AN ADAPTIVE IMMERSIVE PROGRAM TO ASSIST STUDENTS WITH ATTENTION DIFFICULTIES IN STANDARDIZED–TEST VOCABULARY ACQUISITION USING A FIRST-PERSON HEIST GAME

Yibo Hu[1] and Moddwyn Andaya[2]

[1]Saratoga High School, 20300 Herriman Avenue, Saratoga, CA 95070
[2]California State University, Sacramento, 6000 Jed Smith Dr, Sacramento, CA 95819

ABSTRACT

Vocabulary knowledge predicts academic achievement and gates access to higher education through the SAT and ACT, yet conventional preparation depends on sustained attention and verbal working memory — the capacities most impaired in the roughly one in nine United States children who have received an ADHD diagnosis. This paper presents a three-dimensional first-person heist game, built in Unity, in which level progression is gated by recall of standardized-test vocabulary, so that studying constitutes play rather than accompanying it. The system comprises a difficulty-tiered question provider over a pre-generated bank drawn from a twelve-thousand-word corpus, a vault-door state machine binding recall to a consequential in-game action, and a guard behavior system that surfaces additional prompts. A one-week retention study with ten participants found the game group recalled 4.8 words of ten against 3.8 for a matched flashcard control.Although the difference did not reach statistical significance given the limited sample size, the observed effect size suggests promising potential for game-based vocabulary learning and warrants further investigation with a larger participant pool.

KEYWORDS

Game-based learning; serious games; Unity; vocabulary acquisition; attention-deficit/hyperactivity disorder; standardized test preparation; retrieval practice; self-determination theory

1. INTRODUCTION

Vocabulary knowledge is one of the strongest single predictors of academic achievement, and its influence extends well beyond language coursework. Research on diverse middle school populations has demonstrated that knowledge of general academic words explains a considerable and significant share of the variation observed in academic achievement across four separate disciplines [1]. Studies of reading proficiency similarly find that academic vocabulary knowledge is closely bound to comprehension outcomes across item types and proficiency levels, with stronger readers more likely to know target words regardless of the linguistic features of those words [2]. Because the SAT and ACT function as gatekeepers to higher education in the United States, and because both examinations assess words in context rather than in isolation, a student who enters the testing process with a thin vocabulary is disadvantaged before answering a single question. That disadvantage compounds: admission outcomes shape scholarship access, institutional placement, and long-term earning potential.

This burden does not fall evenly. Analysis of the 2022 National Survey of Children's Health found that 11.4 percent of United States children aged three to seventeen had ever received an

David C. Wyld et al. (Eds): BMLI, CACIT, EDUAIT, InWES – 2026
pp. 95-106, 2026. CS & IT - CSCP 2026 DOI: 10.5121/csit.2026.161808

ADHD diagnosis, approximately one in nine [3]. The condition carries well-documented academic consequences, including poor standardized test scores in reading and mathematics, increased grade retention, and relatively low rates of postsecondary education — outcomes that pharmacological and behavioral treatment reduce symptomatically without reliably improving standardized test performance or ultimate educational attainment [4]. A substantial mechanism behind these outcomes is working memory. Children with ADHD score significantly below matched controls on measures of both visuospatial and verbal working memory, with verbal working memory impairment concentrated among younger children [5], and longitudinal work identifies verbal working memory specifically as a mechanism linking ADHD to later vocabulary difficulty [6]. Verbal working memory is the system that holds and rehearses unfamiliar word forms long enough for them to be committed to memory.

The problem, therefore, is structural rather than motivational. Conventional vocabulary preparation relies on flashcards, printed word lists, and repeated silent review — methods that demand sustained voluntary attention and intact rehearsal capacity, the two resources these students have least of. The tools are not merely unappealing to the students who need them most; they are built on the cognitive foundation those students find hardest to supply.

This paper proposes a three-dimensional first-person heist game in which progression through the level is gated by successful recall of standardized-test vocabulary, so that studying is not adjacent to play but is the mechanism of play itself. The design addresses the problem on two fronts. First, it converts passive review into retrieval practice. To open the vault, the player must plant an explosive charge and then correctly answer a vocabulary question drawn from a difficulty-matched bank; an incorrect answer refunds the charge and returns the vault to a locked state, requiring another attempt. Every interaction is therefore an act of recall rather than recognition. This matters because taking a memory test produces substantially better long-term retention than spending the equivalent time restudying [7], an effect that persists across spacing schedules [8] and that has been replicated, albeit at modest magnitude, in classroom vocabulary instruction with primary school learners [9][10].

