

VerbVenture—混合實境與人工智慧 輔助兒童英語搭配詞學習之設計與評估

徐加* 蔡承值** 潘則佑*** 吳宜儒**** 余能豪*****

國立臺灣科技大學設計系

* rossannahsu@gmail.com

***** 通訊作者 jonesyu@ntust.edu.tw

國立臺灣科技大學人工智慧跨域科技研究所

**m11252011@mail.ntust.edu.tw

***typan@ntust.edu.tw

臺北市立大學英語教學系

**** yjarielwu@gmail.com

摘 要

本研究旨在驗證一套應用於兒童英語搭配詞 (collocation) 學習的混合實境 (MR) 教學設計。此設計基於具身認知理論，旨在將日常環境轉化為沉浸式語言學習場域，並用以回應 EFL 兒童因母語影響及缺乏練習情境所面臨的挑戰。為在理想回饋條件下評估此教學設計的有效性，本研究開發了原型 VerbVenture，其學習流程分為兩階段：(1) 詞彙習得：透過人工智慧 (artificial intelligence, AI) 物件辨識，學習者能探索真實物件的英文單字與發音；(2) 搭配詞練習：系統播放情境動畫，並以動態提示引導學習者操作實體或虛擬物件演練動詞—名詞等搭配詞用法，從而強化語言與身體動作的連結。本研究對 14 位國小學童進行了前測—後測—延後測的受測者內實驗，比較操作實體與虛擬物件的學習差異。量化結果顯示，無論操作何種物件，VerbVenture 系統皆能帶來顯著的學習進步與記憶保留效果。雖然兩種互動模式在學習成效上未呈現統計上的顯著差異，但質性分析揭示了其互補價值：實體物件提供真實的觸覺回饋與物理特性，讓學習者感到更直觀、易於掌控，有助於建立初期的動作記憶。虛擬物件憑藉超越現實物理限制的新奇感，顯著提升了學習者的參與動機與想像力，且可能有助於長期記憶保留。本研究的貢獻在於，在理想互動條件下驗證了 MR 具身學習教學設計的成效，並基於實證歸納出一套以具身學習為核心的 MR 學習設計原則，為未來開發更具吸引力與成效的智能化語言學習工具提供具體的設計指引。

關鍵詞：混合實境、人工智慧、EFL 兒童語言學習、具身學習、情境學習

論文引用：徐加、蔡承值、潘則佑、吳宜儒、余能豪 (2025)。VerbVenture—混合實境與人工智慧輔助兒童英語搭配詞學習之設計與評估。《設計學報》，30 (4)，111-134。

一、前言

混合實境 (mixed reality, 以下簡稱 MR) 可以創造動態、互動和真實的學習環境, 讓語言學習者能夠參與如同真實生活般的場景, 並在豐富的情境中練習語言, 從而增強他們的理解和記憶力, 使其在語言學習領域具有巨大的潛力 (Lan, 2025)。同時, 人工智慧 (artificial intelligence, AI) 的發展也為語言學習帶來新的可能性, 現已有許多學習工具將 ChatGPT 等生成式 AI 整合, 為學習者提供即時口語溝通的真實感。隨著 MR 裝置普及至日常生活, 並與各項 AI 工具結合, 這些技術有機會成為新世代的語言學習工具。

本研究旨在驗證一套創新的 MR 具身學習 (Embodied Learning) 系統, 以應對 EFL (English as a foreign language) 兒童學習者在掌握搭配詞 (collocation) 時所面臨的獨特挑戰。所謂「具身學習」, 是指強調認知、身體感官與環境三者間的深度互動, 主張透過肢體動作與感官經驗來輔助抽象概念的內化。搭配詞指語言中某些單詞經常一起出現的語言模式, 這些單詞之間有著固定的聯繫 (Webb & Nation 2017)。這些密不可分的單詞搭配對於流利且自然的表達至關重要, 但由於它們的使用高度依賴語境, 且容易受到母語 (L1) 影響, 學習起來格外具挑戰性 (Bahns & Eldaw, 1993; Chan & Liou, 2005)。對於母語與英語差異顯著的兒童學習者 (如以中文為母語者), 由於詞彙使用習慣不同, 常常在英語中產生搭配語誤用的情形, 進而影響其語意理解與語言表達能力。為此, 本研究開發了原型 VerbVenture, 並前瞻性地探索了整合 AI 技術 (如環境、手勢及動作辨識) 的可能性, 期望將日常環境轉化為沉浸式語言學習場域。

基於具身認知理論, 本研究立足於以下核心假設: 身體經驗與知覺—運動活動在語言學習中扮演關鍵角色。認知科學與教育研究日益強調具身學習對知識獲取的重要性 (Cox et al., 2010; Madan & Singhal, 2012)。研究表明, 身體經驗、感知—運動活動以及與環境的互動對理解與記憶單字具有深遠影響。Madan 和 Singhal (2012) 的研究顯示, 學習者更容易記住能夠進行高互動性操作的具象物體詞語 (例如「相機」), 這項發現表明, 人們與物體互動的方式會影響對這些物體的記憶。此外, Chao 和 Martin (2000) 的研究發現, 觀察並命名物體會啟動與手部運動相關的大腦區域。這些研究結果強調, 語言學習不僅僅是大腦中的抽象過程, 而是與人們的身體經驗與感知—運動行為密切相關。透過身體與手勢動作積極參與環境互動, 學習者能夠實現更深層的理解並構建有意義的知識。雖然具身認知理論強調身體與環境互動的重要性, 但在混合實境環境中, 學習者可以同時與實體物件和虛擬物件互動, 這兩種互動方式可能會啟動不同的認知機制, 並產生不同的學習效果; 然而現行研究尚未明確比較這兩種互動方式在語言學習上的差異與相應的互動設計原則。

VerbVenture 正是基於此理論基礎, 整合 AI 與 MR 技術, 將日常環境轉化為沉浸式語言學習場域。兒童佩戴 MR 頭戴裝置, 開始有趣的學習之旅, 人工智慧物件識別 (object recognition) 技術能夠識別真實世界中的物品, 並觸發動畫場景和互動任務。這些引人入勝的提示透過具身互動指引學習者與物件進行互動, 加深搭配詞與相應行為之間的聯繫。本研究探討以下三個明確的研究問題:

1. 混合實境 (MR) 結合具身學習會有多大程度提升 EFL 兒童學習者的搭配詞習得成效?
2. 在具身學習情境中, 操作實體物件與虛擬物件的差異與特性?
3. 如何於混合實境中設計有效的引導機制, 以促進學習者的具身學習歷程?

本研究的核心目的並非重複驗證 MR 技術優於傳統教學法, 而是聚焦於一個更根本的問題: 當決定採用 MR 技術時, 教育者與開發者該如何設計有效的具身學習體驗? 本研究旨在透過比較 MR 環境中最核心的兩種互動媒介—實體與虛擬物件, 為 AI 與 MR 技術在語言教育中的應用, 提供一份基於實證的

「設計指南」。本文首先回顧相關文獻，接著詳述系統設計與實驗方法，然後呈現研究結果並進行討論，最後總結貢獻並提出未來研究方向。

二、文獻探討

儘管語言學習工具不斷創新，EFL 兒童在實際教育場域中，特別是那些學習成就較低的學生，仍面臨著顯著的學習挑戰。例如：本研究在前期英語教育專家的訪談中了解到，雖然學校會建議學習進度落後的學生參與免費的學習輔助課程，但由於非硬性規定，家長和學生的參與意願普遍不高。這些低成就的學生，往往本身對英語缺乏興趣，或學習態度較不積極，導致即使資源存在也難以有效利用。實際上，儘管有需要學習輔助的學生，但願意主動參與的比例卻相當有限。這突顯了傳統教學模式在吸引和激勵這些學生方面存在著困難，促使教育者尋求更具吸引力、能激發學習動機的創新方法，本研究所探索的 MR 輔助語言學習系統即是回應此一需求的可能解方之一。

本研究旨在透過 VerbVenture 系統展示英語搭配詞學習，探討與實體及虛擬物件互動的特性，並研究如何在 MR 中設計有效的動態提示，以促進具身學習，從而提升學生在搭配詞習得上的學習成果。為了釐清設計考量，本文首先回顧了有關 AR/MR 輔助語言學習之案例、虛擬實境中操作虛擬物件的學習效果，以及具身學習於實體與虛擬環境中應用的相關研究。

2-1 利用擴增實境（AR）與混合實境（MR）進行語言學習

近年來，擴增實境（Augmented Reality，以下簡稱 AR）與 MR 技術為語言學習領域的應用開創了新的可能性。這些技術構建了一個互動式學習環境，支持學習者在日常情境中獲取知識。在多媒體的輔助下，學習者能夠將真實環境中的物件與相關的虛擬動態提示（例如：文字、圖像、動畫與 3D 模型）相互結合，從而實現一種高效且自然的語言學習體驗（Hsu et al. 2023, 2024）。AR 與 MR 技術在激發學習動機與促進學習成效上扮演著關鍵角色（Hadid et al., 2019; Liono et al., 2021）。Yilmaz 等人（2022）設計了三款 AR 學習遊戲，以比較幼兒使用 AR 學習與傳統閃卡教學法在學習成效及態度上的差異；結果顯示，使用 AR 學習英語詞彙的學生，其學習成效顯著提升，且對 AR 輔助學習抱持更積極的態度。Ibrahim 等人（2018）開發的 ARbis Pictu 系統利用 HoloLens 將虛擬標籤即時疊加在真實物件上，支持多種語言；研究發現，與傳統閃卡教學相比，ARbis Pictus 能提供更有效率且愉悅的學習體驗。此外，AR 技術能同步呈現真實物件上的文字與圖像，使學習者能在真實情境中理解詞彙意涵，並增強記憶效果。VocabulARy 採用 HoloLens 將英語與日語詞彙、關鍵詞及 3D 動畫等動態提示疊加在真實物件上（Weerasinghe et al., 2022）；研究證明，在 AR 環境中學習詞彙較平板學習能更有效提升短期記憶表現。在 Draxler 等人（2020）研究中，學習者透過 AR 系統瀏覽環境中物件的標籤，並從中選取兩個實體物件；該系統不僅呈現德語詞彙，還測驗學習者如何以德語表達此兩物件之間的相對關係，讓學習者在日常生活情境中自然地學習。這些研究顯示，AR/MR 技術能夠創造情境化的學習環境，並支持學習者在真實情境中理解和應用語言。

2-2 虛擬實境（VR）中操作虛擬物件對語言學習的影響

虛擬實境（virtual reality，以下簡稱 VR）技術提供沉浸式的語言學習體驗，使學習者能夠在虛擬環境中與物件進行互動。VR 建構了一個多感官且互動的學習環境，有助於詞彙習得與記憶保持。Fuhrman 等人（2021）發現，在 VR 環境中執行與物件相關的動作（例如：使用掃把掃地）相比於進行無關動作或僅觀察物件，更能促進外語詞彙的學習；此結果顯示，有意義的動作型互動較僅觀察或執行無關動作，對新詞彙的學習更具效益。此外，透過動作與物件互動有助於減緩一週後詞彙記憶的衰退，表明 VR 環

