

Dec 29, 2025

Version 2

Multimodal Instruction with AI-Generated Images for Noun Retention: Exploring Semantic Scene and Materiality Effects V.2

 [PLOS One](#)

DOI

<https://dx.doi.org/10.17504/protocols.io.j8nlk1bx6g5r/v2>

Ye Gaojie¹, Yan Shibo^{2,3}

¹Public Course Teaching Department, Anhui Vocational College of Defense Technology, Lu'an, China;

²Faculty of Engineering, Science and Technology, Kuala Lumpur University of Science and Technology, Kuala Lumpur, Malaysia;

³Department Of Information Technology, Anhui Vocational College of Defense Technology, Lu'an, China

zw27f



Yan Shibo

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.j8nlk1bx6g5r/v2>

External link: <https://doi.org/10.1371/journal.pone.0334778>

Protocol Citation: Ye Gaojie, Yan Shibo 2025. Multimodal Instruction with AI-Generated Images for Noun Retention: Exploring Semantic Scene and Materiality Effects. **protocols.io** <https://dx.doi.org/10.17504/protocols.io.j8nlk1bx6g5r/v2>

Manuscript citation:

Ye G, Yan S (2026) Multimodal instruction with AI-generated images for noun retention: Exploring semantic scene and materiality effects. PLOS One 21(4). doi: [10.1371/journal.pone.0334778](https://doi.org/10.1371/journal.pone.0334778)

License: This is an open access protocol distributed under the terms of the **[Creative Commons Attribution License](#)**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

We use this protocol and it's working

Created: December 20, 2025

Last Modified: December 29, 2025

Protocol Integer ID: 235525

Keywords: visual content into english noun vocabulary instruction, images for noun retention, multimodal instruction with artificial intelligence, multimodal instruction with ai, multimodal instruction, contextual features that facilitate memory, integrating multimodal instruction, multimodal materials with memory, facilitate memory, english noun vocabulary instruction, meaningful vocabulary learning, multimodal presentation, visual instruction as an end, noun retention, treating visual instruction, memory retention, memory formation, multimodal group, cognitive principles of memory formation, large effect sizes for memory retention, memory, generative image technology, semantic scene, exploring semantic scene, semantic rating task, instructional design in language education, semantic understanding, multimodal condition, language education, combining multimodal material, generated image, visual material, english learner, pedagogical potential, recall test, cognitive principle, text, contextual feature, image, delayed r

Abstract

This study explores the effectiveness of integrating multimodal instruction with artificial intelligence (AI)-generated visual content into English noun vocabulary instruction, as compared to text-only instruction. Rather than treating visual instruction as an end in itself, the approach leverages generative image technology to create contextually relevant stimuli that align with cognitive principles of memory formation. A controlled experiment (text-only vs. text + AI-generated images) was conducted with 40 English learners recruited from China. Participants completed immediate and delayed recall tests, definition selection, image-to-word matching (available only in the multimodal condition), and semantic rating tasks. Results revealed that the multimodal group significantly outperformed the text-only group across all measures, with large effect sizes for memory retention and semantic understanding. However, the study design does not allow us to attribute this advantage to the AI-generated nature of the images, as no condition with traditional images was included. These findings indicate that multimodal presentation can support durable and meaningful vocabulary learning when visual materials are designed to reflect perceptual and contextual features that facilitate memory. The study highlights the pedagogical potential of combining multimodal materials with memory-informed instructional design in language education.

Session 1 (Traditional Condition)

1 Learners studied 200 nouns(20 groups, one group every week) using text-only materials, including word,phonetic transcription, definition, and example sentence. No images were provided.

1.1 1. grinder /'graɪndə(r)/

释义：n. 研磨机

例句 1：The coffee grinder can make fresh coffee powder at home.

2. steamer /'sti:mə(r)/

释义：n. 蒸锅；蒸汽器

例句 1：The steamer can cook vegetables without losing nutrients.