Second, the design targets the engagement failure directly. Self-determination theory attributes the motivational power of games to their satisfaction of the psychological needs for autonomy and competence [11], and meta-analytic evidence confirms that gamified learning reliably improves perceived autonomy and relatedness, though its effect on perceived competence is minimal [12]. A heist scenario supplies goal structure, escalating stakes, and immediate feedback without requiring the student to generate motivation independently — the demand that conventional study materials place squarely on the learner. This approach is preferable to existing gamified vocabulary tools because those tools gamify the periphery: points, streaks and badges are rewards layered atop an unchanged drilling task. Here the vocabulary question is intended to be load-bearing, such that removing it removes the ability to progress.

2. CHALLENGES

2.1. Generating a valid question bank on a scale

The program draws from a vocabulary corpus of roughly twelve thousand entries, which is far too large to author multiple-choice questions by hand. Automated generation introduces a quality problem: distractors selected at random are frequently too easy, allowing a player to answer correctly by elimination without knowing the target word, and occasionally too hard, if a randomly drawn distractor happens to be a near-synonym of the answer. Either failure invalidates the question. This could be adsample by constraining distractor selection to words sharing the

target's part of speech while excluding words from the same semantic grouping, and by validating a sampled subset of generated questions against human judgment before release. One could additionally reject any question whose distractors fall below a minimum semantic-distance threshold from the answer.

2.2. Integrating assessment without breaking immersion

A vocabulary question that interrupts gameplay risks destroying the engagement the game exists to create. If players experience the question panel as a quiz bolted onto an unrelated activity, the design collapses into a flashcard application with expensive scenery. There is also a risk in the opposite direction: interesting but instructionally irrelevant material imposes extraneous cognitive load and produces a small but reliable negative effect on recall. A heist environment is rich in exactly such material. This could be mitigated by binding each question to a diegetic action the player already intends to perform, so that answering is the act of defusing rather than a pause in it, and by distributing vocabulary exposure across ambient objects rather than concentrating it at a single blocking prompt.

2.3. Calibrating difficulty to the individual learner

A fixed difficulty setting serves almost nobody. A bank of questions calibrated for an advanced student produces repeated failure for a struggling one, and repeated failure at a gated objective is indistinguishable from a broken game. Conversely, questions well below a player's level provide no learning value while consuming play time. The challenge is compounded because the target population is defined partly by difficulty sustaining effort under frustration. This could be resolved by tracking per-player accuracy across a rolling window and shifting the difficulty tier when performance crosses defined thresholds in either direction, and by maintaining a per-player record of previously missed words so that those items resurface at expanding intervals rather than at random.

3. SOLUTION

The program is built in Unity using C# and is organized around three linked components: a vocabulary question system, a progression-gating mechanic, and a non-player character (NPC) behavior system. These components do not operate independently; they communicate through shared singleton references and event listeners so that a single player can move between exploration, combat, and assessment. The flow of the program begins with the player exploring a heist-themed level while avoiding or engaging guard NPCs. Reaching a locked objective requires the player to hold a specific item and interact with the objective, which triggers a vocabulary question drawn from the question provider. A correct answer starts a thirty-second detonation timer that opens the vault, while an incorrect answer reverts the interaction and returns the item to the player, preserving the requirement that progression is earned through successful recall rather than passive exposure.