境中有意義的動作有助於學習者長期記憶新詞彙。此外，在 VR 環境中，學習者可體驗到與真實世界不同的回饋機制。Ratcliffe 等人（2021）研究了兩組學習者在動詞編碼任務中的學習成效。他們開發了一套用於學習日語動詞的 VR 系統，其中一組透過基於動作的互動（操作與動詞相對應的虛擬物件）進行學習，另一組則採用基於手勢的互動（不操作物件）；結果顯示，採用基於動作編碼的學習者獲得了顯著更佳的學習成效，因為操作物件更貼近真實體驗，能提供更豐富的感官回饋並引發更深層次的認知處理。Macedonia 等人（2023）進一步證實，抓握虛擬物件能顯著提升低語言能力學習者的外語詞彙學習成效；此外，即使在缺乏觸覺回饋的情況下，抓握虛擬物件也比僅觀察或聆聽詞彙更有效於記憶保持。儘管具備上述優勢，VR 系統的設計仍須考量虛擬環境中的空間因素及其對學習成效的影響。Feng 與 Ng（2024）強調，虛擬物件的位置以及與學習者互動的頻率均會影響記憶表現。雖然該研究未直接探討物件操作，但結果顯示，在 VR 環境中與物件互動的頻率與物件的性質，對英語作為外語學習者的詞彙習得具有顯著影響；這一發現間接支持具身學習理論，指出頻繁且具體的物件互動可促進學習成效。綜合上述研究，操作或與虛擬物件互動於 VR 環境中不僅能提供更豐富的感官回饋，還能觸發更深層次的認知處理，並促進與虛擬環境的有意義互動，從而顯著提升學習成效，尤其在詞彙習得與記憶保持方面。然而，現有研究尚缺乏針對學習者操作虛擬物件之互動設計與行為設計的具體建議。

2-3 真實與虛擬環境中具身學習的效果

具身學習主張知識的建構是透過身體與環境間的互動而產生，強調感知運動經驗在學習中的重要性（Cox et al., 2010）。Schmidt 等人（2019）探討在課堂中融入身體活動對小學生注意力與外語詞彙習得的影響。他們設計了三種實驗情境，讓兒童學習法語動物名稱：（1）與詞彙意義相關的身體活動、（2）無關的身體活動，以及（3）無身體活動；結果顯示，具身學習組在記憶保持上優於無活動組，支持了具身學習能夠提升學習成效的假設。同樣地，Toumpaniari 等人（2015）探討了手勢與身體動作對學前兒童學習義大利語詞彙的影響；研究結果指出，執行與目標詞彙相對應的動作能帶來最佳的學習成效，凸顯出具身學習對知識習得的正向影響。Mathias 等人（2022）進一步探究在 12 歲與 14 歲學習者中，豐富化學習（結合圖片與手勢）對詞彙習得效果的差異；研究結果顯示，兩個年齡層皆從豐富化學習中獲益，顯示其具有跨發展階段的普遍適用性。Lindgren 等人（2016）與 Kelliher 等人（2009）則展示了 SMALLab 這一套系統之 MR 學習環境在課堂中的應用，研究結果顯示，全身互動模擬在概念理解、參與度與對科學的態度上表現更佳。MR 全身互動系統的優勢在於能夠積極讓學習者參與科學現象，藉由動態視覺效果與類比強調關鍵概念，進而促進概念理解。

2-4 研究缺口與定位

綜合上述文獻，過去已有研究比較擴增實境或虛擬實境與傳統教學方法的學習成效。例如：Ibrahim 等人（2018）的研究發現，與傳統閃卡教學相比，其開發的 ARbis Pictus 系統能提供更有效率且愉悅的學習體驗。Weerasinghe 等人（2022）也證明，在 AR 環境中學習詞彙，相較於使用平板學習，更能有效提升短期記憶表現。此外，Ebert 等人（2016）在 VR 環境中設計一套 Ogma 系統讓使用者在一個虛擬公寓中學習瑞典語詞彙，並透過使用者實驗與傳統的列表及閃卡學習進行比較，研究結果亦顯示 VR 學習組在一週後的長期記憶保留率顯著更高，且使用者普遍認為 VR 方法比傳統方法更有效、更有趣。然而，這些研究主要聚焦於比較新科技與傳統媒介的差異，針對在具身學習情境下，使用虛擬物件與實體物件的學習效果，卻尚未有研究進行深入比較。此外，虛擬與實體物件在材質、重量、觸感與操作回饋上存在本質上的不同，這也引發思考以下問題：這兩種互動媒介所帶來的學習成效是否會有差異？以及在互動設計上應注意哪些細節以優化學習體驗？

基於此，現有研究尚有以下缺口：

1. 雖然 AR／MR 在語言學習中的效益已被證實，但多數研究集中詞彙習得，較少探討搭配詞學習；
2. 在具身學習情境下，對於 MR 環境中實體與虛擬物件互動的比較研究有限；
3. 缺乏針對 MR 環境中引導設計的探討，尤其是如何支持具身學習。

本研究旨在設計有效的英語搭配詞學習流程與動態提示，促進具身學習並提升搭配詞學習成效。基於上述見解，本研究設計了 VerbVenture 系統，期望推動 MR 環境中具身學習的發展。並透過比較與實體及虛擬物件的互動，建立有效的互動與視覺提示設計原則，以彌補現有研究在該議題上的研究空白。

三、系統原型設計

VerbVenture 的設計是基於 Mayer (2002) 的多媒體學習認知理論，該理論強調學習設計應與人類認知機制相符。本研究特別著重於現實搭配詞的具身學習，旨在將動詞－名詞以及動詞－介系詞－名詞搭配與相關的身體活動無縫整合。此設計原則源自於具身認知理論，認為透過直接經驗與環境互動能夠建構更穩固的知識結構。系統設計遵循以下核心原則：

1. 多感官參與：整合視覺、聽覺與動作元素，創造沉浸式學習體驗；
2. 情境化學習：在真實情境中呈現語言元素，促進理解與應用；
3. 主動參與：鼓勵學習者透過手勢、動作積極參與學習過程；
4. 即時回饋：提供即時視聽回饋，強化學習動機與成就感。

這些原則共同支持 VerbVenture 的教學目標：模擬自然語言習得過程，讓兒童在探索現實物件的同時，學習並應用英語搭配詞。

3-1 互動流程

VerbVenture 的核心架構結合了手勢、語音、物件、動作辨識技術與 MR 視覺呈現，創造流暢且具沉浸感的互動學習體驗。圖 1 展示了互動流程，包含以下關鍵階段：

1. 探索與選擇：學習者可自由探索環境中物件，並透過手勢選取，被指向的物件會以黃色外框線做為提示，並在選取得觸發單字泡泡，從而促進學習者的參與與感知體驗。
2. 詞彙習得階段：當學習者選擇物件後，系統會觸發單字泡泡，介紹目標單字的發音並提示學習者進行複誦，以此方式幫助練習與記憶。系統利用語音辨識技術驗證學習者發音的正確性。當學習者正確複誦後，便會進入搭配詞學習階段，進一步學習物件的應用方式。
3. 搭配詞習得階段：系統透過情境動畫引導學習者進行動詞－名詞或動詞－介系詞－名詞的搭配學習，進而凸顯關鍵互動點。透過操作實體或虛擬物件，以鞏固語言學習的內容與肢體動作之間的連結。此階段系統整合手勢與動作辨識，以判斷學習者進行正確的對應動作。
4. 回饋與關鍵互動：系統回饋設計包括：透過指向手勢進行物件選取、藉由動態提示促進詞彙習得，以及以動作為基礎的搭配詞練習。每個關鍵互動均遵循 Mayer 的多媒體學習認知理論，減少冗

餘資訊，並採用雙通道（視覺與聽覺）呈現，針對不同呈現方式提供適切內容。例如：文字資訊以語音方式傳遞，而動畫與動態提示則引導學習者關注視覺內容。此方法能降低認知負荷、促進主動參與，並提升情境理解。完成任務後，學習者會獲得虛擬獎牌與獎盃，強化學習者的成就感。

Hsu 等人（2024）基於上述流程探索了 MR 結合具身互動之可能性，本研究進一步開發了原型系統以進行後續實驗，主要步驟如下所示：首先，兒童戴上 MR 頭盔後，透過指向實體物件的方式探索互動物件（如圖 1a）。接著，一旦選定實體物件，該物件上方便會出現單字泡泡，並播放目標單字的語音；同時，系統也會以音效與視覺提示，提醒兒童複誦（如圖 1b）。在兒童正確複誦目標單字後，泡泡便會消失。隨後，系統會播放一連串具情境脈絡的動畫，為兒童提供直覺式的虛擬情境，讓他們實際演繹搭配詞的行為。例如：目標單字是「umbrella」時，可依序練習「put up」、「fold」、「roll」等動名詞搭配。情境動畫「烏雲密布，接著下雨」，此時系統會以動態提示凸顯可互動的實體物件（雨傘），提示兒童拿起雨傘（如圖 1c）。在該情境中，系統會顯示打開雨傘的動態提示，引導兒童執行同樣的動作（如圖 1d）。之後，顯示包含「put up the umbrella」的搭配詞泡泡，與單字泡泡流程相同，輔助兒童聆聽搭配詞語音並複誦搭配詞。當系統辨識到正確的動作與複誦後，便會進入下一個情境動畫「雨停後出現彩虹」，透過動態提示引導兒童收起雨傘與複誦搭配詞；接著「雲散後出現太陽」，提示兒童將雨傘捲起與複誦搭配詞。每完成一次搭配詞學習，可獲得一枚獎牌（如圖 1e），並在完成該物件單字所有搭配詞後得到一座獎盃（如圖 1f）。最後，系統會以動態提示提醒兒童將物件歸位，即結束本次學習流程。



圖 1. VerbVenture 原型中的互動流程

(a) 兒童戴上 MR 頭盔後，透過指向實體物件的方式探索互動物件；(b) 選定物件後，該物件上會出現單字泡泡並播放語音，同時透過音效與視覺提示引導兒童進行複誦；(c) 兒童正確複誦後，泡泡消失，系統接續播放情境動畫（如下雨了要開傘），引導兒童使用物件進行互動；(d) 系統同步顯示動作提示（如開傘動畫），引導兒童執行對應動作；(e) 當兒童做對動作並唸出正確搭配詞，可獲得一枚獎牌，並出現下一個搭配詞的情境動畫（如彩虹出現要收傘）；(f) 完成該物件單字的三組搭配詞後得到一座獎盃。接著再探索其他物件（重覆 a 步驟）。

（圖片來源：Hsu et al., 2024）

3-2 技術整合

VerbVenture 使用 Meta Quest Pro 為 MR 平台，並結合了多項 AI 技術，用以支持情境式與具身學習。本研究採用 Meta 的互動 SDK 與手部追蹤技術，利用簡單的指向手勢觸發物件單字。為了自動辨識周遭物件，本研究採用了 segment anything model 2 (SAM2) 先對圖像進行物件分割，再使用 YOLOv11 進行物件辨識。SAM2 的資料引擎結合了模型輔助註解、自動遮罩生成以及人工驗證等策略，顯著提升了資料收集的效率與品質，進而使得在 MR 環境中對日常物件的辨識更為精確且自然（Ke et al., 2023; Ravi et al., 2024）。由於每個日常物件均需提供示範發音，而學習者在重複過程中正確發音則為推進學習流程之互動要求之一，因此本研究利用 Meta 的 Voice SDK（由 Wit.ai 支援）實現文字轉語音（TTS）及語音轉

