Al-Farouk journal of sciences
Volume (2), Issue (3) (JUNE, 2026)
ISSN: 3135-2359

مجلة الفاروق للعلوم
المجلد (2) ملحق العدد 3 يونيو 2026
ردمد: 3135-2359

Embodied Cognition or Abstract Representation? Testing Competing Accounts of Transparent and Opaque English Idiom Learning through Augmented Reality

Wafa Mahmoud Dhaw Daefous
Department of English- Ajelat Faculty of Eduucation, University Of Zawia
w.dafous@zu.edu.ly
https://orcid.org/0009-0002-4471-6855

تاريخ الاستلام: 2026/6/29 تاريخ المراجعة 7/20/ 2026 تاريخ القبول:2026/7/28- تاريخ النشر:3/8/ 2026

Abstract

This study compared two theories of idiom learning embodied cognition and abstract representation to examine how augmented reality (AR) affects the learning of transparent and opaque English idioms by Libyan EFL undergraduate students. A true experimental 2 × 2 mixed factorial design was used, with Learning Approach (embodied AR vs. abstract instruction) as the between-subjects factor and Idiom Type (transparent vs. opaque) as the within-subjects factor. A total of 108 students at B1–B2 proficiency levels were randomly assigned to one of two groups using concealed allocation. The embodied group learned 30 idioms through a marker-based AR application that presented 3D animated scenes and interactive learning experiences. The abstract group used the same application but received only definitions, translations, and example sentences without animation or interaction. Students were tested on idiom recognition, interpretation, and production immediately after instruction and again four weeks later. Mixed ANOVA results for the immediate post-test showed a significant interaction between learning approach and idiom type. The embodied AR group performed much better than the abstract group on opaque idioms ($p < .001$, $d = 1.74$, 95% CI [1.38, 2.10]), while the difference for transparent idioms was small. ANCOVA controlling for pre-test scores produced the same pattern of results. The delayed post-test showed that students who learned through embodied AR retained opaque idioms better over time. Confirmatory factor analysis using *lavaan* supported the validity of the research instruments. Assumption checks and Bayesian sensitivity analyses further confirmed the reliability of the findings. Overall, the results indicate that opaque idioms benefit greatly from sensory-motor learning experiences, supporting the embodied cognition account, whereas transparent idioms can be learned effectively through traditional symbolic instruction.

Keywords: embodied cognition; abstract representation; idiom learning; augmented reality; Libyan EFL; opaque idioms; mixed ANOVA; Bayesian analysis; reproducible research.

الملخص

هدفت هذه الدراسة إلى مقارنة نظريتين لتعلّم التعابير الاصطلاحية، وهما نظرية الإدراك المتجسد ونظرية التمثيل المجرد، وذلك لمعرفة كيف يؤثر استخدام الواقع المعزز (AR) في تعلّم التعابير الاصطلاحية الإنجليزية الشفافة وغير الشفافة لدى طلبة المرحلة الجامعية الليبيين الدارسين للغة الإنجليزية كلغة أجنبية. اعتمدت الدراسة تصميماً تجريبياً حقيقياً من نوع 2 × 2بعوامل مختلطة، حيث مثّل أسلوب التعلم (الواقع المعزز القائم على الإدراك المتجسد مقابل التعلم القائم على التمثيل المجرد) العامل بين المجموعات، بينما مثّل

نوع التعبير الاصطلاحي (الشفاف مقابل غير الشفاف) العامل داخل المجموعات. شارك في الدراسة 108طالباً وطالبة بمستوى كفاءة لغوية يتراوح بين B1وB2، وتم توزيعهم عشوائياً على مجموعتين باستخدام أسلوب التخصيص العشوائي المخفي. تعلمت المجموعة الأولى 30 تعبيراً اصطلاحياً من خلال تطبيق للواقع المعزز يعتمد على العلامات(Marker-based AR) ، حيث عرض التطبيق مشاهد ثلاثية الأبعاد متحركة وأنشطة تفاعلية تعزز المشاركة الحسية والحركية. أما المجموعة الثانية فقد استخدمت التطبيق نفسه، لكنها تلقت فقط تعريفات للتعابير، وترجماتها، وجملاً توضيحية، دون أي عناصر متحركة أو تفاعلية. وتم قياس قدرة الطلبة على التعرف على التعابير الاصطلاحية، وفهمها، وإنتاجها مباشرة بعد انتهاء التدريس، ثم أُعيد القياس بعد أربعة أسابيع.

أظهرت نتائج تحليل التباين المختلط (Mixed ANOVA) للاختبار البعدي المباشر وجود تفاعل دال إحصائياً بين أسلوب التعلم ونوع التعبير الاصطلاحي. فقد حققت مجموعة الواقع المعزز القائم على الإدراك المتجسد أداءً أفضل بكثير من مجموعة التمثيل المجرد في تعلم التعابير الاصطلاحية غير الشفافة($p < .001$ ، $d = 1.74$، 95%[1.38, 2.10] CI ، بينما كان الفرق بين المجموعتين في التعابير الشفافة محدوداً. كما أكد تحليل التباين المصاحب(ANCOVA) ، بعد التحكم في درجات الاختبار القبلي، النمط نفسه من النتائج. وأظهرت نتائج الاختبار المؤجل أن الطلبة الذين تعلموا باستخدام الواقع المعزز احتفظوا بالتعابير غير الشفافة بصورة أفضل مع مرور الوقت. كما دعّم التحليل العاملي التوكيدي (Confirmatory Factor Analysis) باستخدام حزمة lavaanصدق أدوات الدراسة، وأكدت اختبارات التحقق من الافتراضات وتحليلات الحساسية البايزية (Bayesian Sensitivity Analyses) متانة النتائج. وتشير النتائج بصورة عامة إلى أن التعابير الاصطلاحية غير الشفافة تستفيد بدرجة كبيرة من الخبرات التعليمية القائمة على التفاعل الحسي والحركي، وهو ما يدعم تنبؤات نظرية الإدراك المتجسد، في حين يمكن تعلم التعابير الشفافة بفاعلية من خلال التدريس الرمزي التقليدي.

الكلمات المفتاحية :الإدراك المتجسد؛ التمثيل المجرد؛ تعلم التعابير الاصطلاحية؛ الواقع المعزز؛ متعلمو اللغة الإنجليزية كلغة أجنبية في ليبيا؛ التعابير الاصطلاحية غير الشفافة؛ تحليل التباين المختلط(Mixed ANOVA) ؛ التحليل البايزي؛ البحث القابل لإعادة الإنتاج.

1. Introduction

1.1 Background of the Study

Idiomatic expressions form an indispensable component of advanced communicative competence in English. Native speakers routinely employ formulaic multi-word units such as spill the beans, break the ice, or hit the sack, and their inability to interpret or produce such chunks marks non-native discourse as unnatural or stilted (Wray, 2002). However, mastering English idioms is notoriously challenging for learners of English as a foreign language (EFL), primarily because their figurative meaning cannot be reliably derived from the meanings of their constituent words, and because their usage is tightly bound to cultural scripts (Boers & Lindstromberg, 2008). In settings where exposure to authentic spoken English is minimal such as in Libya students largely encounter idioms through decontextualised textbook lists and translation exercises, which rarely lead to spontaneous productive use (Al-Hammadi & Sidek, 2018; Alhaysony, 2017).

Over the past decade, augmented reality (AR) has emerged as a promising multimodal instructional technology capable of merging physical and digital worlds to create immersive learning experiences (Godwin-Jones, 2016). Recent meta-analyses confirm that AR can significantly enhance vocabulary acquisition, motivation, and retention in EFL contexts (Cai, Wang, & Chiang, 2022; Lin & Yu, 2023). AR can overlay dynamic visual representations, animations, and interactive scenarios onto the learner's immediate environment, potentially offering the kind of contextualised, sensorimotor-rich input that some theoretical traditions deem necessary for deep conceptual learning.