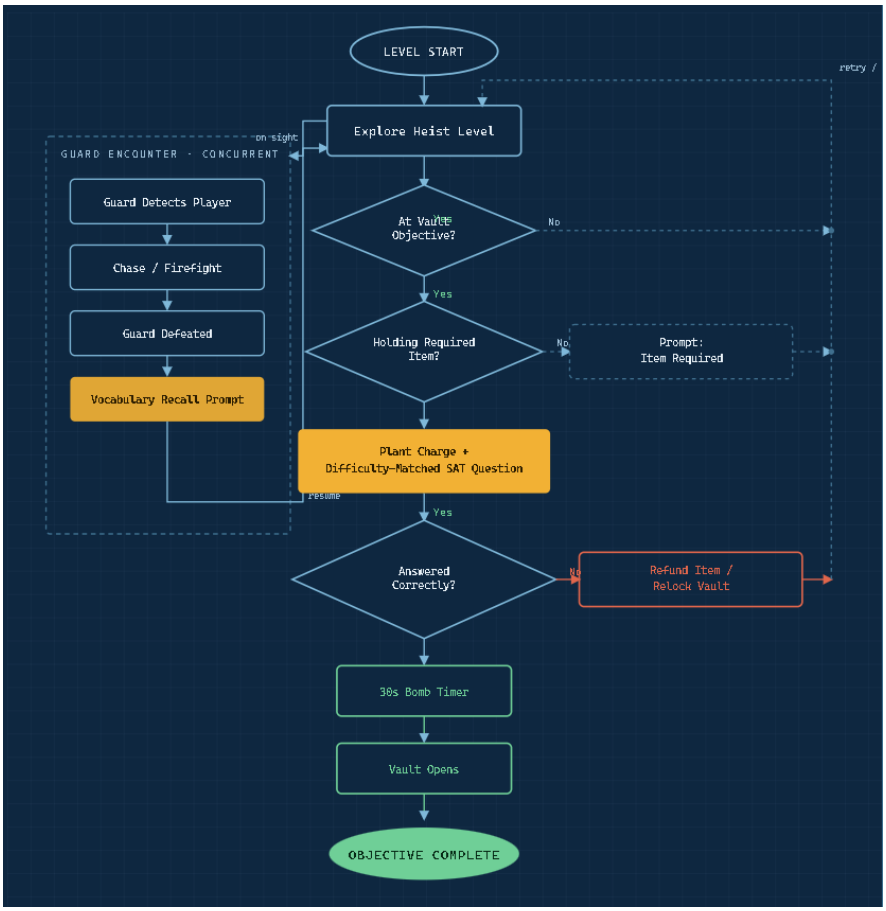


Figure 1. Program flowchart. The vault objective is gated by a difficulty-matched vocabulary question; an incorrect answer refunds the charge and relocks the vault. Guard encounters run concurrently and surface a secondary recall prompt on defeat.

The **progression-gating mechanic** is implemented through a VaultDoor class and determines when questions are presented and what consequence follows a response. It is structured as a finite state machine, a design concept in which a single object's behavior changes according to a small set of defined modes, which prevents invalid actions during each phase. The vault is authored in the Unity editor as a prefab carrying a required item reference, a door animator, an inventory reference, and a configurable bomb duration.

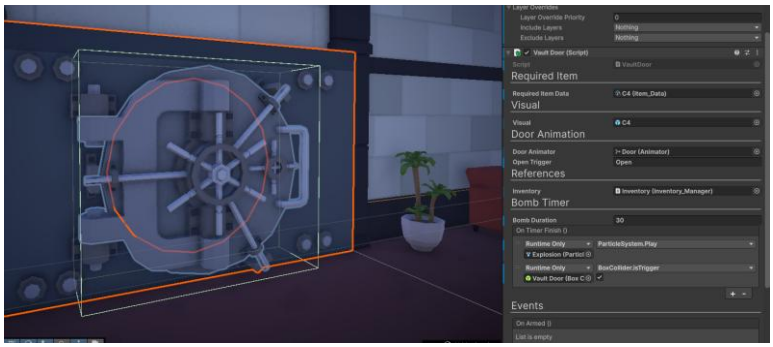


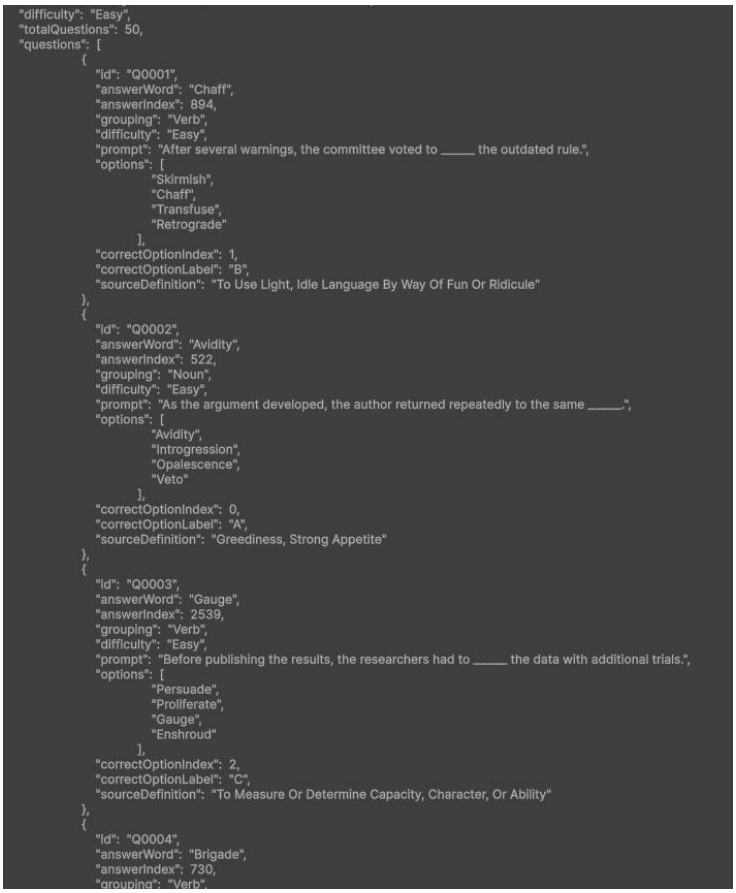
Figure 2. The vault objective in the Unity editor, with the VaultDoor script inspector showing the required item reference, the thirty-second bomb duration, and the timer-completion events that play the explosion and disable the door collider.

```
public class VaultDoor :MonoBehaviour
{
   enum State { Locked, AwaitingAnswer, Ticking, Opened }
   State state = State.Locked;

   void HandleInteracted()
   {
      if (state != State.Locked || !PlayerHasRequiredItem())
         return;

      SATQuestion question = SATQuestionProvider.Instance
.GetRandomQuestion(currentDifficulty);

QuestionUI.Instance.Show(question, OnAnswered);
      state = State.AwaitingAnswer;
   }

   void OnAnswered(bool wasCorrect)
   {
      if (state != State.AwaitingAnswer)
         return;

      state = wasCorrect ?State.Ticking : State.Locked;
      if (!wasCorrect)
RefundItemToPlayer();
   }
}
```

