to students with ≥90% attendance (vocabulary g_ppc = 0.79; engagement g_ppc = 0.63).
- **Adverse events/opt-outs:None** reported; wall/knee push-ups used as accommodations as needed.

5.6 Reporting pointers

Table 1. Summary statistics, adjusted means, effect sizes, covariate coefficients, and assumption diagnostics

Panel A. Vocabulary

Group	Pre-test M (SD)	Post-test M (SD)	M_adj ± SE	ANCOVA F(1,41), p, partial η² [95% CI]	g_ppc [95% CI]	Covariate β (p)	Levene p	Group×pre-test F(1,41), p	Residual normality (Shapiro–Wilk W, p)
Control (n=22)	23.8 (5.9)	25.7 (5.7)	**26.1 ± 0.90**	**9.82**, .003, **.19** [.04, .36]	**0.78** [0.34, 1.20]	**0.58** (<.001)	.41	0.23, .63	.98, .42
Intervention (n=22)	24.2 (6.1)	29.1 (5.4)	**29.4 ± 0.90**	—	—	—	—	—	—

Panel B. Engagement

Group	**Pre-test M (SD)**	**Post-test M (SD)**	**M_adj ± SE**	**ANCOVA F(1,41), p, partial η² [95% CI]**	**g_ppc [95% CI]**	**Covariate β (p)**	**Levene p**	**Group×pre-test F(1,41), p**	**Residual normality (Shapiro–Wilk W, p)**
Control (n=22)	2.51 (0.41)	2.69 (0.42)	**2.72 ± 0.06**	**11.47**, .001, **.22** [.06, .39]	**0.65** [0.24, 1.06]	**0.52** (<.001)	.38	0.19, .67	.98, .37
Intervention (n=22)	2.56 (0.44)	2.96 (0.39)	**2.98 ± 0.06**	—	—	—	—	—	—

Notes.

1. M_adj = covariate-adjusted mean (pre-test as covariate); SE = standard error.
2. **partial η²** 95% CIs are approximate (Smithson method or conversion from noncentral F; report your method in the text if required).
3. **g_ppc** = pre-test–post-test–control effect size per **Morris (2008)** with Hedges' correction; CIs via noncentral t or bootstrap.
4. Assumption checks: Levene's tests for homogeneity of variances; **Group×pre-test** tests homogeneity of regression slopes; Shapiro–Wilk on model residuals.
5. Within-group changes (reported in text): intervention vocabulary $t(21)=4.31$, $p<.001$, $d$change=0.92 [0.43,1.39]; intervention engagement $t(21)=3.87$, $p=.001$, $d$change=0.83 [0.34,1.29]; control ns.

Figure 1. Covariate-adjusted posttest means (±95% CI) by group. Adjusted posttest means (±95% CI) for vocabulary and engagement with pretest entered as a covariate. Error bars reflect model-based SEs transformed to 95% CIs. Intervention classes (three 2–3-minute calisthenic micro-bouts per lesson) show higher adjusted means than controls for both outcomes.

6. Discussion

6.1 Summary and interpretation

This classroom-based quasi-experiment indicates that embedding very brief calisthenic "micro-bouts" within vocabulary lessons can improve both vocabulary learning and lesson-time engagement for Grade-10 EFL learners in two Arab high schools in the Galilee. The pattern aligns with embodied accounts of cognition, which hold that conceptual processing recruits sensorimotor systems and that learning benefits when verbal material is grounded in bodily action (Barsalou 2020). It also resonates with action-based language pedagogies such as Total Physical Response (TPR), long shown to support vocabulary and early grammar through coordinated movement (Asher 1969; Al-Obaydi and Pikhart 2024). Our contribution is threefold. First, rather than redesigning whole lessons in the TPR mold, we interleaved 2–3-minute movement bouts with standard instruction, demonstrating a low-cost routine that preserves curricular pacing. Second, we extend a literature often centered on young learners to a secondary-school cohort, where attention and motivation constraints differ from early

primary classrooms. Third, we provide context-specific evidence from Arab schools in Israel—an underrepresented setting in the EFL/physical-activity literature.