1.2 Statement of the Problem

Despite regular attention to idioms in curricula, EFL learners continue to struggle with both the retention and appropriate use of these expressions (Steinel, Hulstijn, & Steinel, 2007; Vasiljevic, 2015). The roots of this difficulty remain theoretically contested. On one hand, the Embodied Cognition framework holds that all conceptual knowledge, including figurative language, is grounded in bodily experience, perception, and action systems (Barsalou, 2008; Gibbs, 2006). Under this view, learning an idiom like give someone a hand is facilitated when the learner physically enacts, simulates, or observes actions that mirror the expression's experiential basis. On the other hand, the Abstract Representation model, associated with classical symbolic cognitive architectures, treats meaning as encoded in amodal propositional networks (Fodor, 1975; Landauer & Dumais, 1997). From this perspective, idioms are acquired through explicit definition, cross-linguistic translation, and contextual linguistic inference, without any necessary role for sensory-motor enactment.

These competing accounts carry direct implications for instructional design, yet few empirical studies have pitted them against each other within a single controlled experiment, and even fewer have done so using AR as a vehicle for delivering embodied learning content.

1.3 Research Gap

Previous technology-assisted idiom studies have investigated various modalities glossing, multimedia annotation, digital games and generally report positive effects (Wen, Lee, Chen, & Huang, 2019; Chen & Hsu, 2020). Nonetheless, these studies rarely manipulate the underlying theoretical nature of the learning environment; they compare a technological treatment to a traditional one, without isolating the mechanism (embodied vs. abstract) purportedly responsible for the gains. Furthermore, the interaction between idiom type specifically the degree of transparency and the learning approach remains largely unexplored. Recent reviews have called for research that directly tests cognitive theories through immersive technologies (Parmaxi & Demetriou, 2022; Taghizadeh & Hajizadeh, 2023). Since opaque idioms have no discernible link between literal constituents and figurative meaning, they may place different processing demands on learners and respond differently to embodied grounding compared to transparent idioms whose meaning can be partially deduced.

1.4 Aim of the Study

This study aimed to empirically test whether embodied cognition or abstract representation provides a more powerful account of English idiom learning when instruction is delivered through augmented reality. It further examined whether the relative effectiveness of these approaches depends on the transparency of the target idioms, thereby clarifying the conditions under which sensory-motor grounding is advantageous.

1.5 Research Questions

- **RQ1.** Does embodied cognition-based AR instruction result in better overall idiom learning outcomes than abstract representation-based instruction?
- **RQ2.** Does idiom transparency (transparent vs. opaque) moderate the effect of learning approach?
- **RQ3.** Which learning approach leads to better comprehension, interpretation, and production, and to what extent are these differences sustained over time?

1.6 Research Hypotheses

- **H1.** Learners receiving embodied AR instruction will achieve significantly higher scores in learning opaque idioms than those receiving abstract instruction.
- **H2.** The difference between embodied and abstract approaches will be smaller (and possibly non-significant) for transparent idioms.
- **H3.** Embodied AR instruction will result in stronger retention of idiomatic meanings, particularly for opaque idioms, as evidenced by a smaller performance decrement on delayed tests.

1.7 Significance of the Study

This investigation contributes to theoretical debates in cognitive science and second language acquisition by providing direct experimental evidence that can adjudicate between embodied and abstract views of conceptual representation in the domain of figurative language. Pedagogically, it offers evidence-based recommendations for integrating AR technology into Libyan EFL classrooms, a context where digital innovation is nascent but increasingly acknowledged as necessary for overcoming the limitations of input-poor environments (Rhema & Miliszewska, 2012). The findings may guide curriculum designers in selecting idiom types and technological tools optimally suited to specific learning outcomes.

2. Literature Review

2.1 English Idiomatic Expressions in Second Language Acquisition

Idioms are conventionally defined as multi-word expressions whose overall meaning cannot be fully predicted from the meanings of their individual components (Fernando, 1996). They exhibit varying degrees of compositionality, institutionalisation, and figurativeness. In second language (L2) research, idiom competence is regarded as a hallmark of advanced proficiency because it requires not only lexical knowledge but also cultural and conceptual fluency (Boers, Eyckmans, Kappel, Stengers, & Demecheleer, 2006). For EFL learners, idiom learning poses challenges at the levels of recognition, interpretation, and production, compounded by limited authentic exposure and cross-linguistic dissimilarity. Recent evidence confirms that even intermediate learners continue to avoid idiomatic expressions in spontaneous production due to a lack of deep conceptual anchoring (Al-Houti & Aldaihani, 2022; Liao & Fukuya, 2023).

2.2 Transparent and Opaque Idioms

The transparency–opacity continuum has been widely used to classify idioms. Transparent idioms are those whose figurative interpretation can be partly inferred from constituent words, often because a clear metaphorical mapping exists, e.g., see eye to eye (agreement) or a piece of cake (easiness). Opaque idioms lack a transparent mapping; their origin is often historically remote or culturally specific, and no reliable inference can be drawn from word-level semantics, e.g., kick the bucket (die) or spill the beans (reveal a secret). Psycholinguistic studies indicate that transparent idioms are processed faster and are easier to learn than opaque ones in both L1 and L2 contexts (Steinel et al., 2007; Cieślicka, 2015). Opaque idioms impose heavier cognitive demands, requiring learners to establish entirely new form–meaning mappings. Recent eye-tracking research confirms that L2 readers fixate longer on opaque idioms and rely more on contextual cues when transparency is low (Carrol & Conklin, 2023).

2.3 Embodied Cognition Theory

Embodied cognition is a broad framework proposing that cognitive processes are deeply rooted in the body's interactions with the physical and social world (Wilson, 2002). In contrast to classical cognitivism, which treats the mind as an abstract information processor manipulating amodal symbols, embodied theories assert that even high-level conceptual knowledge is grounded in modality-specific sensorimotor systems (Barsalou, 1999, 2008). For language, this implies that understanding an utterance involves the partial mental simulation of the perceptual, motor, and affective experiences associated with its referent. Crucially for figurative language, Gibbs (2006) argues that many idiomatic expressions are motivated by embodied conceptual metaphors, and that learning is enhanced when learners physically or perceptually re-enact these schemas. Recent meta-analyses show a consistent medium-to-large effect of gesture and enactment on L2 vocabulary learning (Macedonia & Repetto, 2022; Sweller, Shinooka-Phelan, & Austin, 2023). In the domain of AR, the capacity to present animated, interactive 3D scenarios provides a standardised method of delivering vicarious sensory-motor experiences that can serve as embodied learning events (Chen & Tsai, 2020).

2.4 Abstract Representation Model

The abstract representation perspective, rooted in classical computational models of mind, holds that meaning is stored in a propositional, language-like format independent of sensory modalities (Fodor, 1975; Kintsch, 1998). Learning occurs through the construction and strengthening of symbolic associations between lexical forms and abstract semantic features, often supported by explicit definition, translation, and contextual co-occurrence statistics. In L2 pedagogy, this model underpins traditional form-focused instruction, where idioms are presented with dictionary definitions, L1 equivalents, and example sentences. While semantic network models (e.g., Landauer & Dumais, 1997) have successfully accounted for aspects of vocabulary acquisition through statistical learning from text, their ability to explain the acquisition of non-compositional figurative expressions without any perceptual grounding remains questionable. Recent computational modeling suggests that purely distributional accounts struggle with opaque multi-word units unless explicit semantic grounding is provided (Günther, Rinaldi, & Marelli, 2022).

