

**SnapStory: A Mobile Learning App Designed for Kids' English Learning through Snap
Photos and AI-Generated Storytelling**

Jiayu Liu
Katy Ye
Rae Ma
Sijie Wu

Technology, Media, and Learning,
Teachers College, Columbia University

Professor Jooeun Shim
December 17th, 2025

Current English as a Foreign Language (EFL) pedagogies are often characterized by rote memorization and passive input. These learning approaches fail to meet the cognitive demands of the digital-native generation. Furthermore, a significant efficiency loss persists when learners attempt to migrate linguistic knowledge from structured classroom environments to authentic, real-world contexts (Lightbown & Spada, 2013). To address these problems, this paper proposes a solution by integrating the adaptive capabilities of generative AI with a mobile application. Grounded in the cognitive theory of multimedia learning and strategies for fostering intrinsic motivation, we introduce a "photo-to-story" pipeline. This approach transforms real-world exploration into personalized learning materials, offering a critical pathway for resolving current pain points in children's EFL education.

This paper is structured to guide the reader from problem analysis to design implementation. We begin by analyzing the current English-learning challenges and the specific learning characteristics of EFL learners aged 6 to 12. Based on these learning needs, we identify the pedagogical gaps present in existing learning environments. Then, the study proposes a novel solution that a pipeline transforms the learning mode from textbook-driven instruction to curiosity-driven, contextualized exploration. The theoretical section then outlines the learning theories that inspire our workflow. Following this, the design section presents the application's development in two parts: a general structural flow and a detailed explanation of specific functions, demonstrating how users interact with the prototype. Finally, we discuss current limitations and outline a strategy for future iterative design to further enrich the application.

Background and Learning Challenges

Learning Needs

The target audience is children aged 6 to 12 elementary school students learning English as a Foreign Language. This cohort belongs to Generation Alpha who are the children born after 2010. As McCrindle and Fell (2020) note, this generation exhibits a reduced reliance on text and a significantly enhanced preference for visual content, being accustomed to acquiring knowledge through images, videos, and interactive media. For beginners, fundamental EFL learning for this age group requires extensive vocabulary acquisition, contextualized reading, and listening practice. Neuroscientific evidence suggests that for children in this developmental stage, embodied learning modalities such as handwriting, gesturing, or physical interactions like photography facilitate the connection between visual and motor networks, thereby enhancing memory retention (James, 2010). Consequently, effective instructional design must align with these specific characteristics and interests, improving upon existing tools to cultivate a long-term interest in English learning.

Problems

I. De-contextualization and the "Transfer Gap" in Traditional Pedagogy

Historically, vocabulary acquisition has been misunderstood as a process determined solely by mechanical repetition, based on the assumption that words are constantly transferred from short-term to long-term memory. However, cognitive neuroscience indicates that mere repetition establishes only shallow phonological and orthographic connections, failing to construct deep semantic and situational links (Craik & Lockhart, 1972). Stahl and Fairbanks (1986) argued that when EFL learners study English words without contextual clues, the

mechanism for deep memory encoding in the brain is not triggered. Consequently, learners find it difficult to utilize the foreign language in real-life scenarios.

This disconnect in EFL vocabulary acquisition can be addressed as the "transfer gap." When learners are immersed in a highly controlled, decontextualized, and limited-exploration environment, their memory encoding process treats language as isolated symbols. These isolated symbols represent a significant challenge in traditional EFL teaching and learning. Harnad (1990) identified this issue as the "Symbol Grounding Problem": if a symbol system (such as English vocabulary) is explained merely by other symbols (such as Chinese translations or abstract images), these symbols remain "rootless" to the learner, lacking an intrinsic connection to the physical world. For example, for a young child learning the word "apple" solely from a screen image is far less effective than holding a real apple to perceive its weight and flavor. The latter process allows for vivid and authentic grounding, transferring the learning status from floating cognitive concepts to a grounded understanding within a real-world context.

II. Passive Learning and Lack of Engagement

The current target audience (Gen A) is often referred to as "screenagers" due to their extensive exposure to mobile devices. After the COVID-19 pandemic, online learning has reshaped many learners' learning habits. Those who are tired of physical textbooks and classroom confinement have increasingly turned to digital resources on platforms such as YouTube and other online learning platforms. Although this approach offers flexibility, the interaction is often limited to mechanical "clicking and watching." This passive consumption leads to boredom and a rapid decay in motivation. A study of young EFL learners in Chile found that insufficient comprehensible input and limited opportunities to encounter new vocabulary outside the classroom are key factors hindering vocabulary acquisition (Barahona, 2016). More specifically, when learners are viewed passively as containers of knowledge rather than as actors who own their learning, the engagement declines significantly. This disengagement heavily impacts the sustainability of their long-term English learning.