Figure 3. Excerpt from the VaultDoor class. The Locked and AwaitingAnswer states make the vocabulary question a precondition of progression: a wrong answer returns the state to Locked and refunds the charge.

This code runs whenever the player interacts with the vault while the door is still locked. The HandleInteracted method first confirms that the vault is in the Locked state and that the player is holding the required item, since neither condition being met should produce any effect. It then requests a question from the question provider, matches the player's current difficulty, and hands that question to the interface responsible for displaying it, passing a callback method to run once the player responds. The state variable is updated to AwaitingAnswer so that a second interaction cannot be triggered mid-question. The OnAnswered method receives the result of that response and moves the state forward to Ticking on a correct answer or back to Locked on an incorrect one, in which case the required item is returned to the player. No external server is involved; all evaluation happens locally within the running application.

The **vocabulary question system**, centered on a SATQuestionProvider class, supplies the content the rest of the program depends on. It is implemented as a singleton, a pattern in which a class guarantees that only one instance of itself exists and is globally accessible, which allows any other component to request a question without a direct reference wired through the scene. On first use it reads difficulty-tiered data files, parses each entry into a question object containing a prompt, an answer word, and a set of choices, and builds a lookup structure indexed by both difficulty and answer word so that later requests are fast rather than requiring a fresh search through the entire bank.



```
"difficulty": "Easy",
"totalQuestions": 50,
"questions": [
    {
      "id": "Q0001",
      "answerWord": "Chaff",
      "answerIndex": 894,
      "grouping": "Verb",
      "difficulty": "Easy",
      "prompt": "After several warnings, the committee voted to _____ the outdated rule.",
      "options": [
            "Skirmish",
            "Chaff",
            "Transfuse",
            "Retrograde"
          ],
      "correctOptionIndex": 1,
      "correctOptionLabel": "B",
      "sourceDefinition": "To Use Light, Idle Language By Way Of Fun Or Ridicule"
    },
    {
      "id": "Q0002",
      "answerWord": "Avidity",
      "answerIndex": 522,
      "grouping": "Noun",
      "difficulty": "Easy",
      "prompt": "As the argument developed, the author returned repeatedly to the same _____.",
      "options": [
            "Avidity",
            "Introgression",
            "Opalescence",
            "Veto"
          ],
      "correctOptionIndex": 0,
      "correctOptionLabel": "A",
      "sourceDefinition": "Greediness, Strong Appetite"
    },
    {
      "id": "Q0003",
      "answerWord": "Gauge",
      "answerIndex": 2539,
      "grouping": "Verb",
      "difficulty": "Easy",
      "prompt": "Before publishing the results, the researchers had to _____ the data with additional trials.",
      "options": [
            "Persuade",
            "Proliferate",
            "Gauge",
            "Enshroud"
          ],
      "correctOptionIndex": 2,
      "correctOptionLabel": "C",
      "sourceDefinition": "To Measure Or Determine Capacity, Character, Or Ability"
    },
    {
      "id": "Q0004",
      "answerWord": "Brigade",
      "answerIndex": 730,
      "grouping": "Verb",
```


Figure 4. Structure of a difficulty-tiered question file. Each entry carries a cloze prompt, an answer word with its index into the source corpus, a part-of-speech grouping, four options, and the source definition. Each tier file declares fifty questions.

```
public SATQuestion GetRandomQuestion(Difficulty difficulty)
{
List<SATQuestion> pool = _questionsByDifficulty[difficulty];
int index = Random.Range(0, pool.Count);
return pool[index];
}
```

Figure 5. The question-retrieval method of the SATQuestionProvider singleton. Selection within a difficulty tier is uniformly random and stateless.

When called, GetRandomQuestion retrieves the pre-built list of questions matching the requested difficulty and returns one at random. The pool variable represents that filtered list, and index represents the randomly chosen position within it. Two properties of this method qualify claims made elsewhere about the system. First, selection is uniformly random and carries no memory of prior responses, so neither the adaptive difficulty adjustment nor the resurfacing of previously missed words described in Section 2.3 is implemented; currentDifficulty is fixed before play rather than shifted by performance. Second, each tier file declares fifty questions, so the bank in its present form covers roughly two hundred items drawn from the twelve-thousand-entry corpus. The generation pipeline constrains distractors by the part-of-speech grouping field visible in the data, but no semantic-distance filtering or human validation pass has been applied, so the distractor-quality safeguards proposed in Section 2.1 remain partially unrealized.