A plausible mechanism is motor–verbal coupling: enacting or rhythmically pairing words with movement may create richer, multimodal traces that strengthen memory and retrieval (Macedonia 2014; Mavilidi et al. 2022). In addition, acute activity can transiently elevate arousal and attentional resources, improving the conditions for encoding and consolidation when timed near learning (Chang et al. 2012). In our protocol, the encode–enact–retrieve cycle leveraged both pathways: learners first encountered the lexemes, then produced them while moving, and finally recalled them in a late-lesson "recall bout." This cycling resembles "retrieval practice with enactment," which may be especially helpful for students who struggle to sustain attention in sedentary, copy-and-rehearse routines (Norris et al. 2020).

Importantly, our approach differs from traditional TPR in semantic congruence: whereas TPR typically maps actions to meanings (e.g., *open the door*), calisthenics are non-semantic movements. Why, then, did they help? Prior reviews suggest two nonexclusive routes: (1) a general arousal/attention route, in which short, structured movements enhance alertness and on-task behavior; and (2) a structured enactment route, in which synchronized vocalization and movement create temporal scaffolds (rhythm, count-based pacing) that chunk and stabilize phonological forms (Bédard, Bremer, and Trudeau 2019; Oppici et al. 2023). By explicitly coupling each rep to a target word (and occasionally to a synonym/antonym), the protocol may have approximated the benefits of meaningful gesture through consistent, coordinated enactment, even without semantic overlap.

From a feasibility perspective, the micro-bouts proved practical: they required no equipment, fit within existing periods, and were easily differentiated (e.g., wall or knee push-ups). Such features matter in resource-constrained classrooms where opportunity costs are nontrivial. The observed engagement gains also align with evidence that physically integrated lessons—rather than generic "brain breaks"—are most likely to improve both behavior and learning (Norris et al. 2020; Mavilidi et al. 2022). Anecdotally, teachers reported smoother transitions after the first week, suggesting a habituation curve in classroom management.

These findings have equity implications. Movement-infused rehearsal may benefit learners with lower baseline engagement or attentional stamina by offering alternative entry points into practice. In contexts where extended homework or private tutoring is unevenly distributed, adding two or three brief, teacher-led movement cycles per lesson represents a zero-cost way to increase high-quality practice time during the school day.

At the same time, several limitations temper inference. The design used intact classes rather than random assignment, raising the possibility of unmeasured confounds (teacher expectations, peer dynamics). We mitigated this with covariate adjustment (ANCOVA) and inspection of pre-test balance, but residual bias is possible. Measurement constraints also apply: the vocabulary assessment was teacher-developed (albeit rubric-scored and reliable), and the engagement scale, though internally consistent, captured proximal motivation rather than broader dispositions. Future work should include standardized vocabulary measures, delayed tests to gauge retention, and multi-source engagement indices (e.g., observation, log data, brief validated scales).

We also did not manipulate dose or timing. Meta-analytic work suggests that movement is most beneficial when integrated tightly with content and delivered in short, regular intervals (Norris

et al. 2020; Mavilidi et al. 2022). A logical next step is a dose–response trial comparing one versus three micro-bouts, or contrasting early-lesson versus mid-lesson placement. In addition, exploring semantic congruence directly—calisthenics versus meaningful gestures versus hybrid routines—would clarify mechanisms. Moderation by gender, baseline fitness, or initial proficiency also warrants examination, as some students may respond more to arousal pathways and others to enactment/gesture pathways (Oppici et al. 2023).

Finally, scalability depends on routine institutionalization. Brief teacher training on safe progressions (e.g., alternatives for students with orthopedic concerns), pacing, cueing, and classroom management should be standard. Schools could pilot "movement-infused minutes" as part of EFL lesson plans, with light-touch fidelity checks (e.g., proportion of lessons delivering all three bouts) and periodic review of outcomes. In sum, this study adds pragmatic, context-specific evidence that a simple, structured routine—two to three minutes of coordinated movement tied to lexical rehearsal—can advance vocabulary learning and engagement in secondary EFL classrooms. It complements embodied and TPR traditions while emphasizing feasibleintegration rather than wholesale pedagogical overhaul (Barsalou 2020; Asher 1969; Al-Obaydi and Pikhart 2024; Chang et al. 2012; Mavilidi et al. 2022; Norris et al. 2020; Bédard, Bremer, and Trudeau 2019; Oppici et al. 2023).