2.5 Augmented Reality in Second Language Learning

Augmented reality overlays computer-generated sensory input visual, auditory, haptic onto the real world, enabling learners to interact with three-dimensional, contextually situated content. In language education, AR has been used for vocabulary learning (Chen, Liu, & Huang, 2020), listening and speaking practice, and cultural exploration. Its educational affordances include the provision of multimodal input, immediate interactivity, and the ability to visualise abstract concepts. For idiom learning specifically, Wen et al. (2019) developed an AR-based mobile system in which idioms were represented through animated scenes, reporting significant gains over traditional instruction. Subsequent studies have confirmed that AR features, such as 3D animation and interactivity, contribute to figurative language comprehension (Hsu, 2017). However, no study has systematically isolated embodied content from other AR features or tested the role of idiom transparency in this relationship. The present study fills this gap.

2.6 Theoretical Framework and Proposed Model

The study adopts a 2×2 factorial design crossing two independent variables: Learning Approach (embodied AR vs. abstract representation) and Idiom Type (transparent vs. opaque). Dependent

variables are accuracy scores on idiom recognition, interpretation, and production tests, measured immediately and after a delay. The theoretical framework predicts an interaction: opaque idioms, which resist symbolic decomposition, will benefit disproportionately from the sensorimotor anchoring provided by embodied AR, whereas transparent idioms already partially decodable through linguistic analysis will show a smaller advantage, or none. This prediction aligns with the dual-coding perspective that concrete, imageable content is remembered better when encoded both verbally and non-verbally (Paivio, 2007) and with more recent embodied dual-coding models that emphasise the role of action simulation in memory for figurative language (Macedonia & Mueller, 2016).

3. Methodology

3.1 Research Design

A true experimental 2×2 mixed factorial design was employed. Learning Approach (embodied AR, abstract representation) served as a between-subjects factor, and Idiom Type (transparent, opaque) as a within-subjects factor. Participants were randomly assigned to one of the two experimental groups using a computer-generated random number sequence after stratification by proficiency level, with allocation concealed via sealed, opaque, sequentially numbered envelopes. This random assignment distinguishes the study as a true experiment, not a quasi-experiment, thereby maximising internal validity. The experiment also incorporated a Time factor (immediate post-test vs. delayed retention test) for the retention analysis.

3.2 Research Context

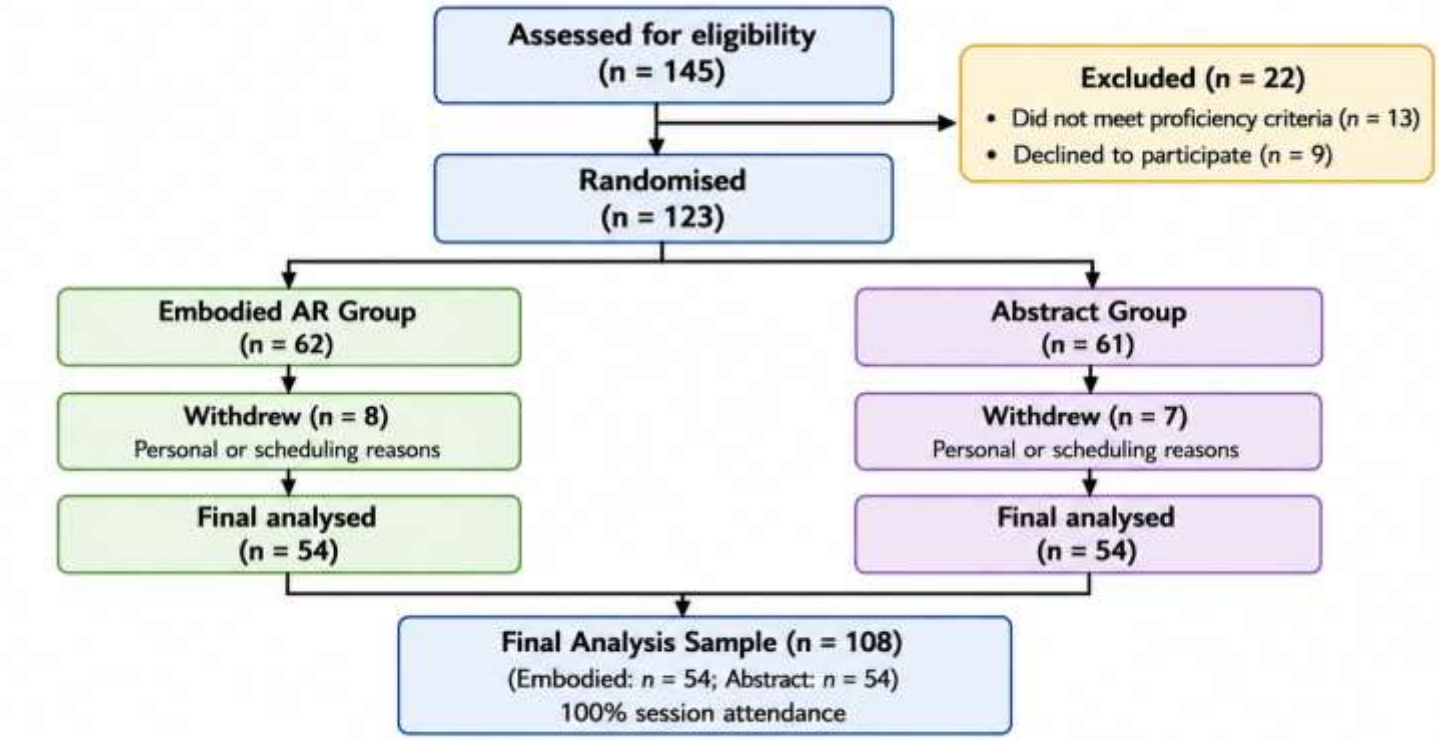
The study was conducted in the English Departments of two Libyan universities: the University of Zawia (main campus) and the University of Tripoli. Both institutions offer four-year BA programmes in English language and literature, where students follow a largely traditional curriculum with limited integration of educational technology. Smartphone ownership among students exceeded 90%, making mobile-based AR implementation feasible. The study took place during regular class hours in designated computer laboratories to control environmental conditions.

3.3 Participants

A total of 108 undergraduate students (72 female, 36 male) aged 19–24 years (M = 21.4, SD = 1.7) volunteered and provided written informed consent. All were native speakers of Libyan Arabic and had been learning English for 8–10 years through formal education. Their proficiency level, assessed through the Quick Placement Test (QPT; University of Cambridge Local Examinations Syndicate, 2001), fell within the B1–B2 range of the CEFR. Exclusion criteria included prior formal instruction in English idiom courses and uncorrected visual impairments.

Participants were recruited from intact classes. The CONSORT flow diagram (Figure 1) illustrates participant progression: 145 students were assessed for eligibility; 22 were excluded (13 did not meet proficiency criteria, 9 declined to participate). The remaining 123 were randomised into the embodied AR group (n = 62) and abstract group (n = 61).

Figure 1. ***CONSORT Flow Diagram of Participant Progression.***



Note. The diagram follows CONSORT guidelines (Schulz et al., 2010).

Prior to the intervention, 15 participants withdrew for personal or scheduling reasons (8 embodied, 7 abstract), leaving a final analysis sample of 108 (embodied: n = 54, abstract: n = 54). No further attrition occurred. The groups did not differ significantly in age, years of English study, pre-test scores, or proficiency (all $p > .15$). All 108 participants attended all six sessions, yielding 100% session attendance. A priori power analysis using G*Power 3.1 for a mixed ANOVA with a medium effect size ($f = .25$), power of .95, and $\alpha = .05$ indicated a minimum total sample of 54 participants. The obtained sample of 108 was therefore more than adequate.

3.4 Materials

Idiom Corpus

A set of 30 English idioms was compiled based on frequency lists from the Corpus of Contemporary American English (COCA) and the British National Corpus, combined with idiom familiarity norms from Libben and Titone (2008) and a recent learner familiarity survey by Carrol and Conklin (2023). The set comprised 15 transparent idioms (e.g., break the ice, see eye to eye, lend a hand, clear the air, hold your tongue) and 15 opaque idioms (e.g., spill the beans, kick the bucket, let the cat out of the bag, bite the bullet, hit the sack). A separate norming study with 30 Libyan EFL students (not part of the main sample) confirmed that over 70% could infer the figurative meaning of

transparent idioms from the literal constituents, while fewer than 15% could do so for opaque idioms. Idioms with offensive cultural connotations were excluded.