The **NPC behavior system** governs guard characters that patrol the environment using Unity's navigation mesh, detect the player, and pursue or engage. It relies on a finite state machine similar in concept to the vault door, cycling an NPC between idle, wandering, chasing, shooting, and dead states. Beyond movement, it functions as a secondary trigger for vocabulary content: when an NPC is defeated, its corpse becomes interactable, and interacting with it presents an additional question rather than simply removing the character from the level.

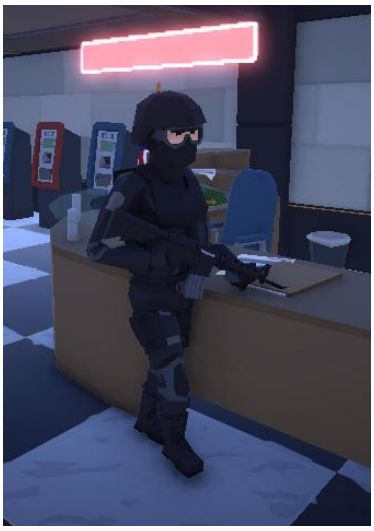


Figure 6. A guard NPC patrolling the heist level. Defeating a guard enables interaction with the corpse, which surfaces a further vocabulary prompt.

```
void HandleHealthDeath()
{
   state = NPCState.Dead;
StopMovement();
EnableCorpseInteraction();
}

void OnCorpseInteracted()
{
   SATQuestion question = SATQuestionProvider.Instance
.GetRandomQuestion(currentDifficulty);
QuestionUI.Instance.Show(question, null);
}
```

Figure 7. Death and corpse-interaction handlers from the NPC controller. The question is displayed with a null callback, so the response carries no consequence.

HandleHealthDeath runs once the character's health value reaches zero, halting its movement and enabling interaction on its corpse. OnCorpseInteracted runs afterward, when the player chooses to interact with that corpse, and requests and displays a question in the same manner as the vault door. The two paths differ in one respect that the user interface does not reveal: the vault door passes OnAnswered as a callback, so the answer determines whether the state machine advances, whereas the corpse handler passes null. Nothing is therefore evaluated, and no consequence follows a correct or incorrect response. The claim that vocabulary recall is load-bearing holds for the vault objective, which cannot be completed without a correct answer, but does not hold for

the guard component, where the prompt is presently decorative. Both methods are executed locally on the player's device.

4. EXPERIMENT

4.1. One-week retention against conventional flashcard study

A possible blind spot is whether gating progression behind vocabulary recall actually produces better one-week retention than conventional flashcard study, since the entire value of the design depends on structural integration outperforming decorative gamification. To test this, ten participants were split into two matched groups of five. The heist game group studied ten difficulty-matched SAT vocabulary words by completing the vault and guard interactions built around those words. The flashcard control group studied the same ten words for an equal amount of time using traditional restudy flashcards. One week later, both groups completed an identical fill-in-the-blank recall test. The design follows the restudy-versus-retrieval paradigm used in the primary school vocabulary work this project draws on, holding study time constant so that the mode of engagement rather than exposure duration is the variable under test.

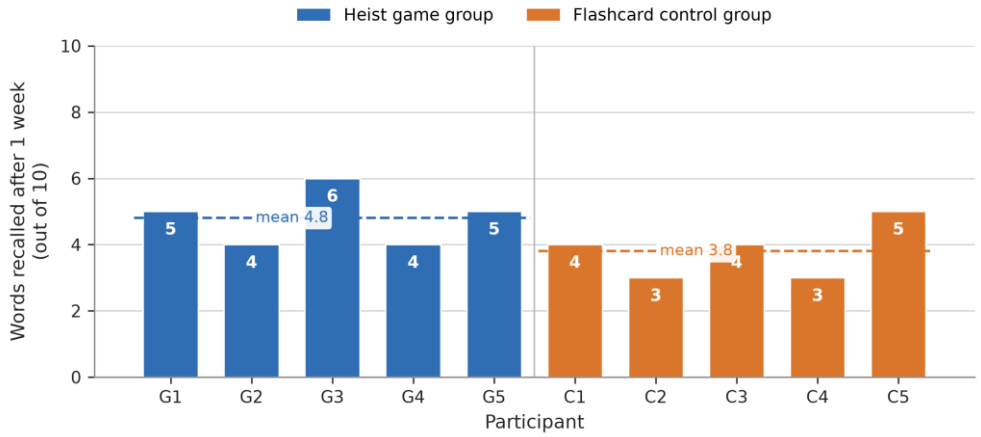


Figure 8. Words recalled after one week by individual participant. Dashed lines mark group means. Every heist-group participant met or exceeded the control mean.