文字 (STT) 的功能。採用文字轉語音而非預錄音檔，旨在符合隨時隨地彈性學習的理念。系統將學習者的語音轉換為文字，並與日常物件的文字表示進行比對；若學習者複誦錯誤，系統將重播音訊，引導其再度聆聽並重複。正如 Kumar 等人 (2012) 與 Yaniafari 等人 (2022) 所強調的，語音辨識在語言學習環境中對發音改善具有正向影響。在搭配詞習得階段，本研究運用情境動畫與動態提示，指引學習者在真實情境中練習動詞－名詞與動詞－介系詞－名詞的搭配。此方法根據 Halbig 與 Latoschik (2024) 的觀點，認為虛擬環境中的空間存在感與具身性密切相關，共享影響兩者體驗的共同提示。大腦會根據預期與實際刺激之間的差異來評估虛擬環境的真實感，因此設計符合使用者期望的提示在 MR 環境中極為關鍵。例如：情境動畫與動態提示的呈現，以及虛擬物件的互動方式，都應與現實世界原則高度契合。這種一致性可降低學習者的認知衝突，從而增強空間存在感與具身性。為此，本研究設計了情境動畫以模擬動詞－名詞與動詞－介系詞－名詞搭配情境，並以動態提示強調互動重點，指引學習者的行動。這些提示不僅能聚焦學習者注意力，還確保其行為與預期語言練習相符。本研究進一步結合 AI 技術，採用 MViTv2 模型進行即時動作識別，這是 Hsu 等人 (2024) 的原型中所沒有整合的部分。MViTv2 模型在影像分類、物件檢測和視頻識別任務上展現出卓越性能，本研究利用 RGB 攝影機來捕捉學習者視角數據，並根據搭配詞動作構建自訂數據集訓練 MViTv2 模型，以支持具身學習之互動歷程。由於研究進行當下 Meta 規範的限制，無法直接從 Meta Quest Pro 擷取 passthrough 攝影機數據（即真實環境影像）。為了解決此問題，本研究在 Meta Quest 3 上連接外接網路攝影機，並將捕捉到的網路攝影機影像傳至 PC，進行 YOLOv11 與 SAM 2 物件辨識，處理結果再回傳至 Meta Quest 3。

四、實驗設計

本研究招募了 18 名受測者，以評估 VerbVenture 系統在促進 EFL 兒童學習搭配詞之成效，並探討實體與虛擬物件互動在操作特性與學習成果上的差異，研究同時針對實體與虛擬物件互動過程中的具身學習行為及關鍵設計考量進行分析，最後歸納具身學習之互動設計原則。

4-1 研究假設

本研究假設源自於先前研究的發現，特別是 Fuhrman 等人 (2021) 與 Ratcliffe 等人 (2021) 關於物件互動對認知處理的影響，以及 Ibrahim 等人 (2018) 與 Draxler 等人 (2020) 關於視覺引導在語言學習中的作用。基於上述見解，本研究提出下列假設：

- H1. 與實體物件互動將比與虛擬物件互動產生更佳的學習成果。
- H2. 情境動畫與動態提示能有效吸引學習者注意力、提升學習者對語義與情境用法的理解。

4-2 受測者

本研究採用受測者內設計，招募 18 位國小五年級非英語母語學生。參與者選擇基於以下標準：

1. 未曾使用過 MR 進行語言學習，確保無先前經驗影響；
2. 前測評估中對目標搭配詞知識有限，確保學習效果來自實驗本身；
3. 具有相似的英語程度，減少個體差異的影響。

研究在臺北市立大學附設實驗國民小學進行，實驗前，本研究先進行前測評估受測者對 MR 系統中呈現之目標英語搭配詞的陌生度，以篩選具有相似英語程度的受測者。前測包含 12 項文意選擇題（如表

1)，並由班導師在課堂中進行施測。前測成績滿分 12 分中，低於 4 分者才被納入本研究受測者 ($N=18$)。為減少潛在之順序效應，本研究採用平衡設計，並以隨機方式將受測者分配至不同的實驗條件序列中。

本研究參與均屬自願，所有受測者皆於實驗前提供知情同意書。本研究已通過國立臺灣大學行為與社會科學研究倫理委員會審查（倫委會案號：202405ES026）。所有受測者皆被告知有權隨時退出研究，且此舉不會對其學業成績產生任何影響。

表 1. VerbVenture 中的 12 組搭配詞的情境動畫與動態提示

					
Put up the umbrella.		Fold the umbrella.		Roll the umbrella.	
					
Pick up the phone.		Answer the phone.		Put down the phone.	
					
Take the bottle.		Open the bottle.		Fill the bottle.	
					
Dress the doll.		Hug the doll.		Store the doll.	

4-3 實驗設計

4-3.1 單字與搭配詞設計

本實驗所採用之教學材料是根據臺灣國小英語課程指導方針設計，並在詞彙選擇上與英語教育專家充分討論，確保所選詞彙對高年級學生具有適當的學習挑戰性。本研究最終選定四個具代表性的英語物件名詞：umbrella、phone、bottle 與 doll。每個名詞均搭配三個常見動作，共形成 12 組動詞—名詞與動詞—介系詞—名詞搭配詞，如表 1 所示。本研究教學材料融入 12 項以動作為基礎的任務設計，透過具身操作練習搭配詞，以促進學生的語詞理解與記憶。在實驗設計上，亦特別重視學習安全與兒童的身心健康，雖然現有研究尚未指出混合實境裝置對兒童視力或認知發展有直接危害，本研究仍採取嚴格措施控制使用時間，每次配戴 Meta Quest Pro 的時間均限制於 15 分鐘以內，以降低長時間沉浸可能帶來的不適風險。鑑於學習時間有限，經與教育專家討論評估後，本研究在每次學習階段僅安排學習兩個物件名詞與對應的六組搭配詞組，以避免認知負荷過高。此設計兼顧兒童學習能力與教學成效，使教學活動在有限時間內能有效進行，同時維持良好的學習體驗。為控制學習歷程中的先學效應，避免兒童因先前學習而影響另一條件下的表現，每位參與者在兩種學習條件中皆學習六組不同的搭配詞，所有詞彙的難易度已與專家協調並加以平衡，確保兩條件下的學習內容具有等效挑戰性。

4-3.2 任務設計

受測者使用本研究開發之 MR 英文搭配學習應用程式，根據系統提供之動態提示與情境動畫進行發音與身體動作練習。理想上，系統會透過物件辨識、語音辨識及動作辨識提供即時回饋。然而，在正式實驗前本研究進行了一項前導實驗，發現基於當前 AI 動作辨識技術（MVITv2 模型）的準確度，尚無法完全滿足教學上即時回饋的需求。考量本實驗的重點在於驗證研究假設，而非系統開發的完善度，因此本研究最終採取 Wizard of Oz（WoZ）實驗設計法來處理動作相關的回饋。在此設計下，由一名實驗人員實時觀察並判斷學習者動作之正確性，再透過無線鍵盤手動控制 MR 系統的回饋內容，以確保回饋的準確性不受系統辨識技術限制。相關 AI 動作辨識技術之改進策略，將於後續章節中進一步討論。

本系統的視覺設計融合了兩大核心元素：動態提示與情境動畫。動態提示以動畫方式指引學習者進行正確動作，每組搭配詞皆設計對應之提示內容，如表 1 所示。情境動畫則模擬搭配詞於日常生活中的實際應用情境，以強化語意理解與情境聯想。上述視覺提示同時結合聽覺刺激，提供視覺與聽覺的雙重感官回饋，進一步提升學習沉浸感與認知參與度。



圖 2. 「Dress the doll.」學習任務中，操作實體物件與操作虛擬物件的受測者畫面

搭配詞的情境動畫分為兩類：漂浮在空中的虛擬情境與放置在地面的虛擬情境。具體而言，雨傘（umbrella）與天氣情境相連結，手機（phone）以雲朵形式呈現問候，水瓶（bottle）置於沙漠環境中，而娃娃（doll）則位於森林雪景。這些情境動畫有助於學習者理解搭配在現實情境中的運用，並能提升趣味性。為探討與虛擬物件及實體物件互動之間的差異，本研究為每個物件名詞製作了一個虛擬物件版本，如圖 2 所示。虛擬模型在尺寸上與實體物件保持一致，教學內容、動作設計、動態提示與情境動畫在虛擬與實體物件版本中皆保持不變。

4-4 實驗流程

實驗為期三週，第一周進行教學實驗並於實驗後進行測試，兩週後（第三週）進行延後測。實驗在非正課期間的兩個時段進行，分別為早自習及午休期間，每個時段學習不同的搭配詞組。參與者隨機分配至不同的操作順序（先實體後虛擬，或先虛擬後實體）。實驗流程包括：（1）系統熟悉階段：使用示範詞彙「cup」熟悉操作歷程；（2）學習階段：依序學習 6 組搭配詞，每組重複兩次；（3）後測評估：

完成口語測驗與文意測驗；(4) 問卷與訪談：評估認知負荷與學習體驗。實驗過程中，設置兩台攝影機進行錄影，如圖 3 所示。其中一台攝影機將固定於環境中，以第三人稱視角捕捉受測者的動作，如圖 3c 所示；另一台則將安裝在受測者胸前，以第一人稱視角記錄手部動作，如圖 3b 所示。同時，也錄製 MR 畫面，如圖 3a 所示，供後續分析使用。在進入學習階段之前，研究員會先示範如何佩戴頭戴裝置，並說明 MR 環境中將會進行的學習重點，隨後協助受測者佩戴裝置。進入 VerbVenture 系統後先進入練習階段，以詞彙「cup」及其搭配詞「raise the cup」進行互動練習使受測者熟悉學習環境與操作方式。練習階段結束後，研究員確認受測者是否理解互動流程並準備好進入正式學習，經確認後，受測者便開始學習階段。