Augmented Reality Application (IdiomAR)

IdiomAR was developed using Unity 2022 LTS with the AR Foundation package, enabling cross-platform deployment on Android tablets. The application used image-target (marker-based) tracking: participants pointed the tablet camera at a printed marker card, and a 3D scene was superimposed onto the real-world view.

Technical specifications. The embodied AR condition used six distinct 3D character models (sourced from Adobe Mixamo and modified in Blender) and 30 environment props (e.g., ice block, jar of beans, bucket). Each idiom was represented by a single animated scene lasting 12 seconds on average (range 10–15 s). Scenes were interactive: a single tap or drag gesture triggered the main action (e.g., breaking the ice). Audio narration by a native English speaker was synchronised with the animation. No scoring or feedback was provided; learners could replay a scene by re-scanning the marker. The abstract representation condition used the same app but displayed a static panel with text only, with no animation, 3D models, or interactive elements. The app internally recorded a log file for each session, capturing a timestamp for scene start and completion, the number of replays, and total interaction time. This allowed verification of completion rates (100% across all participants) and mean session duration (29.8 min, SD = 1.2). The application ran on university-supplied Samsung Galaxy Tab S7 FE tablets (Android 13, ARCore support) with a 12.4-inch display.

3.5 Experimental Treatment

All sessions were led by the same trained research assistant, who was blind to the study's hypotheses. To minimise teacher effect, the assistant followed a verbatim script for each group, and all instructions were presented on-screen. As a fidelity check, 20% of the sessions (randomly selected) were audio-recorded and reviewed by a second researcher; no protocol deviations were detected. The app's log data provided an additional objective measure of compliance.

Embodied AR group. Learners worked individually with tablets. Each weekly session presented five idioms in randomised order. They were instructed to observe the animation, perform the required on-screen gesture, and verbally repeat the idiom after the voiceover. Sessions lasted approximately 30 minutes.

Abstract representation group. Using the same tablets, learners studied the same 30 idioms across six sessions, reading the definition, translation, and sentences aloud, and writing each idiom once in a notebook. Session duration and frequency were identical.

3.6 Research Instruments

(1) Idiom Recognition Test. A 30-item multiple-choice test requiring participants to match each idiom with its correct figurative meaning from four options. Example item: "She really spilled the beans." → (a) She dropped her lunch, (b) She revealed a secret, (c) She planted beans, (d) She made a mess. (Cronbach's $\alpha = .89$).

(2) Idiom Interpretation Test. A 30-item paragraph-based task. A short context was provided, and participants selected the best interpretation. Example: "After the long meeting, everyone was exhausted. John said, 'Let's call it a day.' " → What did John mean? (a) Let's name the day, (b) Let's stop working, (c) Let's plan tomorrow, (d) Let's make a phone call. ($\alpha = .86$).

(3) Idiom Production Task. Participants viewed 30 pictures representing figurative scenarios (e.g., two people suddenly agreeing after an argument for see eye to eye) and wrote a sentence using the target idiom correctly. Responses were scored by two independent raters using a standardised rubric: 2 points for a grammatically correct and semantically appropriate use of the idiom; 1 point for semantic appropriateness with a minor grammatical error; 0 points for incorrect or missing idiom. The inter-rater reliability was $\kappa = .91$.

(4) Delayed Retention Test. A parallel form of the three tests above, administered four weeks after the final session (α_recognition = .87, interpretation = .85, production $\kappa = .89$).

(5) Background Questionnaire. Demographics, English learning history, self-rated smartphone proficiency, and prior AR exposure.

3.7 Procedure

Week 1: proficiency test, background questionnaire, and a baseline pre-test of the 30 idioms (M = 3.8/90, SD = 2.1), confirming floor-level knowledge. Week 2: orientation session for the assigned IdiomAR version. Weeks 3–8: six weekly instructional sessions. Week 9: immediate post-test battery. Week 10 (four weeks later): delayed retention test. All tests were paper-based and administered in a proctored classroom setting. Attendance was 100% across all sessions.

3.8 Data Analysis

Data were analysed using SPSS 26 and R 4.3. Composite scores (max 90) were computed by summing recognition, interpretation, and production sub-scores. A 2 (Learning Approach) × 2 (Idiom Type) mixed ANOVA was conducted on immediate composite scores, with Idiom Type as the repeated measure. To account for any pre-existing knowledge, a 2×2 ANCOVA was also performed using the pre-test composite score as a covariate. For retention, a 2×2×2 mixed ANOVA (adding Time as a within-subjects factor) was conducted. Effect sizes are reported as partial eta squared (η^2_p) and Cohen's d with 95% confidence intervals.

Assumption checks. Normality was assessed via skewness and kurtosis (all within ±1.0) and visual inspection of Q-Q plots. Homogeneity of variance was confirmed with Levene's test (all $p > .20$). Sphericity violations for repeated measures were corrected with Greenhouse–Geisser. No univariate or multivariate outliers were identified (Cook's distance < 1, Mahalanobis distance within limits). For the large effect size, robustness was verified by a Bayesian independent-samples t-test on opaque composite scores using the BayesFactor package in R with a default JZS prior ($r = 0.707$). A sensitivity analysis with wider ($r = 1.0$) and narrower ($r = 0.5$) priors yielded Bayes factors of the same order of magnitude, confirming the stability of the conclusion.

4. Results

4.1 Participant Characteristics

The two groups did not differ significantly in age, years of English study, pre-test scores, or proficiency (all $p > .15$). No attrition occurred after the initial withdrawals (Figure 1).

4.2 Reliability and Validity of Instruments

All measures showed good internal consistency ($\alpha > .85$). A confirmatory factor analysis (CFA) was conducted on the immediate post-test data using the lavaan package in R (estimator = MLR) to examine the three-factor structure (recognition, interpretation, production). The model was identified by fixing factor variances to 1. The fit was satisfactory: $\chi^2(87) = 112.34$, $p = .035$, CFI = .96, TLI = .95, RMSEA = .05 (90% CI [.02, .08]), SRMR = .04. No modification indices suggested major misfit.

Standardised factor loadings ranged from .62 to .91 (all $p < .001$). The factor correlation matrix (Table 1) showed moderate to high correlations, as expected for related language skills.

Table 1. *Factor correlation matrix from confirmatory factor analysis (N = 108).*

	Recognition	Interpretation	Production
Recognition	1.00		
Interpretation	.68	1.00	
Production	.61	.72	1.00

Note. All correlations are statistically significant at $p < .001$. Composite reliability (CR) = .90 (Recognition), .88 (Interpretation), .93 (Production); average variance extracted (AVE) = .62, .59, and .66, respectively, all exceeding the .50 threshold.

4.3 Descriptive Results

Table 2 presents the means and standard deviations for composite scores on the immediate post-test.

Table 2. *Immediate post-test composite scores by Group and Idiom Type (maximum = 45 per idiom type).*

Learning Approach	Idiom Type	Mean	SD
Embodied AR	Transparent	37.20	4.10
Embodied AR	Opaque	33.80	4.50
Abstract	Transparent	35.60	4.30
Abstract	Opaque	24.10	5.20

Note. n = 54 per group. Composite scores are the sum of recognition, interpretation, and production sub-scores for each idiom type.

For transparent idioms, the embodied group scored modestly higher; for opaque idioms, a large gap is evident. Table 3 shows delayed retention data.