The heist game group recalled a mean of 4.8 words out of 10 after one week, with a median of 5, the lowest score of 4, and a highest score of 6. The flashcard control group recalled a mean of 3.8 words, with a median of 4, the lowest score of 3, and a highest score of 5. Across all 10 participants combined, the mean was 4.3 and the median was 4. The roughly one-word gap between groups is consistent in direction with the retention advantage reported in the retrieval practice literature this design was modeled on, which is reassuring given the small sample. What stood out was that one control participant, C5, matched the heist group's median and exceeded two of its members, showing that individual variance can rival the effect of condition at this scale. The most likely explanation for the overall gap is that the game forces repeated, consequential retrieval attempts before progression is allowed, while flashcards permit passive rereading before a learner chooses to self-test.

4.2. Robustness of the group difference at this sample size

The mean comparison in Section 4.1 leaves a second blind spot: whether a one-word gap between two groups of five reflects a genuine effect of condition or an artifact of who happened to be

assigned where. This matters because the project's central claim is a causal one, and a difference that disappears under scrutiny of its distribution would not support it. The same ten participants' scores were therefore re-analyzed at the distribution level rather than collected afresh, comparing the spread of each group, the range over which the two overlap, and standardized measures of separation. Examining the recorded distribution directly is appropriate here because with five observations per group the mean is sensitive to any single participant, and the degree of overlap reveals how much of the apparent advantage survives that sensitivity.

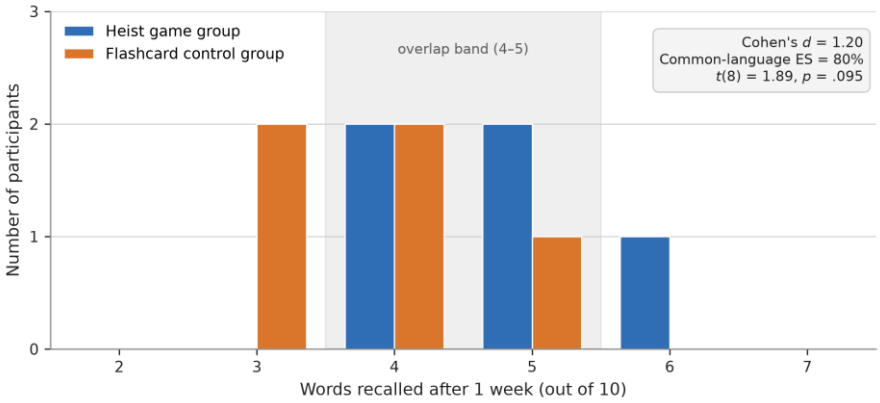


Figure 9. Score distributions of the two groups. The shaded band marks the range in which both groups have members. Effect-size and significance statistics are computed from the same ten observations.

The two groups share identical standard deviations of 0.84, and their scores overlap across the four-to-five band, which contains four of five heist participants and three of five controls. Standardized separation is nonetheless large: Cohen's d is 1.20, and the common-language effect size is 80 percent, meaning a randomly chosen heist participant outscores a randomly chosen control in four of every five pairings. An independent-samples t-test, however, returns t(8) = 1.89 with p = .095, which does not meet the conventional .05 threshold. The most consequential finding is thus a tension between two readings of the same data: the effect is large by the standardized measure and consistent in direction, yet the sample is too small to exclude chance. The overlap band explains why. With only five observations per group and scores confined to integers between three and six, a single participant moving one word in either direction shifts a group mean by 0.2 or materially changes the test statistic. The result supports continuing the design but does not by itself establish the effect.

5. Related Work

McGregor and colleagues conducted a randomized trial of a commercial serious game built explicitly on principles of learning and memory, with forty undergraduates preparing for the Graduate Record Examinations [13]. The intervention worked at the level it targeted: participants mastered an average of one hundred twenty-four words, and definitional accuracy showed no significant decline after a one-month retention interval. Its limitations are consequential. The game produced no measurable gain in verbal practice scores, indicating that word mastery inside the system did not transfer to the high-stakes assessment the participants cared about. The study also examined motivated adult volunteers, a population whose persistence cannot be assumed for adolescents with attention difficulties. This gap is characteristic rather than isolated: meta-analysis across seventy-seven learning comparisons finds serious games more effective than conventional instruction for learning and retention while not more motivating [14], and a systematic review of K–16 game research reports a comparable modest advantage [15]. The

present project targets transfer by drawing questions from the same corpus and closing question format that standardized examinations employ.