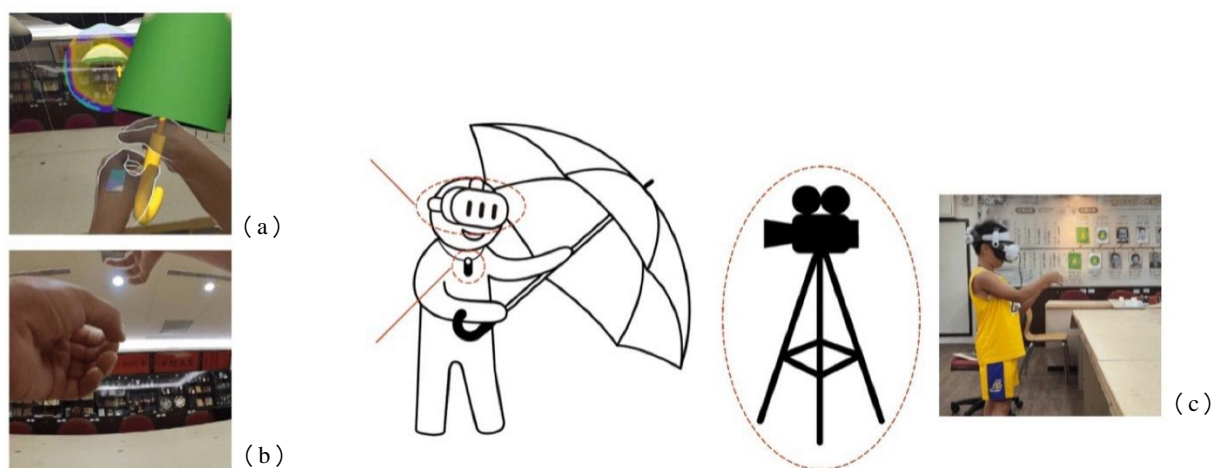


圖 3. 本研究實驗記錄配置

- (a) 第一人稱視角記錄 MR 畫面；(b) 第一人稱視角記錄受測者手部動作；
(c) 第三人稱視角記錄受測者的動作

對於實體物件互動的受測者，MR 環境同時呈現真實場域及真實物件。學習流程由受測者以左手食指指向真實物件開始，此動作會觸發物件出現黃色輪廓外框，並於 0.5 秒後顯現單字泡泡。受測者接著會聽到該單字的英文發音，並需大聲複誦；若發音正確，泡泡將會破掉。隨後，系統呈現第一組搭配詞的情境動畫，再出現動態提示指示所需執行的動詞動作。接著，受測者拿起單字物件，依據提示執行動作；當動作正確完成後，系統會顯示搭配詞泡泡，泡泡內包含搭配詞短句與獎牌，並播放該搭配詞的發音，受測者亦需大聲複誦。若複誦正確，泡泡將破掉，受測者即可獲得泡泡內的獎牌。當受測者完成與特定單字物件相關的三組搭配詞後，便會獲得一座該物件的虛擬金色獎盃。完成六組搭配詞學習後，研究員會指示受測者再次重複一次學習流程。虛擬物件互動的學習程序則採用相同結構，唯一的差異在於互動方式為操作虛擬物件，但所需完成的任務動作皆相同。此學習階段約 15 分鐘（含練習與教學）。

完成學習階段後，研究員協助受測者摘除頭戴裝置，接著受測者使用研究員提供的電腦進行後測評量。第一部分為口說測驗，受測者觀看靜態圖片後口頭回答相對應的英文搭配詞，研究人員則對其發音準確性進行評估。第二部分為文意測驗，測試中所用的素材與前測一致。評測結束後，本研究針對受測者進行半結構式訪談，以探討受測者對學習態度、動機及對 VerbVenture 原型系統的體驗感受。完成訪談後，第一個時段的實驗便結束，實驗總時長約 30 分鐘。同日中午，受測者將參與第二個時段的實驗，學習另外六組搭配，其流程與第一場實驗完全相同。第一週實驗結束後隔兩週，受測者將於第三週參加延後測，以評估單字記憶保留情況。該階段中，受測者將再次接受口說測驗與文意測驗。結束後進行半結構式訪談。與第一次實驗不同，延後測不使用 MR 頭戴裝置或任何基於 MR 的互動方式，僅使用電腦進行測驗與訪談，總時長約 12 分鐘。

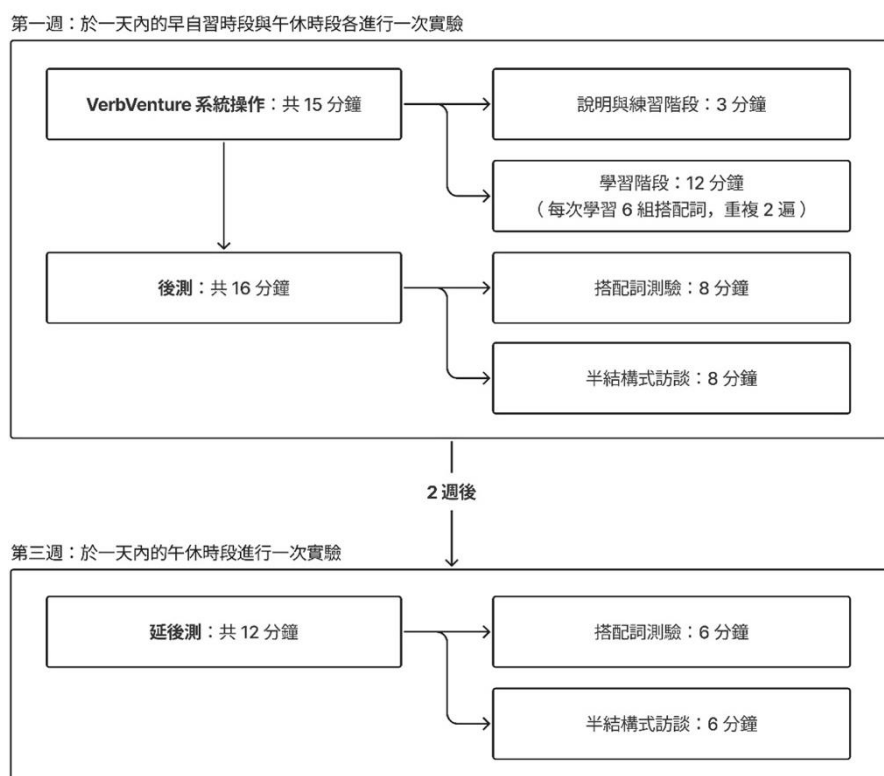


圖 4. 本研究實驗流程圖

4-5 學習成效量化分析

本研究的量化資料分析基於 14 位國小五年級學生的完整數據（原始招募人數為 18 位，其中 4 位因資料不全而排除）。每位參與者皆完成兩種條件下的前測、後測與延後測，研究主要的依變項為每次測驗之答題正確率。為探討條件（condition）與時間（time）對學習成效的影響，本研究採用 2（條件：實體互動／虛擬互動）× 3（時間：前測／後測／延後測）雙因子重複測量變異數分析（repeated measures ANOVA）。其中，條件與時間皆為受測者內因子，以各測驗階段的答題正確率作為依變數，用以分析主效應（時間效應、條件效應）及其交互作用。首先，Mauchly 球形檢定結果顯示，「時間」主效應違反了球形假設（ $W=0.591, p<0.05$ ），而「時間×操作條件」的交互作用則未違反此假設（ $W=0.958, p>0.05$ ）。因此，針對時間主效應的 F 值，本研究將報告採用 Greenhouse-Geisser 校正後的結果。

時間主效果達到顯著水準，Greenhouse-Geisser 校正後 $F(1.42, 18.46)=27.942(p<0.001, \eta^2p=0.682)$ ，顯示出大效應量。為了在控制整體第一類錯誤率的前提下，探究是哪些時間點之間存在差異，本研究採用了 Bonferroni 校正後的事後比較，結果顯示所有時間點間均存在顯著差異：後測分數（ $M=63.10, SD=25.40$ ）顯著高於前測分數（ $M=20.24, SD=15.94$ ），平均差異為 42.86（ $p<0.001$ ）；延後測分數（ $M=48.81, SD=29.37$ ）同樣顯著高於前測分數，平均差異為 28.57（ $p<0.01$ ）；此外，後測分數亦顯著高於延後測分數，平均差異為 14.29（ $p<0.01$ ）。此結果表明，VerbVenture 學習方法不僅帶來了立即且顯著的學習進步，其成效也得以部分保留至兩週後，儘管在延後測時出現了統計上顯著的遺忘現象。這與具身認知學習的前期研究相符，Ratcliffe 和 Tokarchuk（2020）發現，在沉浸式虛擬環境中，結合具身控制與口語的學習方式，能顯著提升日語詞彙（特別是動詞）的記憶表現，顯示具身認知本身對記憶有正面影響，而 Macedonia 和 Knösche（2011）指出，促動學習（enactment learning）透過結合多種感官與運動訊息，使大腦能以多模態方式處理資訊，從而增強語言記憶。

操作條件主效果未達顯著， $F(1, 13) = 1.631, p > 0.05, \eta^2 p = 0.111$ ，顯示出中等效應量。實體（P）互動（ $M = 40.87, SD = 29.72$ ）的整體表現略低於虛擬（V）互動（ $M = 47.22, SD = 30.11$ ），但差異未達統計顯著水準。此結果表明，原先假設 H1『實體互動優於虛擬互動』並未獲得統計支持。然而其效果量卻達到了中等水平，此現象很可能源於本研究較小的樣本數（ $N = 14$ ）所導致的統計考驗力不足，兩種操作條件的學習軌跡可能存在潛在差異：P 條件呈現較大的即時效益，但其保留效果的下降也較為明顯；相對地，V 條件則呈現較為穩定的學習曲線。這些模式可能反映實體與虛擬物件在具身學習中的不同特性。例如：P 條件中手持真實物體提供的物理回饋能觸發深刻的運動痕跡（motor trace），有助於短期記憶編碼（Bara & Kaminski, 2019），而 V 條件中虛擬物件特別醒目的視覺呈現帶來更豐富的記憶表徵則可能支援較穩定的記憶保留（Schurgin, 2018），本研究將在後續的質化分析與結果討論中進一步探討。

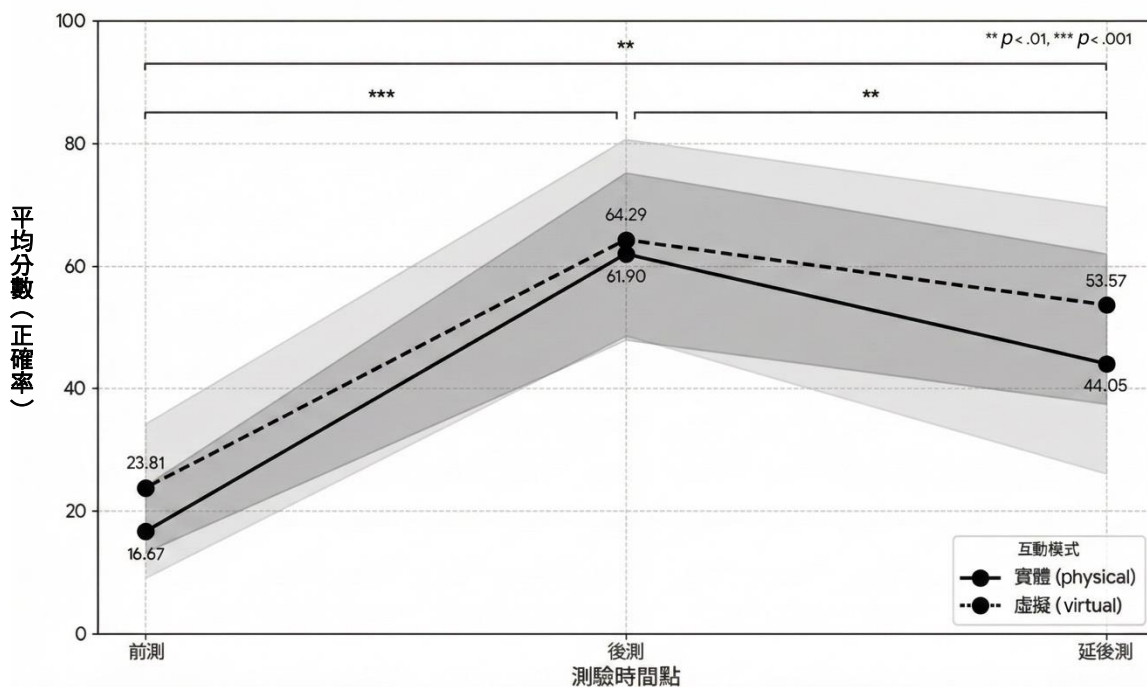


圖 5. VerbVenture 中 P 條件（操作實體物件）和 V 條件（操作虛擬物件）在不同時間點的平均正確率

此折線圖縱軸表示正確率（%），橫軸表示時間（前測、後測、延後測）。誤差棒表示 95% 信賴區間（CI）

時間×操作條件交互作用未達顯著， $F(2, 26) = 0.361, p > 0.05, \eta^2 p = 0.027$ ，顯示出小效應量。此結果意味著兩種條件在不同時間點的學習曲線趨勢相似。雖然從描述性數據來看，V 條件在延後測的記憶保持狀況略優於 P 條件，但此趨勢並未達到統計上的顯著差異。如圖 5 所示，P 條件從前測（ $M = 16.67, SD = 13.07$ ）至後測（ $M = 61.91, SD = 23.05$ ）呈現較陡峭的上升趨勢，但從後測到延後測（ $M = 44.05, SD = 31.08$ ）下降幅度較大；V 條件則從前測（ $M = 23.81, SD = 18.16$ ）至後測（ $M = 64.29, SD = 28.39$ ）上升較平緩，延後測（ $M = 53.57, SD = 27.87$ ）下降幅度較小。然而，交互作用不顯著，意味著兩種條件在不同時間點的學習曲線相似，如圖 5 所示。