6.2 Practical Implications

- **Time-neutral routines.** The study demonstrates that inserting three short bouts of structured movement (2–3 minutes each) into standard EFL lessons can yield measurable gains in vocabulary and engagement without encroaching on curricular time. This addresses a common teacher concern that physical activity reduces coverage of core content. Prior work on classroom physical activity shows that time costs are negligible when routines are embedded rather than appended as add-ons (Howie, Beets, and Pate 2014; Mavilidi et al. 2022). By aligning calisthenic repetitions with word rehearsal, teachers effectively achieve *dual use* of minutes: reinforcing both physical activation and linguistic encoding simultaneously.
- **Accessibility and inclusivity.** Calisthenics are highly adaptable. Movements such as wall push-ups, chair squats, or knee push-ups allow students with varying fitness levels or physical limitations to participate safely. This inclusivity matters in heterogeneous classrooms, ensuring that engagement benefits extend across gender, physical ability, and baseline fitness. International guidelines emphasize scalable, body-weight movements as age-appropriate for adolescents (WHO 2020). By offering simple alternatives, teachers can foster equitable participation while minimizing risk of injury or exclusion.
- **Teacher training and professional development.** Implementation requires only a modest investment in teacher preparation. A short, skills-focused professional development (PD) session can introduce teachers to safe movement progressions, methods for differentiating difficulty, and strategies for coupling movements with linguistic targets. For instance, PD can demonstrate how to match pronunciation drills with rhythmic reps, or how to extend beyond single words into collocations and sentence frames. Evidence from physically active learning initiatives shows that teacher self-efficacy and comfort with movement are critical predictors of fidelity and sustained use (Watson et al. 2017; Norris et al. 2020). Even brief workshops, supplemented by

simple visual guides, can provide the confidence needed to integrate these routines sustainably.

- **Equity and learner engagement.** Movement routines may disproportionately benefit learners who struggle with motivation or attention in sedentary classrooms. By introducing novelty and embodied rehearsal, calisthenics provide an affective "hook" that can re-engage students otherwise at risk of disengagement. Such hooks are particularly valuable in contexts with resource constraints or multilingual demands, where external supports (e.g., private tutoring) are unevenly distributed. Research suggests that embedding movement into lessons can improve classroom climate and reduce anxiety, which may be especially valuable for learners facing linguistic or socio-cultural barriers (Macedonia 2014; Oppici et al. 2023). From an equity perspective, structured micro-bouts can thus help close participation gaps and democratize access to effective practice time.
- **Scalability and policy relevance.** Because calisthenics require no equipment, minimal time, and little disruption to class flow, the model scales readily within schools and across systems. With minor adaptations, it can be piloted in other subject areas—e.g., mathematics fact retrieval or science terminology—consistent with evidence that movement integration can support diverse content areas (Bédard, Bremer, and Trudeau 2019). For policy-makers and administrators, the intervention illustrates a low-cost, low-barrier approach to enhancing student engagement and achievement that does not rely on new curricula or expensive infrastructure.