Loewen and colleagues examined Duolingo as a case study in mobile-assisted language learning, tracking learners through a semester of independent use and finding a moderate positive correlation between time spent and learning gains [16]. The approach is genuinely effective at scale, and its motivational architecture — streaks, experience points, tiered progression — maps onto the autonomy and competence needs that self-determination theory identifies as the engine of intrinsic motivation [11], needs that gamification does measurably serve for autonomy and relatedness if not for competence [12]. Two limitations bear on the present problem. Gamification is extrinsic to the task: streaks and points reward the act of drilling without altering it, so a learner who stops caring about the streak is left with an unchanged drill. The platform is also oriented toward second-language acquisition rather than the low-frequency academic vocabulary that standardized examinations sample. This design instead makes vocabulary recall the mechanism of progression rather than a scored activity beside it.

The STARS-ADHD trial evaluated AKL-T01, a video-game-based digital therapeutic delivered at home for twenty-five minutes daily across four weeks, in a randomized, double-blind, controlled study of children with ADHD [17]. The intervention improved an objective attention measure relative to digital control, evidence that a game format both sustains adherence and produces measurable change in this population. Systematic reviews of the broader serious-games literature report similar patterns of improvement in attention and executive function alongside consistently positive participant attitudes [18][19], and game-based training has produced generalizable gains in adjacent academic skills such as letter knowledge in beginning readers [20]. The limitation is one of scopes: such systems train attention as a general capacity and deliver little or no academic content, so any benefit to vocabulary is indirect and unmeasured. Occupying the space between these approaches requires care, since instructionally irrelevant but engaging detail raises extraneous cognitive load and produces a small, reliable decrease in recall [21] — a risk that a heist environment carries by construction. This project attempts to deliver the academic content itself through the engagement structure these interventions demonstrate to be effective.

6. CONCLUSIONS

The experiment tested whether gating level progression behind vocabulary recall produces stronger long-term retention than conventional flashcard study, since the project's core claim depends on structural integration outperforming decorative gamification layered on top of an unchanged drilling task. Ten participants were split into two matched groups of five, one studying ten difficulty-matched words through the heist game's vault and guard interactions and one studying the same words for an equal duration using restudy flashcards, with an identical fill-in-the-blank test administered one week later. The game group recalled a mean of 4.8 words against 3.8 for the control. A distribution-level re-analysis of those same scores found a large, standardized effect, with Cohen's d of 1.20 and a common-language effect size of 80 percent, but also substantial overlap between groups and a t-test that did not reach significance at $p = .095$. The direction of the result is consistent with the retrieval practice literature; its magnitude cannot yet be trusted.

The most significant limitation is the weak coupling between the game layer and the study layer. Vocabulary recall is bound to a single consequential interaction, the vault charge, while navigation, guard avoidance and inventory management ignore vocabulary entirely, and the guard prompt evaluates nothing at all. The question panel therefore behaves as an interruption rather than a mechanic, the mechanism by which structurally irrelevant material degrades learning.

With more time, recall would be distributed across the level through keypad codes derived from definitions, safe combinations tied to word meanings, and documents recalled later, making vocabulary the level's connective tissue rather than a gate at its end. Game flow is correspondingly thin: a run consists of one objective, one success state, and no recovery loop after failure, so the escalating challenge that sustains engagement is absent. Staged objectives, each gated by a different vocabulary interaction, would supply that arc. Remaining gaps include a bank of roughly two hundred questions drawn from twelve thousand entries, difficulty fixed before play rather than adapted to performance, no persistent record of missed words to resurface at expanding intervals, and a sample too small to support a statistical claim.

Vocabulary preparation currently asks the students who struggle most to supply the sustained attention they have least of. This project relocates the demand: rather than requiring students to bring motivation to the material, it embeds the material inside a structure intended to generate motivation on their behalf. The approach warrants further evaluation at larger scale and over longer intervals.

REFERENCES

[1] Townsend, Dianna, Alexis Filippini, Penelope Collins, and Gina Biancarosa. "Evidence for the importance of academic word knowledge for the academic achievement of diverse middle school students." The Elementary School Journal 112.3 (2012): 497-518.

[2] Lawrence, Joshua F., Rebecca Knoph, Autumn McIlraith, Paulina A. Kulesz, and David J. Francis. "Reading comprehension and academic vocabulary: Exploring relations of item features and reading proficiency." Reading Research Quarterly 57.2 (2022): 669-690.

[3] Danielson, Melissa L., Angelika H. Claussen, Rebecca H. Bitsko, Samuel M. Katz, Kimberly Newsome, Stephen J. Blumberg, Michael D. Kogan, and Reem M. Ghandour. "ADHD prevalence among U.S. children and adolescents in 2022: Diagnosis, severity, co-occurring disorders, and treatment." Journal of Clinical Child & Adolescent Psychology 53.3 (2024): 343-360.

[4] Loe, Irene M., and Heidi M. Feldman. "Academic and educational outcomes of children with ADHD." Journal of Pediatric Psychology 32.6 (2007): 643-654.