本研究進一步針對學習歷程的兩個關鍵階段—「立即學習增益」與「長期記憶保留」—進行成對樣本 t 檢定，以深入探討 P 條件（實體物件操作）和 V 條件（虛擬物件操作）對學習成效的影響。

本研究比較了兩組的學習增益分數（後測減去前測），如圖 6a 所示。結果顯示，實體組的平均進步幅度（ $M = 45.24, SD = 24.83$ ）與虛擬組（ $M = 40.48, SD = 29.75$ ）之間，並未達到統計上的顯著差異， $t(13) = -0.618, p > 0.05$ 。其效果量 Cohen's d 為 -0.165，屬於微小效果量，這表示在立即學習成效上，兩種模式

效果相當。本研究接著比較了兩組的遺忘分數（後測減去延後測）（如圖 6b）。結果顯示，實體組的平均遺忘分數（ $M=17.86, SD=23.98$ ）與虛擬組（ $M=10.71, SD=18.03$ ）之間的差異，未達到統計顯著水準 $t(13)=-0.822, p>0.05$ 。其效果量 Cohen's d 約為 -0.220 ，屬於小效果量。此結果表明，兩種互動模式在長期記憶的保留效果上，並無顯著差異。綜合來看，延伸分析的結果與先前的 ANOVA 結果一致，亦即實體與虛擬互動皆能帶來顯著的學習，但在學習的效率與長期保留效果上，兩者間並無顯著優劣差異。

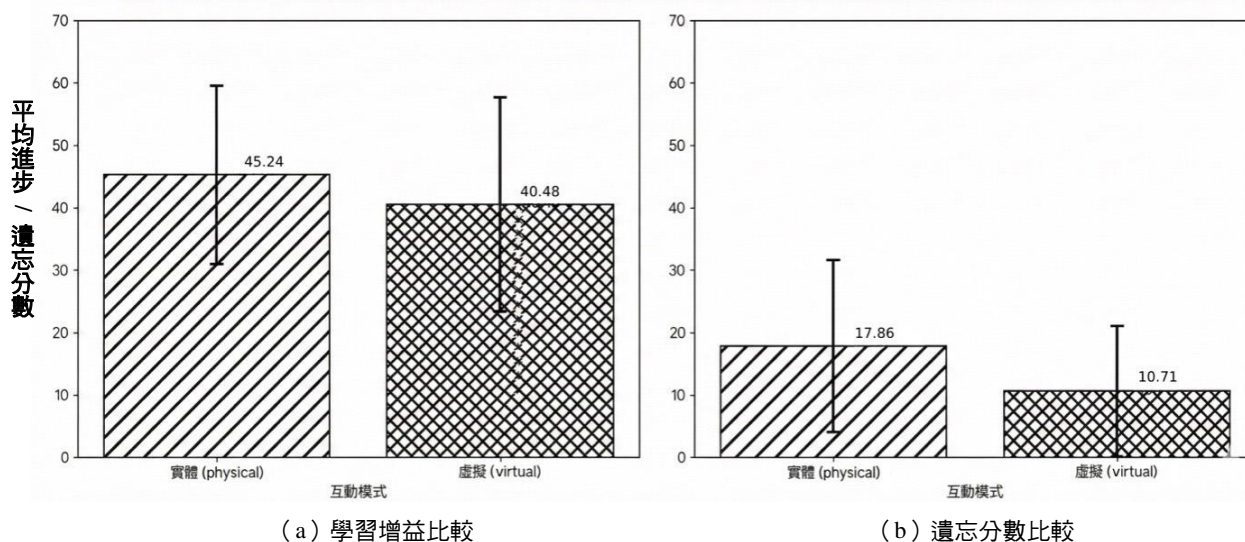


圖 6. 學習增益與長期記憶保留比較圖 誤差棒表示 95%信賴區間 (CI)

4-6 使用者體驗與行為質性分析

4-6.1 使用者體驗訪談及偏好

本研究採用主題分析法 (Braun & Clarke, 2006)，著重於受測者對 VerbVenture 英語搭配詞的學習體驗，特別關注雙手操作的互動體驗以及具身學習在互動中對學習成效的影響。在實驗中兩種操作條件僅標示為 P（實體物件操作）與 V（虛擬物件操作），受測者資訊以 PX 表示，其中 X 為受測者代號。

在 14 位參與者中，有 8 位（P4、P8、P10、P11、P13、P14、P15、P17）偏好操作虛擬物件，其主要吸引力在於超越現實物理限制所帶來的新奇感與高自由度。P17 的回饋極具代表性：「虛擬物件可以漂浮在空中，可以放在不同的地方，覺得很好玩，讓我對學習更有興趣。」相對地，偏好操作實體物件的 5 位參與者（P6、P7、P9、P12、P18），則更看重真實的物理互動。他們認為，真實物件的觸感與可預測的物理屬性，使其更易掌控並有助於動作記憶。P9 便指出：「……有實際拿起來的感覺，比較清楚有沒有拿到物件……」，這凸顯了觸覺回饋在具身學習中的重要性。

儘管兩種操作模式各有其吸引力，參與者也回饋了在混合實境環境中所面臨的具體挑戰。虛擬物件的挑戰主要源於其非物理性與技術限制：虛擬物件因缺乏觸覺回饋與物理屬性，導致難以掌控或出現浮空等非預期狀況，進而影響學習專注度。此問題直接導致了認知上的干擾，如 P7 所述：「我覺得我比較會分心，因為娃娃會一直亂飛。」在執行細緻動作時，現行技術限制增加了操作難度。P16 便提到，在執行「把雨傘捲起來。」等動作時，因細小繩帶不易抓取而感到操作麻煩。P11 也指出：「抓起虛擬物件的方式不符合直覺。」顯示抓取虛擬物件的手勢與現實經驗不符，可能增加認知負荷。

實體物件的挑戰則主要與 MR 技術限制和人因相關。部分受測者反映，透過 MR 頭盔觀看實體物件時，視覺提示與真實物件之間存在大小或位置上的落差，影響了操作的直覺性，如 P11 表示：「在頭盔下操作實體物件的時候會覺得不好拿，把物件照指示放回去的時候圖標跟對象大小對不上。」此外，長

時間手持實體物件會造成身體疲勞，這一點在 P13 的回饋得到了深入的闡述。他對比了兩種模式的差異，解釋了為何更偏好虛擬物件：「因為實體物件需要一直拿著會比較累...（虛擬物件）拿起來只是做『抓』的動作，如果覺得累，手放開的話物件會浮在空中，虛擬物件還是比較好。」P13 的觀點不僅突顯了虛擬互動在減輕身體勞損上的優勢，更進一步揭示了其在擴展學習內容範疇上的巨大潛力。他提到，虛擬物件打破了物理現實的限制，允許使用者與現實中不可能操作的物體互動，例如：「可以拿起一些很重的東西（例如：房子）」或是「拿很輕的東西（例如：雲），一些平常抓不到的東西。」。此回饋表明，虛擬物件的優勢不僅在於提升操作的舒適度與自由度，更關鍵的是它能極大擴展具身學習的內容廣度，將學習從可觸及的實體物件，延伸至抽象、巨大或無形的物件，這為教學設計提供了更豐富的可能性。

儘管訪談設計中並未直接引導受測者與傳統教學法進行比較，但受測者普遍的反饋表明，VerbVenture 系統所提供的沉浸式、具身化學習體驗，相較於課堂學習更加提升他們的學習樂趣與參與度。多位受測者頻繁使用「有趣」、「好玩」等詞彙描述學習感受，並直接表達這種體驗能「讓我對學習更有興趣」（P17）。這種積極的情緒回饋強烈暗示，相較於傳統相對靜態的學習模式，MR 所創造的互動環境能更有效地激發兒童的學習動機，使其主動投入學習過程，這對於提升潛在低成就學生的學習意願尤其重要。

4-6.2 使用經驗質性分析

根據問卷與訪談結果，VerbVenture 的視覺與互動設計在多方面獲得了參與者的正面肯定，同時也揭示了關鍵的改進方向。

1. 動畫與情境設計：無論是大規模的環境特效或小範圍的物件動畫，都成功地提升了使用者的沉浸感與參與度。例如：「Put up the umbrella」中大範圍的下雨動畫因其真實感而讓 P9、P13 與 P16 印象深刻；「phone」情境中天空出現的「Hello」字樣及電話互動動畫，也普遍吸引了多位參與者（P4、P6、P7、P10、P12、P13、P17）的注意並激發其互動意願；而在「Fill the bottle」任務中，P9 與 P13 也特別讚賞了虛擬注水動畫的流暢感。這些回饋共同表明，生動且具情境感的視覺動畫是吸引使用者注意力的有效設計。在核心互動體驗方面，「Fill the bottle」的倒水動作設計證實其核心玩法具有跨越媒介的吸引力，無論是使用實體物件（P4、P6）或虛擬物件（P9、P13），參與者都認為該任務有趣且好玩。然而，虛擬物件的互動體驗也凸顯了物理屬性的重要性。部分參與者（P12、P17）反映，虛擬物件因缺乏物理特性而出現的浮空或卡頓等異常狀況，會引發負面感受。此問題不僅容易造成分心（P7），也直接影響了操作的流暢性（P18），顯示為虛擬物件導入擬真的物理屬性，是提升互動真實感與順暢度的必要功能。
2. 互動設計：在互動設計層面，本研究發現 VerbVenture 的核心具身學習機制獲得了參與者的直接肯定，而系統的引導回饋與獨特的互動自由度，則進一步強化了整體的正面體驗。首先，核心的具身學習機制被證實能有效輔助記憶。P9 在訪談中明確指出，透過指向特定物件並執行相應動作來學習單字，確實能加深印象，這直接驗證了此互動設計在強化學習效果上的潛力。其次，系統的引導與獎勵機制成功地提升了參與感與成就感。P13 的回饋是其最佳例證，他對互動中出現的閃亮提示特效，以及完成任務後的獎盃動畫表達了顯著的偏好。最後，關於互動自由度的回饋，則揭示了使用者對虛擬物理一種更為細膩的期待。P12 與 P13 對於能讓物件浮空或輕鬆舉起大型物件等超越現實的操作感到新奇，這顯示了賦予使用者「超能力」的互動自由度是極具吸引力的。此觀點與先前使用者抱怨物件卡頓等缺乏擬真物理的負面回饋看似矛盾，實則不然。使用者抱怨的是不可預期的、破壞流暢感的系統錯誤；而他們享受的，是可預期的、帶來賦權感的設計。這表明，成功的互動設計關鍵不在於絕對的物理擬真，而在於建立一套清晰一致的內部規則，既能