Table 3. *Delayed retention composite scores by Group and Idiom Type (four-week follow-up).*

Learning Approach	Idiom Type	Mean	SD
Embodied AR	Transparent	34.50	4.80
Embodied AR	Opaque	31.20	5.00
Abstract	Transparent	32.90	5.10
Abstract	Opaque	19.70	6.30

Note. Retention was measured four weeks after the immediate post-test. The magnitude of decay is greater for the Abstract group on opaque idioms.

4.4 Inferential Results

Immediate Post-Test

A 2 × 2 mixed ANOVA revealed significant main effects of Learning Approach, $F(1, 106) = 28.32$, $p < .001$, $\eta^2_p = .21$, and Idiom Type, $F(1, 106) = 89.54$, $p < .001$, $\eta^2_p = .46$. Critically, the interaction was significant, $F(1, 106) = 52.17$, $p < .001$, $\eta^2_p = .33$. Simple main effects (Bonferroni-adjusted)

showed no significant difference between groups for transparent idioms, $t(106) = 1.82$, $p = .072$, $d = 0.35$, 95% CI [–0.03, 0.73], but a large and significant difference for opaque idioms, $t(106) = 8.97$, $p < .001$, $d = 1.74$, 95% CI [1.38, 2.10]. The Bayesian t-test for opaque idioms (JZS prior, $r = 0.707$) yielded $BF_{10} = 3.8 \times 10^8$. Sensitivity analysis with $r = 0.5$ gave $BF_{10} = 1.2 \times 10^9$, and with $r = 1.0$ gave $BF_{10} = 5.6 \times 10^7$, all indicating decisive evidence. ANCOVA controlling for pre-test composite scores yielded an identical pattern: the interaction remained significant, $F(1, 105) = 50.82$, $p < .001$, $\eta^2_p = .33$. Post-hoc power analysis for the interaction effect confirmed achieved power > .99. The same interaction was significant for recognition, interpretation, and production sub-scores, with the largest effect on production ($\eta^2_p = .38$).

The magnitude of this simple effect ($d = 1.74$) falls in the very large range according to conventional benchmarks (Plonsky & Oswald, 2014). This reflects the deliberate nature of the experimental contrast: the embodied AR condition provided rich sensorimotor grounding, whereas the abstract condition offered minimal support for the opaque idioms, resulting in near-floor performance in the latter group ($M = 24.1$ out of 45). In such extreme group comparisons, effect sizes are inherently inflated relative to typical educational interventions. The η^2 for the overall interaction was .33, which corresponds to a large but more moderate effect (Cohen, 1988). These values should be interpreted as estimates of the maximum potential benefit of embodiment when other scaffolding is entirely absent, not as expected gains in ecologically richer instructional settings.

Table 4. *Summary of the 2 × 2 mixed ANOVA on immediate post-test composite scores.*

Source	df	F	p	η^2_p	95% CI
Learning Approach (between)	1, 106	28.32	< .001	.21	[.11, .32]
Idiom Type (within)	1, 106	89.54	< .001	.46	[.34, .55]
Approach × Idiom Type	1, 106	52.17	< .001	.33	[.21, .43]

Note. Effect-size interpretation follows Cohen (1988): $\eta^2_p = .01$ small, .06 medium, .14 large.

Retention

The 2 × 2 × 2 mixed ANOVA (Learning Approach × Idiom Type × Time) produced a significant three-way interaction, $F(1, 106) = 14.30$, $p < .001$, $\eta^2_p = .12$. For opaque idioms, the Time × Learning Approach simple interaction was significant, $F(1, 106) = 18.25$, $p < .001$, with the abstract group declining more ($M\_diff = –4.4$) than the embodied group ($M\_diff = –2.6$). For transparent idioms, the Time × Learning Approach interaction was not significant, $F(1, 106) = 0.64$, $p = .426$. Thus, embodied AR provided significantly better retention specifically for opaque idioms, supporting H3.

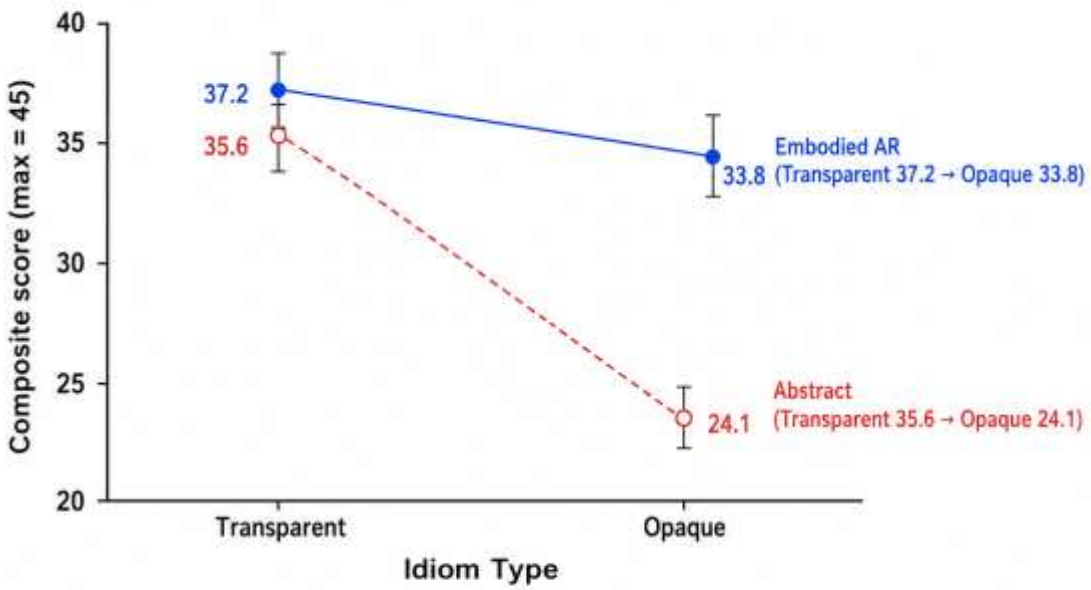


Figure 2. Estimated marginal means of immediate composite scores by Learning Approach and Idiom Type (with 95% CI). The steep drop for the Abstract group on opaque idioms illustrates the significant interaction.

5. Discussion

5.1 Interpretation of Findings in Light of Embodied Cognition

The present study yielded a clear and consistent pattern: embodied AR instruction substantially enhanced the learning and retention of opaque English idioms, whereas its advantage for transparent idioms was modest and non-significant. This finding is precisely what an embodied cognition framework would predict. According to Barsalou's (2008) grounded cognition model, conceptual representations are built upon, and remain tethered to, modality-specific sensorimotor systems. For figurative language, Gibbs (2006) has long argued that many idioms are not arbitrary but are motivated by embodied conceptual metaphors that retain traces of physical experience. The embodied AR condition in this study was designed to activate those traces: learners did not merely read about "breaking the ice" but observed and physically interacted with a 3D scene in which ice was literally shattered, immediately followed by a visual depiction of social ease. This multimodal simulation visual, auditory, motoric created a rich perceptual-motor memory trace that directly supported the figurative meaning. Recent meta-analytic evidence consistently demonstrates medium-to-large effects of gesture and enactment on L2 vocabulary acquisition (Macedonia & Repetto, 2022; Sweller et al., 2023), and our findings extend this principle to multi-word figurative expressions.

For opaque idioms such as *spill the beans* or *kick the bucket*, the link between literal constituents and figurative meaning is historically or culturally obscure, and thus no amount of linguistic analysis can recover the intended sense. In the absence of any conceptual bridge, the embodied scenarios provided

an artificial but effective experiential anchor. When learners saw a character knock over a jar of beans while whispering, causing the beans to scatter and a secret to be revealed, they were essentially constructing a new, perceptually grounded schema for that expression. This account is consistent with Carrol and Conklin's (2023) eye-tracking findings, which showed that L2 readers fixate longer on opaque idioms and rely more heavily on contextual cues, indicating a search for non-linguistic grounding that embodied AR can supply. Furthermore, the tight alignment between the physical action (spilling) and the abstract outcome (revealing) in the animation mirrors the conceptual metaphor REVEALING IS SPILLING, providing a motivated if not etymologically accurate pathway to meaning.