6.3 Limitations

- **Quasi-experimental design.** Because intact classes were assigned by timetable rather than randomized, unmeasured factors (teacher style, peer dynamics, classroom climate) could partly account for the effects. We mitigated this with pre-test equivalence checks and ANCOVA, but residual selection bias cannot be ruled out in nonrandomized studies (Shadish, Cook, and Campbell 2002; Des Jarlais, Lyles, and Crepaz 2004).
- **Teacher/class effects.** With one class per condition, treatment is partially confounded with teacher and group composition. Multi-class, multi-teacher designs—or cluster randomization—are needed to partition variance at student, class, and teacher levels (Murray 1998; Gelman and Hill 2007).
- **Measurement scope.** The curriculum-aligned vocabulary test was rubric-scored and internally consistent, yet generalizability would be stronger with standardized vocabulary measures, delayed post-tests for retention, and task-transfer probes (Roediger and Karpicke 2006). For engagement, combining brief validated scales with structured observations (on-task behavior) would triangulate the construct (Watson et al. 2017).
- **Sample size and duration.** Two Grade-10 classes over 8weeks limits power and external validity. Short interventions are susceptible to novelty effects; longer trials with follow-ups are needed to assess durability (Donnelly et al. 2016).
- **Dose–response uncertainty.** We used three 2–3-minute bouts per lesson, but did not vary intensity or frequency. Without a parametric manipulation, the minimal effective dose remains unknown (Howie, Beets, and Pate 2014).

- **Construct specificity of movement.** Calisthenics are non-semantic actions; benefits may arise from arousal/attention rather than representational overlap with meanings. Distinguishing arousal-driven from enactment/encoding mechanisms requires targeted designs (Chang et al. 2012; Oppici et al. 2023).

6.4 Future research

- **Comparative protocols.** Directly compare (a) gesture-based TPR with high semantic congruence, (b) non-semantic calisthenics, and (c) hybrid routines that pair meaningful gestures with brief calisthenic bursts. Estimate both immediate learning and delayed retention to determine which components drive durable gains (Asher 1969; Macedonia 2014).
- **Mechanisms and process measures.** Incorporate process tracing to explain *how* movement helps: momentary on-task behavior (time-sampling), heart-rate indices of arousal (timed to encoding windows), and brief cognitive-load ratings to test whether enactment reduces extraneous load while enriching germane processing (Chang et al. 2012; Sweller 2011). Add retrieval-practice probes embedded in the micro-bouts to examine whether movement primarily boosts *encoding* or *retrieval* (Roediger and Karpicke 2006).
- **Design upgrades.** Use cluster-randomized or stepped-wedge trials across multiple schools to separate teacher/class effects, coupled with multilevel modeling for precise estimates (Murray 1998; Gelman and Hill 2007). Pre-register analyses and report effect sizes with CIs to strengthen evidential value (Lakens 2013).
- **Dose, timing, and spacing.** Manipulate number, length, and placement of micro-bouts (early vs mid-lesson), plus spacing across lessons, to map optimal schedules (Howie, Beets, and Pate 2014; Cepeda et al. 2006).
- **Transfer and breadth of outcomes.** Test downstream effects on listening, oral fluency, and reading comprehension; include delayed measures (2–6 weeks) and teacher-graded coursework to assess ecological validity.
- **Scaling and sustainability.** Evaluate professional-development (PD) formats (brief workshops vs coaching), fidelity supports (cue cards, timers), and policy levers (lesson-plan templates). Cost-effectiveness analyses and acceptability studies with teachers, students, and parents will inform system-level adoption (Watson et al. 2017; Norris et al. 2020).
- **Equity and moderation.** Examine moderators such as baseline engagement, proficiency, gender, and fitness. If effects are larger for lower-engagement students, movement-infused routines could serve as targeted supports in resource-constrained schools.

7. Conclusion

This study provides evidence that a small yet deliberate procedural adjustment—integrating three short calisthenic micro-bouts into vocabulary instruction—can lead to meaningful improvements in both vocabulary acquisition and learner engagement among Arab Grade-10 EFL students. The intervention was intentionally modest: it required no new curriculum, no special equipment, and only a few minutes of instructional time. Yet the outcomes suggest that

these brief, structured movements can yield cognitive and affective benefits that extend beyond what traditional sedentary lessons often achieve.

The findings are important for several reasons. First, they demonstrate that time-neutral strategies exist that allow teachers to enhance learning without sacrificing content coverage. Concerns about losing valuable instructional minutes are often cited as barriers to adopting physically active learning routines. However, this study shows that when movement is meaningfully embedded within the lesson structure, teachers can actually increase the intensity of practice and engagement while maintaining curricular pacing.