避免破壞沉浸感的失控，又能提供增強趣味性與掌控感的互動自由。

3. 學習內容：在學習內容方面，本研究發現，將情境脈絡、生活經驗與具身互動三者有效結合，是提升 EFL 兒童學習動機與記憶成效的關鍵。首先，情境化的具身學習能有效應對學習挑戰。P13 回饋，「umbrella」一詞本身較具挑戰性，但透過親手操作真實雨傘，並搭配逼真的下雨情境動畫，顯著地加深了受測者本人的記憶。此案例表明，豐富的情境線索（視覺、聽覺）與實體互動的結合，能為較困難的學習內容提供強力的記憶鷹架。其次，與生活經驗高度相關的素材能顯著提升學習動機與親切感。P11 指出，「phone」和「doll」等日常物件讓學習過程更具實用性與熟悉感，從而感到有趣。這反映出取材於兒童生活經驗的內容，能有效降低學習門檻，並透過內在連結增強其參與動機。綜合來看，VerbVenture 的內容設計驗證了，當學習素材能讓兒童在貼近真實的情境中，操作與自身經驗相關的物件時，其單字記憶與學習興趣將獲得最有效的提升。

五、結果與討論

本章旨在深入探討研究結果的意涵，將第四章呈現的量化與質性數據與研究問題、相關文獻及理論框架進行連結。考量本研究樣本數較小（有效樣本 $N=14$ ），在統計考驗力上有所限制，因此本章的論述將更多地聚焦於質性分析所揭示的豐富洞見，並將量化結果作為一個輔助性的佐證，用以呈現學習趨勢而非做出強力的因果推論。本章首先討論 MR 具身學習的整體學習成效與挑戰，接著深入詮釋本研究的核心發現—實體與虛擬物件互動的互補價值，並基於此提出 MR 具身學習環境的設計建議（重點回應 RQ3 並檢視 H2）；最後，將探討 AI 技術整合的考量，並總結研究限制。

5-1 MR 具身學習對搭配詞學習的整體成效與挑戰

本研究最核心的量化結果顯示，不論採用何種互動模式，VerbVenture 系統皆能顯著提升學童對英語搭配詞的學習成效。重複測量變異數分析顯示，「時間」的主效果達到顯著水準（ $p<0.001$, $\eta^2 p=0.682$ ），呈現出大效應量。事後比較進一步證實，後測分數與延後測分數均顯著高於前測分數。此結果強烈表明，本研究所設計的 MR 具身學習體驗，不僅能帶來立即且顯著的學習進步，其成效也能部分保留至兩週後。此發現與前期研究的結論相符，即透過身體的實際參與，結合多重感官與運動訊息，能有效增強語言記憶的建立（Cox et al., 2010）。

儘管如此，一個顯著的落差在於，參與者雖能回憶中文詞義與對應動作，卻難以準確地產出英文的口語形式。此挑戰可歸因於以下多重因素的交互作用。

1. 學習者因素：受測的小學生，作為英語學習的基礎階段學習者，其語音感知與模仿能力仍在發展階段。此外，如文獻所述，中文為母語的學習者在學習英語搭配詞時，因語言使用習慣的差異，本就容易產生語用及發音上的錯誤（Bahns & Eldaw, 1993），這需要大量的重複練習才能鞏固。
2. 教學時間限制：考量小學生單次學習的記憶量與專注力，本研究單次實驗的教學時間有限（每次 12 分鐘，重複 2 遍），這對於兒童充分練習並內化搭配詞的標準發音是不足的，特別是對於流利表達而言，需要更大量的重複練習與矯正。
3. 語音示範模型的清晰度與可模仿性：系統預設的語音播放速度，對初階學習者而言可能過快。過快的語速不僅增加了聽辨的難度，更直接阻礙了有效的語音感知與模仿，導致學習者難以掌握準確的發音細節與語音節奏。

4. 語音辨識的準確性與回饋機制：在評估回饋環節，現行 STT (speech-to-text) 語音辨識技術，對於兒童變異性較高的語音特徵（如音高、語速）辨識準確性仍待優化。此技術限制使得系統不易準確判斷學習者的發音錯誤，從而無法提供即時、具體且有效的矯正回饋，削弱了互動練習的成效。
5. 教學設計層面的反思—被動練習的侷限性：除了上述因素，本研究也需從系統本身的教學設計進行反思。目前的互動流程主要停留在「聆聽與複誦」，這本質上仍是一種較為被動的練習形式。雖然此方法有助於建立詞彙發音與動作的初步連結，但對於促進學習者自發性、創造性的口語產出幫助可能有限。一個更理想的設計，應包含更多需要學習者主動建構語言的任務，例如：要求他們用目標搭配詞描述一個新的情境，或對虛擬角色下達指令等。缺乏此類產出導向的練習，可能是導致學習者雖能理解，卻難以流利使用的關鍵原因之一。

綜合而言，VerbVenture 在促進詞義習得與動作連結上成效卓越，但在培養發音精確性與口語流利度方面仍有待加強，這為系統的未來迭代指出了明確的改進方向。

5-2 實體與虛擬互動的互補價值

本研究的另一項關鍵發現是，儘管質性回饋中參與者對「實體 (P)」與「虛擬 (V)」物件的操作模式各有偏好，但量化數據顯示，兩種模式在整體的學習成效上並無統計上的顯著差異 ($p>0.05$)。此結果雖未支持原先假設 H1「實體互動優於虛擬互動」，卻揭示了更深層的意涵：實體與虛擬物件在具身學習中並非單純的優劣之分，而是為學習者提供了具互補價值的學習體驗與路徑。

實體互動具有體感真實性與初始編碼優勢：偏好實體物件的 5 位參與者 (P6、P7、P9、P12、P18)，其主要理由集中於真實的物理互動所帶來的直觀性與掌控感。他們認為，真實物件的觸感與可預測的物理屬性，使其更易操作並有助於動作記憶。P9 便精確地指出：「……有實際拿起來的感覺，比較清楚有沒有拿到物件……」，這凸顯了觸覺回饋在建立穩固的身體感知—運動連結上的重要性。此觀點與 Bara & Kaminski (2009) 研究發現高度一致，即手持真實物體能觸發深刻的運動痕跡 (motor trace)，有助於建立強力的初始記憶編碼。這或許能解釋在描述性數據中，實體組從前測到後測呈現出較為陡峭的學習曲線，暗示其可能在立即學習階段具有潛在優勢。

虛擬互動帶來新奇感、高自由度與長期記憶潛力：相對地，偏好虛擬物件的 8 位參與者 (P4、P8、P10、P11、P13、P14、P15、P17)，則被其超越現實的互動自由度與新奇感所吸引。P17 的回饋極具代表性：「虛擬物件可以漂浮在空中...覺得很好玩，讓我對學習更有興趣。」P13 更進一步提到，虛擬物件不僅能減輕身體負擔，更能擴展學習內容至現實中不可能操作的物件（如房子、雲）。這種由新奇感與賦權感 (empowerment) 驅動的高度動機，可能促使學習者投入更多認知資源。根據 Schurgin (2018) 的理論，這種更深度的認知加工有助於建立細節更豐富、更依賴「回憶」(recollection) 的記憶表徵。一個更穩固的表徵結構或許能抵抗時間造成的遺忘，這也與描述性數據中，虛擬組在兩週後的延後測呈現出較為穩定的記憶保留趨勢相符。

5-3 MR 具身學習環境之設計建議

本研究證據顯示，假說 H3「情境動畫與動態提示能有效吸引學習者注意力、提升對語意與情境用法的理解」僅獲得部分支持。研究發現，引導設計的具體品質是決定其成敗的關鍵，而非其形式本身。

首先，與學習內容高度相關的情境動畫及獎勵性回饋，能有效提升學習動機與語境理解。設計精緻的情境動畫（如「Put up the umbrella」的下雨場景、「Fill the bottle」的倒水流動感）與獎勵性視覺回饋

（如閃亮特效、獎盃）確實能有效吸引學習者注意力，引發正面情緒反應，並可能間接促進對語境的理解與學習動機（P9、P13、P16 等反饋），此部分支持 H3。

然而，引導設計若語意模糊或步驟過於繁瑣，不僅無法有效引導，反而會因增加認知負荷而降低學習效率。現有設計中並非所有動態提示均能達到預期效果。特定操作引導提示（如拿起雨傘與放回雨傘的半透明雨傘模型）因語意不明確而未能被多數學習者正確解讀（P7、P9、P10、P11、P12、P13），P7 看到動態提示後沒有直接做出動作，反而主動詢問研究員：「要拿雨傘嗎？」顯示系統未能有效引導受測者直覺性的將雨傘拿起或放回，從而挑戰了 H3 中關於提示能提升理解的假設。此外，設計欠佳的多步驟複雜操作提示（如「answer the phone」），不僅因資訊過多而增加了認知負荷（P16），更直接導致了操作的分解與錯誤。多位參與者（P4、P9、P15、P16、P18）僅執行了部分動作，例如：只做滑手機（P4、P16）或只做將手機放到耳邊（P9、P15）的動作。這種因引導不清而導致的操作失敗，中斷流暢的學習體驗，並可能引發學習者的挫敗感。更為關鍵的是，在具身學習的框架下，重複執行不完整的動作，可能導致學習者將詞彙的語意與一個錯誤的動作產生連結，從而建立起錯誤的具身記憶（incorrect embodied memory），這對學習的正確性構成潛在風險。