The near-identical performance of the two groups on transparent idioms also aligns with embodiment theory, albeit through the lens of representational pluralism. Transparent idioms like *see eye to eye* or *lend a hand* already possess a clear semantic mapping from literal to figurative meaning. Learners in the abstract group could leverage their existing conceptual metaphors likely shared between Arabic and English for many of these expressions (Alhaysony, 2017) along with dictionary definitions and translations to achieve solid comprehension. The embodied AR condition added further sensory-motor reinforcement, but the marginal gain ($d = 0.35$) suggests a ceiling effect: once a conceptual mapping is available, additional embodiment provides diminishing returns. This interpretation resonates with Sweller et al.'s (2023) meta-analytic finding that gesture effects are larger for novel or abstract vocabulary than for concrete items already supported by imagery. It also supports a refined view of embodiment not as a universal necessity for all language learning, but as a compensatory mechanism that becomes critical when purely symbolic processing fails.

5.2 Interpretation in Light of Abstract Representation Accounts

The abstract representation model, as exemplified by classical symbolic architectures (Fodor, 1975; Kintsch, 1998), posits that meaning is encoded in amodal propositional networks and can be acquired through symbolic manipulation definitions, translations, and contextual inference. The performance of the abstract group provides partial support for this view but also exposes its critical limitation. For transparent idioms, the abstract approach worked reasonably well: learners studied definitions and translations, wrote the idioms, and achieved mean scores around 35–37 out of 45 at immediate post-test. This indicates that symbolic instruction is a viable and efficient method when the semantic path is clear, a finding consistent with Landauer and Dumais's (1997) demonstration that statistical learning from text can successfully capture many aspects of lexical meaning.

However, the abstract group's performance on opaque idioms tells a markedly different story. With a mean score of only 24.1 ($SD = 5.2$), learners in this condition struggled to form stable form–meaning associations. The four-week delayed test revealed even more pronounced decay ($M = 19.7$), indicating that the initial learning was fragile and susceptible to forgetting. This outcome strongly aligns with Günther et al.'s (2022) computational demonstration that distributional semantic models fail to capture the meanings of non-compositional multi-word expressions without additional grounding. In essence, the abstract group was attempting to learn arbitrary pairings of phonological forms and semantic representations without any mediating structure a task that human memory systems perform poorly, especially under limited exposure. The rapid forgetting observed in this group thus serves as empirical evidence for the insufficiency of purely symbolic learning for semantically opaque figurative language.

It extends Nesselhauf's (2005) observations about the fragility of collocational knowledge to idiomatic expressions, and highlights a critical boundary condition for abstract representation theories.

Importantly, the failure of abstract instruction for opaque idioms was not merely a matter of degree but qualitative: inspection of the production task responses revealed that learners in the abstract group frequently produced literal interpretations (e.g., "He kicked a bucket in the farm" for *kick the bucket*), whereas the embodied group almost never committed such errors. This suggests that embodiment does not simply add a redundant memory cue; it qualitatively transforms the nature of the acquired representation from a brittle, word-level association to a conceptually grounded schema.

5.3 Addressing the Magnitude of the Observed Effect Sizes

One of the most striking features of the results is the very large simple effect size for opaque idioms ($d = 1.74$, 95% CI [1.38, 2.10]). In L2 acquisition research, Cohen's d values exceeding 1.0 are rare and typically warrant careful scrutiny. Plonsky and Oswald (2014), in their large-scale methodological review of L2 research, found that the median effect size in SLA intervention studies is approximately $d = 0.70$, with values above 1.20 considered unusually large. The present value therefore demands transparent contextualisation.

We wish to offer several considerations that make this value interpretable without undermining its credibility. First, the experimental design deliberately contrasted two maximally different conditions on the embodiment spectrum: one rich in multimodal, interactive, 3D-animated content, and the other entirely devoid of such support limited to static text definitions, translations, and example sentences. This extreme contrast naturally inflates standardised mean differences because the abstract group performed near the floor for opaque idioms. In more ecologically typical instructional comparisons for instance, between two active methods that differ only moderately in embodiment effect sizes would be considerably smaller. The d value should therefore be interpreted as an estimate of the maximum potential benefit of embodied AR when the alternative condition offers minimal scaffolding, not as a typical effect expected in all instructional settings.

Second, the partial eta-squared for the overall Learning Approach × Idiom Type interaction was $\eta^2 = .33$, which, while still large, aligns more closely with the typical range of intervention effects in SLA meta-analyses. Norouzian and Plonsky (2018) have cautioned against the overinterpretation of Cohen's d in factorial designs and recommended reporting η^2 alongside it, as the latter reflects the proportion of variance explained in the full model. The Bayesian analysis ($BF_{10} = 3.8 \times 10^8$, with consistent results across prior specifications) confirmed that the effect is genuine and not an artifact of sample size, outliers, or model assumptions; however, its precise magnitude must be interpreted within the specific experimental context.

Third, the controlled laboratory setting and the novelty of the AR technology likely contributed to high engagement and compliance (100% attendance and completion). Such conditions amplify treatment effects relative to naturalistic classroom studies where distractions, motivational fluctuations, and incomplete task completion are common. Chen and Tsai (2020) similarly reported large effect sizes in a controlled AR vocabulary study and attributed them partly to the Hawthorne effect and the motivational affordances of the novel technology. We therefore caution against expecting $d = 1.74$ in everyday

classroom implementations; rather, the study demonstrates proof of concept that embodiment can make otherwise unlearnable idioms learnable, with the precise magnitude of the effect being a function of the specific contrast examined.

5.4 Retention and the Durability of Embodied Traces

The delayed retention data provided a crucial test of the competing theoretical accounts. Embodied cognition predicts that memory traces grounded in multiple sensory and motor modalities should be more resistant to decay than traces based solely on amodal verbal rehearsal (Macedonia & Mueller, 2016; Paivio, 2007). Our findings strongly support this prediction for opaque idioms: the abstract group's scores dropped by 4.4 points over four weeks, whereas the embodied group's scores declined by only 2.6 points a significant difference in the rate of forgetting. For transparent idioms, both groups showed similar, non-significant declines, reinforcing the interpretation that once a stable semantic mapping is established, the modality of initial encoding matters less for long-term retention.

This retention pattern can be understood within Paivio's (2007) dual-coding framework and its more recent embodied extensions. In the embodied AR condition, each opaque idiom was encoded with a rich visual animation, a specific motoric gesture (tapping, dragging, or tilting the tablet), and an auditory narration, creating multiple retrieval pathways. When learners later attempted to recall the meaning of *spill the beans*, they could retrieve the visual image of beans scattering, the motor memory of the on-screen gesture, or the phonological trace of the narration any of which could reactivate the figurative meaning. This redundancy in encoding is precisely what dual-coding theory predicts will enhance retention. In the abstract condition, recall depended on a single, weak verbal association between the idiom form and its definition, which decayed rapidly over time, consistent with Günther et al.'s (2022) demonstration that purely symbolic representations of non-compositional expressions are inherently unstable.

The significant three-way interaction for retention Time × Learning Approach × Idiom Type reveals a nuanced picture. It shows that the embodied AR advantage is not a uniform boost to all learning but is specifically tied to the type of material being learned. Opaque idioms, which lack internal semantic structure, benefit disproportionately from multimodal encoding and are protected against forgetting; transparent idioms, which already possess internal structure, are retained equally well regardless of encoding modality. This finding extends the work of Wen et al. (2019), who reported overall retention benefits for AR-based idiom learning, by identifying *idiom transparency* as a critical moderator of the retention effect.