[5] Sowerby, Paula, Simon Seal, and Gail Tripp. "Working memory deficits in ADHD: The contribution of age, learning/language difficulties, and task parameters." Journal of Attention Disorders 15.6 (2011): 461-472.

[6] Gremillion, M. L., T. E. Smith, and Michelle M. Martel. "Verbal working memory as a longitudinal mechanism of vocabulary problems in preschoolers with ADHD." Journal of Psychopathology and Behavioral Assessment 40.1 (2018): 130-138.

[7] Roediger, Henry L., III, and Jeffrey D. Karpicke. "Test-enhanced learning: Taking memory tests improves long-term retention." Psychological Science 17.3 (2006): 249-255.

[8] Karpicke, Jeffrey D., and Henry L. Roediger III. "Expanding retrieval practice promotes short-term retention, but equally spaced retrieval enhances long-term retention." Journal of Experimental Psychology: Learning, Memory, and Cognition 33.4 (2007): 704-719.

[9] Goossens, Nicole A. M. C., Gino Camp, Peter P. J. L. Verkoeijen, Huib K. Tabbers, and Rolf A. Zwaan. "The benefit of retrieval practice over elaborative restudy in primary school vocabulary learning." Journal of Applied Research in Memory and Cognition 3.3 (2014): 177-182.

[10] Goossens, Nicole A. M. C., Gino Camp, Peter P. J. L. Verkoeijen, Huib K. Tabbers, Samantha Bouwmeester, and Rolf A. Zwaan. "Distributed practice and retrieval practice in primary school vocabulary learning: A multi-classroom study." Applied Cognitive Psychology 30.5 (2016): 700-712.

[11] Ryan, Richard M., C. Scott Rigby, and Andrew Przybylski. "The motivational pull of video games: A self-determination theory approach." Motivation and Emotion 30.4 (2006): 344-360.

[12] Li, Liuyufeng, Khe Foon Hew, and Jiahui Du. "Gamification enhances student intrinsic motivation, perceptions of autonomy and relatedness, but minimal impact on competency: A meta-analysis and systematic review." Educational Technology Research and Development 72.2 (2024): 765-796.

[13] McGregor, Karla K., Brittany A. Marshall, Sara K. Julian, and Jacob Oleson. "Learning while playing: A randomized trial of serious games as a tool for word mastery." Language, Speech, and Hearing Services in Schools 50.4 (2019): 596-608.

[14] Wouters, Pieter, Christof van Nimwegen, Herre van Oostendorp, and Erik D. van der Spek. “A meta-analysis of the cognitive and motivational effects of serious games.” Journal of Educational Psychology 105.2 (2013): 249-265.
[15] Clark, Douglas B., Emily E. Tanner-Smith, and Stephen S. Killingsworth. “Digital games, design, and learning: A systematic review and meta-analysis.” Review of Educational Research 86.1 (2016): 79-122.
[16] Loewen, Shawn, Dustin Crowther, Daniel R. Isbell, Kathy Minhye Kim, Jeffrey Maloney, Zachary F. Miller, and Hima Rawal. “Mobile-assisted language learning: A Duolingo case study.” ReCALL 31.3 (2019): 293-311.
[17] Kollins, Scott H., Denton J. DeLoss, Elena Cañadas, Jacqueline Lutz, Robert L. Findling, Richard S. E. Keefe, Jeffery N. Epstein, Andrew J. Cutler, and Stephen V. Faraone. “A novel digital intervention for actively reducing severity of paediatric ADHD (STARS-ADHD): A randomised controlled trial.” The Lancet Digital Health 2.4 (2020): e168-e178.
[18] Lin, J., and W. R. Chang. “Effectiveness of serious games as digital therapeutics for enhancing the abilities of children with attention-deficit/hyperactivity disorder (ADHD): Systematic literature review.” JMIR Serious Games 13 (2025): e60937.
[19] Schultz, Brandon K., Steven W. Evans, Kathryn M. Carter, Amy Dembowski, Kayla Lojinger, Elizabeth R. Murray, and Christy Walcott. “A systematic review and content analysis of serious video games for children with ADHD.” Frontiers in Psychiatry 16 (2025): 1605744.
[20] Glatz, Toni, Wim Tops, Elisabeth Borleffs, Ulla Richardson, Natasha Maurits, Annemie Desoete, and Ben Maassen. “Dynamic assessment of the effectiveness of digital game-based literacy training in beginning readers: A cluster randomised controlled trial.” PeerJ 11 (2023): e15499.
[21] Cheng, C., Y. Wu, and R. Wang. “Seductive details, cognitive load, and learning outcomes: A multi-level meta-analysis and MASEM.” Educational Psychology Review 38 (2026): 28.

©2026 By AIRCC Publishing Corporation. This article is published under the Creative Commons Attribution (CC BY) license.