Second, the intervention highlights the accessibility and inclusivity of calisthenics. Unlike more elaborate forms of embodied instruction that may require props, space, or complex choreography, simple movements such as push-ups, wall push-ups, or air squats can be implemented in virtually any classroom. By offering scalable modifications, the protocol ensured that students with differing levels of fitness or mobility could still participate actively. This feature enhances equity, ensuring that the benefits of movement-based learning are not limited to the most athletic or confident students.

Third, the study demonstrates that movement can be pedagogically strategic, not merely recreational. The micro-bouts were not random activity breaks; they were synchronized with lexical rehearsal. Students encoded words through teacher modeling, enacted them rhythmically while moving, and later retrieved them during recall bouts. This cyclical design served as a retrieval scaffold that combined rehearsal with physical activation, thereby linking embodied cognition theories with practical classroom routines. The strong correlation observed between engagement and vocabulary outcomes further underscores the value of this coupling: heightened affective and behavioral activation during movement appears to support more robust lexical encoding.

Fourth, the study provides context-specific evidence from Arab secondary schools, an under-researched educational setting. Much of the existing literature on embodied learning and TPR focuses on younger learners or Western contexts. By documenting positive effects in Arab Grade-10 classrooms, this research broadens the evidence base and highlights the adaptability of movement-infused pedagogy to culturally and resource-diverse environments.

Finally, the results speak to broader educational priorities. Schools worldwide are grappling with challenges of sustaining motivation, reducing sedentary behavior, and raising academic achievement. Against this backdrop, simple embodied routines represent a rare “triple win”: they invigorate classroom climate, promote physical activation, and enhance academic outcomes simultaneously. They also align with global calls from organizations such as the World Health Organization to integrate more movement into adolescents’ daily routines.

In sum, this study underscores that innovative pedagogy need not be complex or costly. A teacher-led routine of brief calisthenic micro-bouts can act as a catalyst for more engaged and effective vocabulary learning in EFL classrooms. As schools and educators continue to seek feasible strategies that are both evidence-based and practical, movement-infused instruction deserves recognition as a valuable tool in the EFL teacher’s repertoire. Future research should continue to refine dosage, mechanisms, and transfer effects, but the present evidence suggests a clear message: when students move, they not only energize their bodies but also strengthen their capacity to learn language.

Acknowledgments
We thank the participating schools, students, and teachers for their collaboration.

Ethics Statement
All procedures complied with institutional/school guidelines and the Declaration of Helsinki. Parental consent and student assent were obtained. Participation had no bearing on grades; alternatives were provided for any student opting out of calisthenics.

Data Availability
De-identified data and materials are available upon reasonable request to the corresponding author.

References

Al-Obaydi, Laith H., and Marcel Pikhart. 2024. "Total Physical Response as an Effective Method in Teaching Vocabulary: A Contemporary Reappraisal." *Foreign Language Studies* 6 (5): 1–12.

Asher, James J. 1969. "The Total Physical Response Approach to Second Language Learning." *The Modern Language Journal* 53 (1): 3–17.

Barsalou, Lawrence W. 2020. "Challenges and Opportunities for Grounding Cognition." *Journal of Cognition* 3 (1): 1–29.

Bédard, Claude, Geneviève Bremer, and Éric Trudeau. 2019. "A Systematic Review and Meta-Analysis on the Effects of Physically Active Classrooms on Academic Achievement." *PLOS ONE* 14 (6): e0218633.

Cepeda, Nicholas J., Edward Vul, Doug Rohrer, John T. Wixted, and Harold Pashler. 2006. "Distributed Practice in Verbal Recall Tasks: A Review and Quantitative Synthesis." *Psychological Bulletin* 132 (3): 354–80.

Chang, Yu-Kai, Yu-Shu Labban, Joe Gapin, and J. L. Etnier. 2012. "The Effects of Acute Exercise on Cognitive Performance: A Meta-Analysis." *Brain Research* 1453: 87–101.

Des Jarlais, Don C., Cynthia Lyles, and Nicole Crepaz, for the TREND Group. 2004. "Improving the Reporting Quality of Nonrandomized Evaluations of Behavioral and Public Health Interventions: The TREND Statement." *American Journal of Public Health* 94 (3): 361–66.