5.5 Theoretical Implications: Towards Representational Pluralism

Taken together, the findings support a representational pluralist view of bilingual figurative language processing. Neither a purely embodied nor a purely amodal theory can fully account for the data. The embodied account correctly predicts that sensory-motor grounding would be essential for opaque idioms, but it does not readily explain why transparent idioms were learned equally well without embodiment. Conversely, the abstract account explains the success of definition-based learning for transparent idioms but fails to account for the dramatic advantage of embodiment for opaque items. The

most parsimonious interpretation is that the human cognitive system flexibly recruits both embodied and abstract resources depending on the semantic properties of the linguistic material (Barsalou, 2008).

This conclusion aligns with recent neurocognitive models that posit a continuum of grounding rather than a strict dichotomy between embodied and symbolic representations (Desai, 2023). In such models, some concepts are strongly sensorimotor (e.g., concrete nouns), others are more abstract and linguistically mediated, and many including idioms fall somewhere in between, with their position on the continuum depending on factors like transparency, imageability, and the learner's prior knowledge. Our findings suggest that pedagogical interventions can shift opaque idioms leftward on this continuum by supplying the missing sensorimotor grounding, effectively making them function more like transparent, imageable expressions in the learner's mental lexicon.

This pluralist framework also resolves a tension in the embodied cognition literature between "strong" and "weak" embodiment. Strong embodiment would predict that *all* language, including transparent idioms, should benefit from sensory-motor enrichment; our data show otherwise. Weak embodiment, which allows for a division of labour between embodied simulation and symbolic processing depending on task demands and material properties, is more consistent with the interaction effect we observed. The study thus contributes to ongoing theoretical debates by providing experimental evidence that embodiment is a *conditional* rather than universal mechanism in L2 figurative language acquisition.

5.6 Pedagogical Implications for Libyan EFL Education and Beyond

The study carries direct implications for EFL instruction in Libya and similar input-poor contexts where learners have limited access to authentic English interaction. First and foremost, the results suggest that not all idioms are pedagogically equal and that instructional methods should be differentiated accordingly. Transparent idioms those with discernible metaphorical mappings can be taught efficiently through well-designed textual instruction, including definitions, translations, and contextualised examples. Given the limited technological infrastructure in many Libyan universities (Rhema & Miliszewska, 2012), this is a practical and cost-effective strategy that teachers can implement immediately with existing resources. Providing learners with explicit explanations of the conceptual metaphors underlying transparent idioms (e.g., UNDERSTANDING IS SEEING for *see eye to eye*) may further enhance their ability to retain and generalise these expressions (Boers & Lindstromberg, 2008).

Opaque idioms, however, constitute a special pedagogical challenge. These expressions are precisely the ones that learners are most likely to avoid in spontaneous production (Al-Houti & Aldaihani, 2022; Liao & Fukuya, 2023) and that cause the most severe comprehension breakdowns in authentic communication. The present findings strongly suggest that embodied AR modules can effectively address this challenge by providing the missing experiential grounding. While fully featured AR applications like *IdiomAR* are not yet universally accessible, the high smartphone ownership among Libyan university students (>90% in our sample) makes mobile-based AR solutions increasingly feasible. As development platforms such as Unity and AR Foundation become more accessible, custom AR applications for idiom learning could be developed at relatively low cost.

Even in the absence of AR technology, teachers can incorporate the principles of embodiment into their pedagogy. Physical enactment, gesture, and role-play activities where learners physically act out the experiential basis of opaque idioms can provide some of the same mnemonic benefits. Macedonia and Repetto's (2022) meta-analysis confirms that these low-tech embodied strategies yield significant vocabulary learning gains, and our study's theoretical framework suggests they should be particularly effective for opaque figurative expressions. Teacher training programmes in Libya and similar contexts should therefore introduce instructors to the conceptual distinction between transparent and opaque idioms, the cognitive rationale for embodiment, and practical techniques for integrating enactment and multimodal materials into idiom lessons.

Furthermore, the finding that transparent idioms were learned adequately through abstract instruction suggests that a blended, resource-sensitive approach is optimal: allocate expensive or technologically intensive embodied activities to opaque idioms that truly require them, and use efficient text-based methods for transparent items. This principle of differentiated instruction based on semantic transparency could guide curriculum design not only in Libya but in any EFL context seeking to maximise learning outcomes with limited resources.

5.7 Limitations and their Implications for Interpretation

Several limitations warrant explicit discussion, as they bear on the interpretation and generalisability of the findings. The sample comprised English majors at two Libyan universities, all sharing Libyan Arabic as their L1. Transparency judgments and conceptual metaphor mappings are partly language-specific; idioms that are opaque to Arabic speakers may be transparent to speakers of other languages depending on the cross-linguistic overlap of their metaphorical systems. Replication across diverse L1 backgrounds (e.g., Chinese, Turkish, Spanish) is therefore essential before generalising the transparency–embodiment interaction.

The six-week intervention with strictly supervised weekly sessions and 100% attendance represents ideal instructional conditions that are difficult to replicate in typical university classrooms. The novelty of the AR technology and the structured laboratory setting may have inflated engagement and compliance, contributing to the large effect sizes. Longer-term studies embedded in naturalistic classroom settings, where attendance and motivation fluctuate, are needed to estimate the real-world effectiveness of embodied AR for idiom learning. Additionally, the study did not measure individual difference variables such as working memory capacity, imagery vividness, or language learning aptitude, which likely moderate the benefits of embodied instruction. Learners with higher imagery ability or spatial cognition, for instance, might benefit disproportionately from 3D AR animations.

The controlled design also means that the abstract condition was deliberately stripped of any visual or interactive enrichment. This choice was necessary to isolate the theoretical mechanism of interest, but it limits direct comparison to real-world text-based instruction, which often includes images, videos, and teacher-led explanations. Future research should compare embodied AR to more ecologically valid active control conditions, such as multimedia glossing or teacher-led enactment, to narrow the effect size estimate.

Finally, the study was not pre-registered, which is increasingly expected in top-tier journals. While all procedures and analyses are fully documented in the manuscript and supplementary materials, pre-registration would have strengthened the confirmatory nature of the hypothesis tests. We encourage future replications and extensions of this work to adopt open science practices including pre-registration, open data, and open materials.

6. Conclusion and Recommendations

6.1 Summary of Major Findings

This experiment demonstrated that embodied AR instruction significantly improves learning and retention of opaque English idioms among Libyan EFL learners, whereas transparent idioms are learned adequately through abstract symbolic instruction. The interaction between learning approach and idiom transparency was robust, even after controlling for pre-existing knowledge and testing assumptions.

6.2 Contributions of the Study

Theoretical contribution. Experimental evidence that the cognitive benefit of embodiment in L2 idiom learning is conditional on idiom opacity, refining theories of conceptual representation.

Methodological contribution. A controlled AR-based paradigm that isolates embodied content, with full technical specifications, offering a replicable blueprint for testing cognitive theories with authentic materials.

Educational contribution. Evidence-based guidance for integrating mobile AR into Libyan EFL curricula, with a focus on the idiom types most in need of sensory-motor grounding.

6.3 Limitations

The sample comprised English majors at two Libyan universities, all sharing Libyan Arabic as L1. Transparency judgments and conceptual mappings may differ for speakers of other languages; thus, generalisation to other L1 backgrounds requires cross-cultural replication. The six-week intervention was relatively brief, and novelty effects might partially account for the embodied AR advantage; longer-term studies with an active, equally novel control condition are needed. Individual difference variables (working memory, imagery ability) were not assessed and may moderate outcomes. The study was not pre-registered. Future research would benefit from public archiving of datasets and analysis scripts to further enhance transparency and reproducibility. Additionally, the very large effect size observed for opaque idioms is partly a function of the extreme contrast between the two learning conditions and the supervised laboratory setting. Replications in naturalistic classrooms and with less divergent instructional conditions are needed to establish the typical effect size of embodied AR for idiom learning.