III. Economic Constraints

Traditional learning materials typically include textbooks, workbooks, and tapes, most of which are commonly used by schools for English instruction. Additional resources, such as picture books or after-school tutoring, often cause extra costs for families. Research highlights a stark disparity in resource access, whereas children in middle-income families possess an average of 13 books per child, this ratio declines to one book for every 300 children in low-income families (Neuman & Celano, 2001). This scarcity of physical resources directly contributes to the "word gap" which means that with the same potential, children achieve vastly different outcomes due to unequal access to materials (Hart & Risley, 1995). Furthermore, exacerbated by the recent cost-of-living crisis, National Literacy Trust(2023) reports indicate that approximately 20% of parents have reduced spending on books, while 36.1% of financially struggling families have cut educational resource expenditures. Thus, the affordability of English-learning materials remains a significant barrier to addressing the real-world problems learners face.

To address these systemic problems, this paper focuses on designing a mobile application as an intelligent learning assistant. The goal is to help 6-12-year-old EFL learners cultivate long-term learning habits through exploration driven by their curiosity and imagination, thereby encouraging their interests and sustaining learning energy. The following sections will provide a brief introduction to our strategy for improving the current situation through rational analysis and theoretical implementation.

Solution Overview

A Photo-Story Pipeline

To create a contextualized and engaging learning tool, we propose "SnapStory," an AI-powered English learning application designed to transform physical exploration into linguistic acquisition. The core intervention is a Photo-to-Story Pipeline. This workflow encourages children to use mobile phone cameras to detect surrounding objects and convert them into "digital realia." Driven by their innate curiosity, the child initiates the learning process through embodied interaction.

Once an object is captured and learned via a digital word card, the system advances to a narrative phase. For instance, when a child photographs their own teddy bear and the system generates an English picture story featuring this specific toy, this pattern induces a sense of ownership. This process leverages the self-reference effect, where information related to the self is processed more deeply and remembered better than neutral information. By integrating the learner's personal physical reality into the digital learning material, the pipeline significantly reduces cognitive load and enhances memory retention. This pedagogical design is theoretically grounded in the Cognitive Theory of Multimedia Learning (Mayer, 2014) and Self-Determination Theory (SDT) (Ryan & Deci, 2000), utilizing intrinsic motivation to design a user flow that mimics authentic scenarios rather than rote drills.

Why an AI-Driven Mobile Application

Our technical implementation uses Google AI Studio, leveraging the multimodal capabilities of the Gemini 3.0 model released in 2025. This AI engine combines Natural Language Processing (NLP) with image generation to create personalized reading materials based on the child's historical learning data and current English level. Beyond the software, the application takes advantage of the flexibility and widespread availability of mobile devices. It uses the phone's camera as a primary input tool for detecting objects in the real world. To make learning fun without pressure, the design includes a non-competitive gamification feature based on sticker collections, which aligns with children's natural preferences. Ultimately, this AI-driven approach allows the app to generate an endless stream of personalized stories tailored to each user's interests. It can be an infinitely updatable library. In short, our solution integrates embodied physical action, real-world context, and adaptive AI storytelling into a single mobile learning application.

Theoretical Framework

To ensure our design decisions are educationally and theoretically reasonable, we grounded our platform in two main established frameworks: Cognitive Theory of Multimedia Learning and Motivation and Engagement theories.

I. Multimedia Learning

Driven by the rapid advancement of AI technology, the cost and efficiency of multimodal content generation have improved dramatically, giving rise to educational products that are capable of generating multimedia content in real-time, and enabling continual expansion in both

form and functionality. At the same time, this technological progress makes personalization for educational products more achievable: systems can instantly generate multimedia materials, such as images, text, audio, etc. based on learners' interests, learning pace, and the content they upload, allowing them to engage more deeply and learn more effectively from materials they enjoy. We selected multimedia learning theory as one of the foundational pillars of our platform design because it systematically explains how human beings process multimodal information and how we should leverage this model for reducing cognitive load to further enhance language learning outcomes.