綜合以上成功與失敗的案例，本研究針對 RQ3 提出以下五項 MR 具身學習引導機制的設計原則：

1. 互動簡潔性與直觀性原則：最大限度地簡化操作步驟，將複雜任務分解為單一、符合使用者既有心智模型的直觀動作，以降低認知負荷。此原則看似通用，但在 MR 具身學習中尤為重要。如本研究中，「answer the phone」因被設計為「滑動後再放到耳邊」的多步驟任務，直接導致多位參與者（P4、P9、P15、P16、P18）操作分解與錯誤的具身記憶連結，此結果證明了在要求身體執行的情境下，任何額外的認知負荷都可能干擾核心的學習目標。
2. 引導提示的清晰性與顯著性原則：視覺與聽覺提示必須語意明確、易於理解，並具備足夠的顯著性以有效引導注意力（Syberfeldt et al., 2015）。此基本原則在 MR 具身學習中尤為關鍵，因為模糊的提示會直接干擾身體的直覺反應。本研究中，「拿起雨傘」的半透明模型提示便是一個深刻的反面教材。其模糊的語意讓多位參與者（P7、P9—P13）感到困惑，P7 甚至在看到提示後主動詢問「要拿雨傘嗎？」，這證明了缺乏清晰性的引導，將迫使學習者從「身體感知模式」切換回「認知解讀模式」，從而破壞了具身學習的流暢性。
3. 情境化與動機化回饋原則：策略性地運用與語意高度相關的情境動畫，並整合即時、具吸引力的獎勵性回饋，以強化語意連結並維持學習動機。在 MR 環境中，情境不僅是背景，更是語意本身的一部分。本研究的成功案例有力地支持了此原則，例如：「Put up the umbrella」的下雨場景（P9、P13、P16）與「Fill the bottle」的倒水流動感，都透過創造一個需要「解決問題」的情境，將搭配詞從抽象符號轉化為具體的身體目標。同樣地，完成任務後的獎盃動畫（P13）等動機化回饋，在沉浸式環境中能提供比傳統介面更強的成就感，從而維持學習者的持續參與。
4. 多模態回饋整合原則：尤其在虛擬物件互動中，應整合擬真的聽覺（Canales & Jörg, 2020）或探索觸覺回饋，以彌補物理感知的缺失，提升沉浸感。此原則的重要性，可從偏好實體物件的參與者回饋中得到反向印證。多位參與者（P6、P7、P9、P12、P18）皆強調「有實際拿起來的感覺」至關重要，這表明在具身學習中，觸覺與本體感覺的回饋是建立穩固記憶連結的關鍵通路。因此，當視覺是唯一的回饋渠道時（如操作虛擬物件），系統就必須策略性地利用聽覺（如碰撞聲、摩擦聲）或其他感官渠道來補償缺失的物理感知，否則將難以建構一個可信、有效的具身學習體驗。
5. 目的性物理與互動自由原則：為虛擬物件建立一套清晰、一致的內部物理規則。此原則的關鍵不

在於追求「完全擬真」，而在於規則的「一致性」與「目的性」。本研究的證據完美地展示了此觀點的雙面性：一方面，參與者對虛擬娃娃會亂飛、卡頓等不可預期的物理錯誤感到困惑與分心（P7、P12、P17）；另一方面，他們卻對能夠輕鬆舉起虛擬重物（房子、雲）等可預期的、可引發賦權感的超現實互動感到新奇與興奮（P13）。這表明，一個成功的 MR 學習物理系統，應以擬真物理作為防止失控的基石，並在此基礎上策略性地賦予使用者提升動機的互動自由。

5-4 AI 技術整合之挑戰與考量

本研究在將 AI 技術整合於 VerbVenture 系統時，揭示了若干關鍵技術挑戰，這些挑戰對於未來開發相似系統具有高度的參考價值：（1）物件辨識：儘管如 SAM2 等先進模型已展現出良好的辨識性能，但實現即時、內建的實體物件辨識，其主要障礙來自於平台層面的限制。在研究當下，主流 MR 設備（如本研究使用的 Meta Quest）出於隱私考量，嚴格限制了對 Passthrough 攝影機原始數據的存取。本研究採用的替代方案（外接攝影機或 PC 端處理）雖驗證了技術可行性，卻犧牲了系統的整合度與潛在的即時性。因此，此功能的未來發展，高度依賴於平台政策的開放與在保護隱私前提下的 API 支持。（2）動作辨識：MViTv2 等模型雖具潛力，但在真實應用中仍面臨三大挑戰：第一，兒童使用者動作的自然變異性；第二，MR 環境中複雜的光照與遮擋；第三，頭戴裝置有限的計算資源。手勢辨識的不穩定性直接影響了互動的流暢性與具身學習體驗的品質。本研究的實驗觀察提供了一個具體的解決思路：與其試圖辨識完整且多變的連續動作（如「撐傘」的完整過程），不如將判斷重點聚焦於動作中最關鍵、最穩定的視覺線索（如「傘面張開」的狀態變化）。未來在訓練模型時，或可透過遷移學習（transfer learning）或微調（fine-tuning）整個模型，例如：準備一個特別標註的數據集，藉此引導模型的自注意力機制（self-attention mechanism）去捕捉最核心的視覺特徵，從而提升辨識的準確性與穩健性。（3）語音處理：在語音處理上，Voice SDK 雖提供了基礎功能，但其挑戰在於對特定用戶群的適配性不足。研究發現，系統預設的語音播放速率對初學兒童而言可能過快，直接影響了模仿學習的效果。這表明，AI 語音模型的可配置性（如語速調整）與針對性優化，對於建立一個對兒童更友善、更有效的語言互動循環至關重要。

5-5 研究限制

本研究雖獲得初步成果，但仍存在若干限制，為結果的詮釋與未來研究方向提供了重要脈絡。首先，在研究設計上，本研究的樣本規模較小（有效樣本 $N=14$ ），且集中於特定區域與年級，這可能限制了研究結果的普遍性。如量化分析所示，儘管實體與虛擬互動條件的主效果未達統計顯著（ $p>0.05$ ），但其中等的效應量（ $\eta^2 p=0.111$ ）暗示，樣本數不足所導致的統計考驗力低下，可能是未能偵測出潛在差異的原因。此外，本研究未引入與傳統教學法比較的對照組，因此無法直接推論 VerbVenture 系統相較於傳統方法的相對效益。其次，在學習目標上，本研究的結果明確顯示，系統雖在促進搭配詞的語意理解與動作連結上成效顯著，但在口語產出形式的掌握上則面臨挑戰，揭示了當前設計側重語意習得，而對口語流利度的支持相對不足。最後，在技術層面，為了確保互動回饋的即時性與準確性，目前的實驗採用 Wizard of Oz（WoZ）研究方法。這意味著實驗中所評估的學習效果，是基於人為控制下、近乎完美的動作辨識回饋。此方法雖有助於驗證教學設計本身的有效性，卻也意味著目前尚無法得知，當內建 AI 在實際運行中遭遇不可避免的辨識不準確問題時，會對學習體驗與成效產生何種影響。在此脈絡下，系統功能的完整實現仍受限於當前的技術瓶頸，包括因 MR 硬體平台對 Passthrough 攝影機數據的存取限制而影響了物件辨識的整合度，以及 AI 模型在面對兒童多變的動作與語音特徵時，其辨識準確性仍待優化。未來的研究應擴大樣本規模、納入對照組設計、將口語流利度設為明確的學習目標，並致力於整合真實、自主運行的 AI 模型，以評估在存在辨識誤差的現實條件下，學習者的體驗與成效會如何變化，從而建立更完整、更有效的具身學習體驗。

六、結論

本研究成功開發並評估了 VerbVenture 系統，一個結合 MR 與 AI 技術的創新英語搭配詞學習平台，旨在為 EFL 兒童提供沉浸式的具身學習體驗。本章將總結研究發現，並針對研究問題與假說提出結論。

針對 RQ1「MR 具身學習在多大程度上提升了搭配詞習得成效？」本研究的量化結果顯示，VerbVenture 系統具有顯著且效果量大的學習成效（時間主效果 $p < 0.001$ ， $\eta^2 p = 0.682$ ）。無論是透過實體或虛擬物件操作，均能有效促進學習，且學習成效得以部分保留至兩週後。此發現強烈支持將具身認知理論應用於語言教學的有效性，證實身體參與及環境互動能顯著增強學童的記憶保留與語意理解。

針對 RQ2「操作實體與虛擬物件的差異與特性為何？」以及 H1「實體互動將比虛擬互動產生更佳的學習成果」，本研究發現 H1 未獲統計支持，兩種模式在學習成效上並無顯著差異（ $p > 0.05$ ）。然而，質性回饋揭示了兩者在學習體驗上具有互補的價值與特性。實體物件操作憑藉其真實的觸覺回饋與物理感受，提供了更直觀、更易掌控的互動，有助於建立穩固的初始記憶編碼；虛擬物件操作則以其超越現實的互動自由度與新奇體驗，在提升學習動機與參與度上表現出巨大潛力。

針對 RQ3「如何設計有效的引導機制？」以及 H2「情境動畫與動態提示能有效吸引學習者」，本研究發現 H2 僅獲部分支持。研究證實，設計精良、與語意高度相關的情境動畫及獎勵性回饋，確實能有效吸引注意力並提升動機。然而，語意模糊或步驟過於複雜的動態提示，不僅無法有效引導，反而會因增加認知負荷而降低學習體驗。基於此，本研究歸納出 MR 具身學習引導的關鍵設計原則，強調互動的簡潔直觀性、提示的清晰性、以及回饋的情境化與動機化。

本研究的主要貢獻在於：（1）設計並實現了一個整合 AI 與 MR 的具身語言學習系統原型；（2）系統性地比較了實體與虛擬物件互動在具身學習中的特性與效果；（3）基於實證數據提出了 MR 環境中視覺引導的設計原則；以及（4）探討了 AI 技術導入此類系統的應用潛力與現實挑戰。

VerbVenture 的設計與研究發現為語言教育實踐提供了多項應用價值。對教師而言，它提供了一種將抽象語言概念轉化為具體身體經驗的創新工具；對教育科技開發者而言，本研究提出的設計原則可作為開發相似系統的參考；對家長與自學者而言，它展示了一種將語言學習融入日常生活的可能性。本研究的貢獻不僅在於當下的發現，更在於其前瞻性價值。此研究結果預見，隨著 AI 與 MR 技術的快速發展與普及，智能化、沉浸式的語言學習工具將成為教育的主流；然而，技術的進步不能取代對教學法本身的深刻理解。本研究透過在理想條件下驗證核心的具身互動設計，為這個即將到來的趨勢提供了一份奠基性的教學設計藍圖。在底層 AI 技術完全成熟之前，先行確立何種互動、何種引導是有效的，是確保未來科技能真正服務於教育、而非淪為技術堆砌的關鍵。隨著科技不斷進步，期待未來能發展出更智慧、適應性更強的 MR 語言學習系統，為學習者創造更豐富、更有效的學習體驗。

誌謝

感謝臺北市立大學附設實驗國民小學吳家瑞校長、翁瑞伶主任、周永泰組長、蘇怡如老師協助實驗順利完成。誠摯感謝審查委員的細心指正以及國科會研究計劃輔助（112-2221-E-011-101-MY3）。