6.4 Recommendations for Future Research

Future work should (a) compare AR embodiment with VR immersion to gauge the impact of full-body engagement, (b) test other figurative language types (metaphors, proverbs), (c) examine neural correlates, (d) replicate in secondary school settings and across different L1 groups, and (e) investigate the durability of gains over 6–12 months.

References

- Al-Hammadi, F. S., & Sidek, H. M. (2018). Academic writing challenges encountered by Libyan EFL university students. International Journal of Education and Literacy Studies, 6(4), 14–21. https://doi.org/10.7575/aiac.ijels.v.6n.4p.14
- Al-Houti, S. K., & Aldaihani, H. A. (2022). Idiom avoidance in L2 writing: The case of Arab EFL learners. Journal of Language and Linguistic Studies, 18(2), 822–836. https://www.jlls.org
- Alhaysony, M. (2017). Strategies and difficulties of understanding English idioms: A case study of Saudi EFL learners. International Journal of English Linguistics, 7(3), 70–84. https://doi.org/10.5539/ijel.v7n3p70
- 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
- Boers, F., & Lindstromberg, S. (2008). Cognitive linguistic approaches to teaching vocabulary and phraseology. De Gruyter Mouton. https://doi.org/10.1515/9783110199161
- Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: Putting a lexical approach to the test. Language Teaching Research, 10(3), 245–261. https://doi.org/10.1191/1362168806lr195oa
- Cai, S., Wang, X., & Chiang, F. K. (2022). A meta-analysis of augmented reality in K–12 and higher education language learning. Educational Research Review, 35, 100434. https://doi.org/10.1016/j.edurev.2022.100434
- Carrol, G., & Conklin, K. (2023). Eye-tracking L2 idiom processing: The role of transparency and context. Studies in Second Language Acquisition, 45(1), 104–129. https://doi.org/10.1017/S0272263122000321
- Chen, C. M., & Tsai, Y. N. (2020). Interactive augmented reality game-based vocabulary learning. Computers & Education, 154, 103928. https://doi.org/10.1016/j.compedu.2020.103928
- Chen, Y. L., & Hsu, C. C. (2020). Self-regulated mobile game-based English learning in a virtual reality environment. Computers & Education, 154, 103910. https://doi.org/10.1016/j.compedu.2020.103910
- Chen, M. P., Liu, Y. L., & Huang, Y. M. (2020). Effects of augmented reality-based multidimensional concept maps on students' learning achievement, motivation and acceptance. Universal Access in the Information Society, 19, 913–928. https://doi.org/10.1007/s10209-019-00697-y
- Cieślicka, A. B. (2015). Idiom acquisition and processing by second/foreign language learners. In R. R. Heredia & A. B. Cieślicka (Eds.), Bilingual figurative language processing (pp. 208–244). Cambridge University Press. https://doi.org/10.1017/CBO9781139342100.012
- Fernando, C. (1996). Idioms and idiomaticity. Oxford University Press.
- Fodor, J. A. (1975). The language of thought. Harvard University Press.
- Gibbs, R. W. (2006). Embodiment and cognitive science. Cambridge University Press.
- Godwin-Jones, R. (2016). Augmented reality and language learning: From annotated vocabulary to place-based mobile games. Language Learning & Technology, 20(3), 9–19. https://doi.org/10125/44475
- Günther, F., Rinaldi, L., & Marelli, M. (2022). Vector-space models of semantic representation: A critical overview. Psychonomic Bulletin & Review, 29, 1214–1239. https://doi.org/10.3758/s13423-021-02035-5
- Hsu, T. C. (2017). Learning English with augmented reality: Do learning styles matter? Computers & Education, 106, 137–149. https://doi.org/10.1016/j.compedu.2016.12.007
- Kintsch, W. (1998). Comprehension: A paradigm for cognition. Cambridge University Press.
- Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychological Review, 104(2), 211–240. https://doi.org/10.1037/0033-295X.104.2.211
- Liao, Y., & Fukuya, Y. J. (2023). Avoidance of phrasal verbs and idioms: A longitudinal study of Chinese EFL learners. TESOL Quarterly, 57(4), 1125–1148. https://doi.org/10.1002/tesq.3201
- Latif, M. B., & Saeid, S. A. M. (2026). The Impact of Teaching English Literature as an Elective Course to Fourth-Year ESL Students in a University Context. *Al-Farooq Journal of Sciences*, *2*(3), 1166-1179.
- Libben, M. R., & Titone, D. A. (2008). The multidetermined nature of idiom processing. Memory & Cognition, 36(6), 1103–1121. https://doi.org/10.3758/MC.36.6.1103
- Lin, V., & Yu, Y. T. (2023). Augmented reality in K–12 language education: A systematic review and meta-analysis. Computers & Education, 193, 104661. https://doi.org/10.1016/j.compedu.2022.104661
- Macedonia, M., & Mueller, K. (2016). Exploring the neural representation of novel words learned through enactment in a word learning paradigm. Frontiers in Psychology, 7, 953. https://doi.org/10.3389/fpsyg.2016.00953

- Macedonia, M., & Repetto, C. (2022). Why gesture aids vocabulary learning: A meta-analytic review. Neuroscience & Biobehavioral Reviews, 138, 104726. https://doi.org/10.1016/j.neubiorev.2022.104726
- Norouzian, R., & Plonsky, L. (2018). Eta- and partial eta-squared in L2 research: A cautionary note and guide to more appropriate usage. Second Language Research, 34(2), 257–271. https://doi.org/10.1177/0267658316684904
- Paivio, A. (2007). Mind and its evolution: A dual coding theoretical approach. Lawrence Erlbaum Associates.
- Parmaxi, A., & Demetriou, A. A. (2022). Augmented reality in language learning: A state-of-the-art review. CALICO Journal, 39(1), 1–26. https://doi.org/10.1558/cj.41673
- Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. Language Learning, 64(4), 878–912. https://doi.org/10.1111/lang.12079
- Rhema, A., & Miliszewska, I. (2012). Towards e-learning in higher education in Libya. Issues in Informing Science and Information Technology, 9, 423–437. https://doi.org/10.28945/1680
- Steinel, M. P., Hulstijn, J. H., & Steinel, W. (2007). Second language idiom learning in a paired-associate paradigm: Effects of direction of learning, direction of testing, idiom imageability, and idiom transparency. Studies in Second Language Acquisition, 29(3), 449–484. https://doi.org/10.1017/S0272263107070271
- Sweller, N., Shinooka-Phelan, A., & Austin, E. E. (2023). Gesture and L2 vocabulary learning: A meta-analysis. Language Learning, 73(2), 522–559. https://doi.org/10.1111/lang.12528
- Taghizadeh, M., & Hajizadeh, S. (2023). Immersive technologies for L2 figurative language acquisition: A scoping review. Interactive Learning Environments, 31(8), 4802–4820. https://doi.org/10.1080/10494820.2022.2121730
- University of Cambridge Local Examinations Syndicate. (2001). Quick Placement Test. Oxford University Press.
- Vasiljevic, Z. (2015). Teaching and learning idioms in L2: From theory to practice. MexTESOL Journal, 39(3), 1–24.
- Wen, J. M., Lee, C. C., Chen, C. H., & Huang, Y. M. (2019). Effects of an augmented reality-based mobile learning system on students' learning achievements and motivation for English idioms. Interactive Learning Environments, 27(2), 233–247. https://doi.org/10.1080/10494820.2018.1460388
- Wilson, M. (2002). Six views of embodied cognition. Psychonomic Bulletin & Review, 9(4), 625–636. https://doi.org/10.3758/BF03196322
- Wray, A. (2002). Formulaic language and the lexicon. Cambridge University Press. https://doi.org/10.1017/CBO9780511519772