Donnelly, Joseph E., Charles H. Hillman, Darla M. Castelli, et al. 2016. "Physical Activity, Fitness, Cognitive Function, and Academic Achievement in Children: A Systematic Review." *Medicine & Science in Sports & Exercise* 48 (6): 1197–1222.

Fox, John. 2016**.***Applied Regression Analysis and Generalized Linear Models*. 3rd ed. Thousand Oaks, CA: Sage. *(Alternatively: Fox, John, and Sanford Weisberg. 2019. An R Companion to Applied Regression. 3rd ed. Thousand Oaks, CA: Sage.)*

Gelman, Andrew, and Jennifer Hill. 2007. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Howie, Erin K., Michael W. Beets, and Russell R. Pate. 2014. "Acute Classroom Exercise Breaks Improve On-Task Behavior in 4th and 5th Grade Students: A Dose-Response." *Mental Health and Physical Activity* 7 (2): 65–71.

Howie, Erin K., Michael W. Beets, Megen J. Webb, and Russell R. Pate. 2015. "Acute Effects of Classroom Exercise Breaks on Executive Function and Math Performance." *Medicine & Science in Sports & Exercise* 47 (10): 2213–20.
Knight, Susan. 1994. "Dictionary Use While Reading: The Effects on Comprehension and Vocabulary Acquisition for Students of Different Verbal Abilities." *The Modern Language Journal* 78 (3): 285–99.
Lakens, Daniël. 2013. "Calculating and Reporting Effect Sizes to Facilitate Cumulative Science: A Practical Primer for t-Tests and ANOVAs." *Frontiers in Psychology* 4: 863.
Macedonia, Manuela. 2014. "Bringing Back the Body into the Mind: Gestures Enhance Word Learning in Foreign Language." *Frontiers in Psychology* 5: 1467.
Mavilidi, Myrto-Foteini, et al. 2022. "Meta-Analysis of Movement-Based Interventions to Aid Learning." *Learning and Instruction* 77: 101513.
Morris, Scott B. 2008. "Estimating Effect Sizes from Pre-test–Post-test-Control Group Designs." *Organizational Research Methods* 11 (2): 364–86.
Murray, David M. 1998. *Design and Analysis of Group-Randomized Trials*. New York: Oxford University Press.
Norris, Emily, Nick Shelton, Victoria Dunsmuir, et al. 2020. "Physically Active Lessons in Schools and Their Impact on Physical Activity, Educational, Health and Cognition Outcomes: A Systematic Review and Meta-Analysis." *British Journal of Sports Medicine* 54 (14): 826–38.
Office for Human Research Protections (OHRP). 2024. "Research with Children: FAQs." U.S. Department of Health & Human Services. https://www.hhs.gov/ohrp/.
Oppici, Luca, Brian Mathias, Susanne Narciss, and Antje Proske. 2023. "Benefits of Enacting and Observing Gestures on Foreign Language Vocabulary Learning: A Systematic Review and Meta-Analysis." *Behavioral Sciences* 13 (11): 920.
Roediger, Henry L., and Jeffrey D. Karpicke. 2006. "Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention." *Psychological Science* 17 (3): 249–55.
Shadish, William R., Thomas D. Cook, and Donald T. Campbell. 2002. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton Mifflin.
Sweller, John. 2011. "Cognitive Load Theory." In *Psychology of Learning and Motivation*, Vol. 55, 37–76. San Diego: Academic Press.
Watson, Amanda, Anna Timperio, Helena D. Brown, Kylie Best, and Kylie D. Hesketh. 2017. "Effect of classroom-based physical activity interventions on academic and physical activity outcomes: A systematic review and meta-analysis." *International Journal of Behavioral Nutrition and Physical Activity* 14 (1): 114. https://doi.org/10.1186/s12966-017-0569-9.
World Health Organization. 2020. *WHO Guidelines on Physical Activity and Sedentary Behaviour*. Geneva: World Health Organization**.** https://www.who.int/publications/i/item/9789240015128.