參考文獻

1. Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations? *System*, 21(1), 101-114. [https://doi.org/10.1016/0346-251X\(93\)90010-E](https://doi.org/10.1016/0346-251X(93)90010-E)
2. Bara, F., & Kaminski, G. (2019). Holding a real object during encoding helps the learning of foreign vocabulary. *Acta Psychologica*, 196, 26-32. <https://doi.org/10.1016/j.actpsy.2019.03.008>
3. Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101. <https://doi.org/10.1191/1478088706qp063oa>
4. Chan, T., & Liou, H.-C. (2005). Effects of web-based concordancing instruction on EFL students' learning of verb-Noun collocations. *Computer Assisted Language Learning*, 18(3), 231-251. <https://doi.org/10.1080/09588220500185769>
5. Chao, L. L., & Martin, A. (2000). Representation of manipulable man-made objects in the Dorsal Stream. *NeuroImage*, 12(4), 478-484. <https://doi.org/10.1006/nimg.2000.0635>
6. Cox, M., Schofield, G., & Kolt, G. S. (2010). Responsibility for children's physical activity: Parental, child, and teacher perspectives. *Journal of Science and Medicine in Sport*, 13(1), 46-52. <https://doi.org/10.1016/j.jsams.2009.02.006>
7. Canales, R., & Jörg, S. (2020). Effects of virtual hand representation on interaction and weight perception. In *Proceedings of Motion, Interaction and Games* (Article 6). Retrieved from <https://dl.acm.org/doi/10.1145/3611659.3615689>
8. Draxler, F., Labrie, A., Schmidt, A., & Chuang, L. L. (2020). Augmented reality to enable users in learning case grammar from their real-world interactions. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-12). Retrieved from <https://dl.acm.org/doi/10.1145/3313831.3376537>
9. Ebert, D., Gupta, S., & Makedon, F. (2016). Ogma: A virtual reality language acquisition system. *Proceedings of the 9th ACM International Conference on Pervasive Technologies Related to Assistive Environments* (pp. 1-5). Retrieved from <https://dl.acm.org/doi/10.1145/2910674.2910681>
10. Feng, B., & Ng, L.-L. (2024). The spatial influence on vocabulary acquisition in an immersive virtual reality-mediated learning environment. *International Journal of Computer-Assisted Language Learning and Teaching*, 14(1), 1-17. <https://doi.org/10.4018/IJCALLT.339903>
11. Fuhrman, O., Eckerling, A., Friedmann, N., Tarrasch, R., & Raz, G. (2021). The moving learner: Object manipulation in virtual reality improves vocabulary learning. *Journal of Computer Assisted Learning*, 37(3), 672-683. <https://doi.org/10.1111/jcal.12515>
12. Hadid, A., Mannion, P., & Khoshnevisan, B. (2019). Augmented reality to the rescue of language learners. *Florida Journal of Educational Research*, 57(2), 81-89.
13. Halbig, A., & Latoschik, M. E. (2024). Common cues? Toward the relationship of spatial presence and the sense of embodiment. In *Proceedings of IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 1117-1126. <https://doi.org/10.1109/ISMAR62088.2024.00128>
14. Hsu, C., Chen, Y., Liu, Y.-J., Chang, Y.-C., & Lee, M.-J. (2023). Spelland: Situated language learning with a mixed-reality spelling game through everyday objects. *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-6). Retrieved from <https://dl.acm.org/doi/10.1145/3544549.3583830>

15. Hsu, C., Sun, Y.-C., Wang, H.-Y., Liu, Y.-J., Wu, Y. (Ariel), Pan, T.-Y., & Yu, N.-H. (2024). Exploring the potential of mixed reality in enhancing language acquisition for EFL children. *Companion of the 2024 on ACM International Joint Conference on Pervasive and Ubiquitous Computing* (pp. 776-780). Retrieved from <https://dl.acm.org/doi/10.1145/3675094.3678388>
16. Ibrahim, A., Huynh, B., Downey, J., Hollerer, T., Chun, D., & O'donovan, J. (2018). ARbis pictus: A study of vocabulary learning with augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 24(11), 2867-2874. <https://doi.org/10.1109/TVCG.2018.2868568>
17. Ke, L., Ye, M., Danelljan, M., Liu, Y., Tai, Y.-W., Tang, C.-K., & Yu, F. (2023). Segment anything in high quality. In *Proceedings of the 37th International Conference on Neural Information Processing System*, 1303, 29914-29934. <https://doi.org/10.48550/ARXIV.2306.01567>
18. Kelliher, A., Birchfield, D., Campana, E., Hatton, S., Johnson-Glenberg, M., Martinez, C., Olson, L., Savvides, P., Tolentino, L., Phillips, K., & Uysal, S. (2009). SMALLab: A mixed-reality environment for embodied and mediated learning. *Proceedings of the 17th ACM International Conference on Multimedia*, (pp. 1029-1032). Retrieved from <https://dl.acm.org/doi/10.1145/1631272.1631504>
19. Kumar, A., Reddy, P., Tewari, A., Agrawal, R., & Kam, M. (2012). Improving literacy in developing countries using speech recognition-supported games on mobile devices. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1149-1158). Retrieved from <https://dl.acm.org/doi/10.1145/2207676.2208564>
20. Lan, Y.-J. (2025). Unlocking language learning with immersive technology: A dive into extended reality (XR). In L. McCallum & D. Tafazoli (Eds.), *The Palgrave encyclopedia of computer-assisted language learning* (pp. 1-10). Switzerland: Springer Nature. https://doi.org/10.1007/978-3-031-51447-0_21-1
21. Lindgren, R., Tscholl, M., Wang, S., & Johnson, E. (2016). Enhancing learning and engagement through embodied interaction within a mixed reality simulation. *Computers & Education*, 95, 174-187. <https://doi.org/10.1016/j.compedu.2016.01.001>
22. Liono, R. A., Amanda, N., Pratiwi, A., & Gunawan, A. A. S. (2021). A systematic literature review: Learning with visual by the help of augmented reality helps students learn better. *Procedia Computer Science*, 179, 144-152. <https://doi.org/10.1016/j.procs.2020.12.019>
23. Macedonia, M., & Knösche, T. R. (2011). Body in mind: How gestures empower foreign language learning. *Mind, Brain, and Education*, 5(4), 196-211. <https://doi.org/10.1111/j.1751-228X.2011.01129.x>
24. Macedonia, M., Mathias, B., Lehner, A. E., Reiterer, S. M., & Repetto, C. (2023). Grasping virtual objects benefits lower aptitude learners' acquisition of foreign language vocabulary. *Educational Psychology Review*, 35(4). Retrieved from <https://link.springer.com/article/10.1007/s10648-023-09835-0>
25. Madan, C. R., & Singhal, A. (2012). Encoding the world around us: Motor-related processing influences verbal memory. *Consciousness and Cognition*, 21(3), 1563-1570. <https://doi.org/10.1016/j.concog.2012.07.006>
26. Mathias, B., Andrä, C., Schwager, A., Macedonia, M., & Von Kriegstein, K. (2022). Twelve- and fourteen-year-old school children differentially benefit from sensorimotor- and multisensory-enriched vocabulary training. *Educational Psychology Review*, 34(3), 1739-1770. <https://doi.org/10.1007/s10648-021-09648-z>
27. Mayer, R. E. (2002). Multimedia learning. *Psychology of learning and motivation* (Vol. 41, pp. 85-139). San Diego, CA: Elsevier. [https://doi.org/10.1016/S0079-7421\(02\)80005-6](https://doi.org/10.1016/S0079-7421(02)80005-6)

28. Yaniafari, R. P., & Olivia V. (2022). The potential of ASR for improving english pronunciation: A review. *KnE Social Sciences*. Retrieved from <https://kneopen.com/kne-social/article/view/10670/>
29. Ratcliffe, J., Ballou, N., & Tokarchuk, L. (2021). Actions, not gestures: Contextualising embodied controller interactions in immersive virtual reality. *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology* (pp. 1-11). Retrieved from <https://dl.acm.org/doi/10.1145/3489849.3489892>
30. Ratcliffe, J., & Tokarchuk, L. (2020). Evidence for embodied cognition in immersive virtual environments using a second language learning environment. In *Proceedings of IEEE Conference on Games (CoG)* (pp. 471-478). Retrieved from <https://ieeexplore.ieee.org/document/9231752/>
31. Ravi, N., Gabeur, V., Hu, Y.-T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K. V., Carion, N., Wu, C.-Y., Girshick, R., Dollár, P., & Feichtenhofer, C. (2024). *SAM 2: Segment anything in images and videos*. <https://doi.org/10.48550/ARXIV.2408.00714>
32. Schmidt, M., Benzing, V., Wallman-Jones, A., Mavilidi, M.-F., Lubans, D. R., & Paas, F. (2019). Embodied learning in the classroom: Effects on primary school children's attention and foreign language vocabulary learning. *Psychology of Sport and Exercise*, 43, 45-54. <https://doi.org/10.1016/j.psychsport.2018.12.017>
33. Schurgin, M. W. (2018). Visual memory, the long and the short of it: A review of visual working memory and long-term memory. *Attention, Perception, & Psychophysics*, 80(5), 1035-1056. <https://doi.org/10.3758/s13414-018-1522-y>
34. Syberfeldt, A., Danielsson, O., Holm, M., & Wang, L. (2015). Visual assembling guidance using augmented reality. *Procedia Manufacturing*, 1, 98-109. <https://doi.org/10.1016/j.promfg.2015.09.068>
35. Toumpaniari, K., Loyens, S., Mavilidi, M.-F., & Paas, F. (2015). Preschool children's foreign language vocabulary learning by embodying words through physical activity and gesturing. *Educational Psychology Review*, 27(3), 445-456. <https://doi.org/10.1007/s10648-015-9316-4>
36. Webb, S., & Nation, I. S. P. (2017). *How vocabulary is learned*. Oxford, England: Oxford University Press.
37. Weerasinghe, M., Biener, V., Grubert, J., Quigley, A. J., Toniolo, A., Pucihar, K. Č., & Kljun, M. (2022). *Vocabulary: Learning vocabulary in AR supported by keyword visualisations* (Version 1). arXiv. Retrieved from <https://arxiv.org/abs/2207.00896>
38. Yilmaz, R. M., Topu, F. B., & Takkaç Tulgar, A. (2022). An examination of vocabulary learning and retention levels of pre-school children using augmented reality technology in English language learning. *Education and Information Technologies*, 27(5), 6989-7017. <https://doi.org/10.1007/s10639-022-10916-w>

VerbVenture: Designing and Evaluating a Mixed Reality and AI-Supported System for Children's English Collocation Learning

Chia Hsu* Cheng-Chih Tsai**

Tse Yu Pan**** Yi Ju Ariel Wu**** Neng Hao Yu*****

Department of Design, National Taiwan University of Science and Technology

* rossannahsu@gmail.com

***** Corresponding author: jonesyu@ntust.edu.tw

College of Industry-Academia Innovation, National Taiwan University of Science and Technology

**m11252011@mail.ntust.edu.tw

***typan@ntust.edu.tw

Department of English Instruction, University of Taipei

**** yjarielwu@gmail.com

Abstract

This study validates an innovative mixed reality (MR) instructional design for children's English collocation learning. Grounded in embodied cognition theory, the design aims to transform everyday environments into immersive language learning spaces, addressing challenges EFL learners face due to first-language interference and limited contextual practice. A prototype system, VerbVenture, was developed to implement this approach through two learning stages: (1) Vocabulary Acquisition—AI-based object recognition helps learners explore real-world objects and learn their English names and pronunciations; and (2) Collocation Practice—contextual animations and dynamic prompts guide learners to manipulate physical or virtual objects, enacting verb–noun collocations to strengthen language-action connections.

A within-subjects experiment with 14 elementary students compared physical and virtual object interactions using pre-, post-, and delayed post-tests. Both modes significantly improved learning and retention, though no statistical difference emerged between them. Qualitative findings

revealed complementary advantages: physical interactions fostered intuitive control and motor memory, while virtual ones enhanced engagement and imagination.

This study contributes empirical evidence for MR-based embodied learning and derives design principles for integrating embodied cognition and AI technologies (e.g., object, motion, and speech recognition) in educational MR systems, informing the design of engaging, intelligent language learning tools.

Keywords: Mixed Reality, AI in Education, Embodied Learning, Contextual Learning, Language Learning.