3. dehumidifier /,di:hju:'mɪdɪfaɪə(r)/

释义：n. 除湿机

例句 1：The dehumidifier can reduce the humidity in the basement.

4. microwave /'maɪkrəweɪv/

释义：n. 微波炉

例句 1：The microwave can heat leftovers in a few minutes.

5.monitor /'mɒnɪtə(r)/

释义：n. 监视器；显示器

例句 1：The security monitor can keep an eye on your home when you are out.

6.charger /'tʃɑ:dʒə(r)/

释义：n. 充电器

例句 1：The fast charger can charge a smartphone in one hour.

7.adapter /ə'dæptə(r)/

释义：n. 适配器；转接器

例句 1：The adapter can convert a two-pin plug to a three-pin plug.

8.controller /kən'trəʊlə(r)/

释义：n. 控制器

例句 1：The remote controller can operate the TV and air conditioner.

9.sensor /'sensə(r)/

释义：n. 传感器

例句 1：The motion sensor can turn on the light when someone enters the room.

10. filter /'fɪltə(r)/

释义：n. 过滤器

例句 1：The water filter can remove impurities from tap water.

2 Session 2 (Multimodal Condition)

2.1 Basic Learning: Learners were presented with the target word, its pronunciation, and an initial AI-generated image designed to support basic recognition. Once the learner felt

the word was memorized, they could manually proceed to the next stage.

1.purifier /'pjʊərɪfaɪə(r)/

释义：n. 净化器

例句 1：This air purifier can effectively remove PM2.5 from the room.



- 2.2 Semantic Expansion: Learners continued studying the same word, accompanied by a series of AI-generated images (typically five or more), each emphasizing different materials, contexts, or perspectives. Images were displayed one at a time, and learners could click to view the next image at their own pace, allowing for individualized semantic exploration.











- 2.3 Contextual Application: Learners engaged with additional AI-generated images paired with simple contextual sentences (typically three), illustrating the word's use in real-world scenarios. As in the previous stage, image and sentence progression was learner-controlled, enabling flexible and personalized semantic generalization.

例句 1：This air purifier can effectively remove PM2.5 from the room.



例句 2 : We need to replace the filter of the purifier every six months.



例句 3 : Many families choose to buy a purifier to improve indoor air quality.



3.Assessment Procedures

3 To evaluate learners' retention and semantic understanding of vocabulary, assessments were conducted at two distinct time points:

- Immediate Recall Phase: administered directly after each learning session;
- Delayed Recall Phase: conducted 48 hours after the initial learning session.

3.1 Immediate Recall Phas:

1. Select the correct definition for the word grinder.

- A. n. 搅拌机（用于混合液体）
- B. n. 研磨机（用于将咖啡、香料等磨成粉末）
- C. n. 切割机（用于将木材切成小块）
- D. n. 冷藏容器（用于低温储存食物）

2. Select the correct definition for the word dehumidifier.

- A. n. 加湿器（增加空气湿度）

- B. n. 除湿机 (降低空气湿度)
 - C. n. 空气净化器 (过滤 PM2.5)
 - D. n. 冷风机 (吹出冷风)
3. Select the correct definition for the word sensor.
- A. n. 开关 (手动控制通断)
 - B. n. 传感器 (自动检测变化)
 - C. n. 插座 (提供电源接口)
 - D. n. 遥控器 (远距离操作)
4. Select the correct definition for the word filter.
- A. n. 漏斗 (导流液体)
 - B. n. 过滤器 (去除杂质)
 - C. n. 搅拌棒 (混合液体)
 - D. n. 保温瓶 (保持温度)
5. 请看图片, 从选项中选出图片所示物品的英文名称。



- A. n. 电磁炉 (用电生热)
- B. n. 烤箱 (烘烤食物)
- C. n. 微波炉 (快速加热或解冻)
- D. n. 电饭煲 (煮饭专用)

释义题（用中文写出词义）

1. Definition Writing : steamer : _____ (中文释义)
2. Definition Writing : monitor : _____ (中文释义)
3. Definition Writing : charger : _____ (中文释义)
4. Definition Writing : adapter : _____ (中文释义)
5. Definition Writing : controller : _____ (中文释义)

3.2 Delayed Recall Phase:

【选择题】（每题仅一个最贴切的中文释义）

1. Select the correct definition for the word **microwave**.
 - A. n. 电饭煲（煮饭专用）
 - B. n. 微波炉（快速加热或解冻）
 - C. n. 烤箱（烘烤食物）
 - D. n. 电磁炉（用电生热）
2. Select the correct definition for the word **sensor**.
 - A. n. 开关（手动控制通断）
 - B. n. 插座（提供电源接口）
 - C. n. 传感器（自动检测变化）
 - D. n. 遥控器（远距离操作）
3. Select the correct definition for the word **grinder**.
 - A. n. 冷藏容器（用于低温储存食物）
 - B. n. 切割机（用于将木材切成小块）
 - C. n. 研磨机（用于将咖啡、香料等磨成粉末）
 - D. n. 搅拌机（用于混合液体）
4. Select the correct definition for the word **dehumidifier**.
 - A. n. 加湿器（增加空气湿度）
 - B. n. 空气净化器（过滤 PM2.5）
 - C. n. 冷风机（吹出冷风）
 - D. n. 除湿机（降低空气湿度）
5. 请看图片，从选项中选出图片所示物品的英文名称。



- A、filter _____
- B、charger _____
- C、steamer _____
- D、grinder _____

【释义题】（用中文写出词义）

1. Definition Writing : steamer : _____ (中文释义)
2. Definition Writing : monitor : _____ (中文释义)
3. Definition Writing : charger : _____ (中文释义)
4. Definition Writing : adapter : _____ (中文释义)
5. Definition Writing : controller : _____ (中文释义)

4.Data Analysis

- 4 In order to compare the effectiveness of traditional text-only instruction and multimodal instruction with AI-generated images, this study applied statistical methods suitable for a within-subjects design. Such methods mainly used paired samples t-tests and repeated measures ANOVA. And the analyses were used to examine differences in learner performance under different instructional conditions and between two time points: immediately after learning and 48 hours later.

4.1 Retention Rate Calculation

To better capture how well participants retained vocabulary over time, a retention rate was calculated for each individual. This metric reflects the proportion of correctly recalled items in the delayed test relative to the immediate test. By using this approach, the study was able to assess not only initial learning outcomes but also the durability of memory across conditions.

$$R_i = C_i / T$$

R_i = retention rate of participant i

C_i = number of correctly recalled words by participant i

T = total number of words learned in that condition (i.e., 20)

Group-level retention rates were then averaged and compared across conditions.

4.2 Semantic Association Rating

The learners' semantic generalization ability was assessed by the mean semantic rating score, which was calculated using, based on a 5-point Likert scale.

$$\bar{S} = \frac{1}{n} \sum_{i=1}^n S_i$$

S_i = rating score given by participant i

n = total number of participants

Standard deviation (SD) was also reported to assess variability.

4.3 Definition Accuracy and Matching Tasks

Accuracy scores were computed using (3), representing the proportion of correct responses in each task, so as to evaluate task-specific performance.

$$A = \frac{C}{N}$$

A = accuracy score

C = number of correct responses

N = total number of items

These scores were compared across conditions using paired t-tests to determine whether Multimodal instruction led to significantly higher performance.

4.4 Statistical Testing

- a) Paired Samples t-Test was used to compare mean scores between the traditional and multimodal(AI-generated images) conditions for each task.
- b) Repeated Measures ANOVA was applied to examine interaction effects between instructional condition and time (immediate vs. delayed).
- c) Effect sizes (Cohen's d) were reported to quantify the magnitude of differences.

Supplemental Materials

5



Data.zip 93.4MB



experiment_results.xlsx 26.1KB