Cognitive theory of multimedia learning brought up by Mayer explains that utilizing various forms of instructions by stimulating learners' multi-sensory memories can help them learn more effectively compared to a single-media design. Specifically, refer to figure 1 below, information first enters the learner's information-processing system through different sensory memories, which is triggered by stimuli presented in the form of text, images, audio, and other modalities (Mayer, 2017). Among all types of sensory memory, visual and auditory are the most commonly used channels for receiving information; however, tactile, olfactory, and gustatory can also be effectively activated when appropriately stimulated, further enriching the learner's sensory input. The coordination of multiple senses can not only deepen memory but also significantly enhances the overall learning outcomes. Learners then select relevant information through attention and perception so that it moves into their working memory for further processing. But only information that is actively processed within about 20s can be kept in working memory, or it is never possible to be transferred further (Mayer, 2017). Next, the learner mentally organizes the incoming information, retrieving prior knowledge from long-term memory, and then integrating the two to store new information into long-term memory. Thus, meaningful learning occurs in this process as learners actively "select, organize, and integrate" new information repeatedly.

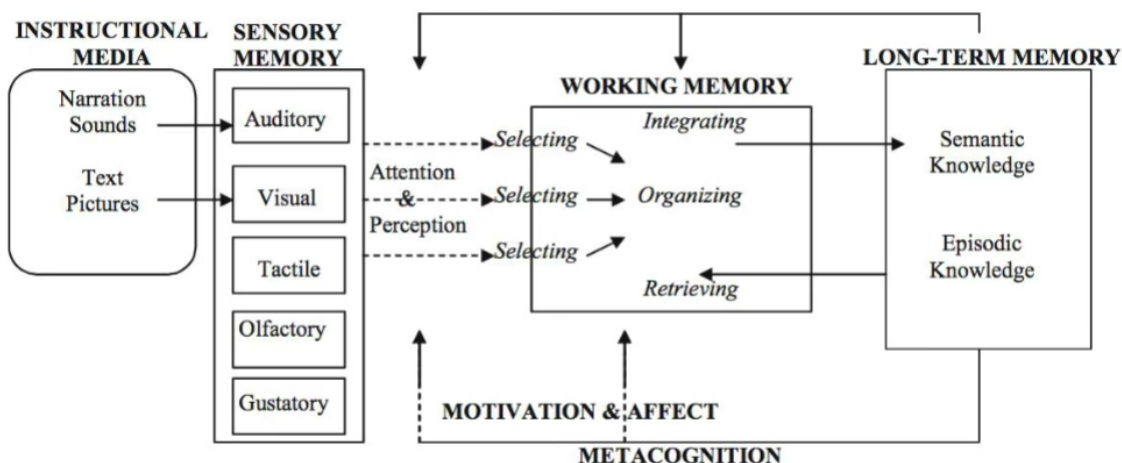


Figure 1 Cognitive-affective model of learning with media

Another important notion emphasized in this theory is that human cognitive load is limited, so it is critical to present the proper amount of information at the right time and in the right format when designing multimedia learning experiences. To prevent cognitive overload and optimize learning, Mayer (2017) proposed a series of design strategies that inform how multimedia content should be structured. Below are the main strategies that guide us in designing our platform SnapStory.

- Mayer's Principles for Reducing Extraneous Processing or Cognitive Load:
 - I.Multimedia Principle: Use graphics and text together
 - II.Coherence Principle: Exclude extraneous material
 - III.Signaling Principle: Highlight essential material
 - IV.Redundancy Principle: Don't overload with graphics, narration AND text
 - V.Spatial Contiguity Principle: Place sign next to corresponding graphic
 - VI.Temporal Contiguity Principle: Present narration and graphics together
 - VII.Segmenting Principle: Allow learners to advance at their own pace
 - VIII.Pre-training Principle: Summarize key elements before
 - IX.Voice Principle: Use a human voice
 - X.Image Principle: Don't include the narrator's image if possible

II. Motivation, Interest and Engagement Framework

While multimedia theory guides how we present information, theories of motivation, interest and engagement explains why a learner chooses to participate and persist, deciding the quality of their involvement. As the definition of engagement stated, it means “the amount of effort one contributes to an activity, but one can be engaged but not interested or motivated.” Thus, in order to activate meaningful engagement, we must look beyond simple participation and involve the learner through promoting high-quality motivation and the development of genuine interest. In this case, we leverage Self-Determination Theory and the Four Phases of Interest Development to guide our platform design.

A. Promoting Intrinsic Motivation via Self-Determination Theory

Motivation is generally categorized into extrinsic motivation, which is driven by external rewards like points, and intrinsic motivation, which refers to “doing something because it is inherently interesting or enjoyable” (Ryan & Deci, 2000). Although well-designed rewarding mechanisms can stimulate users' extrinsic motivation and encourage short-term engagement, sustained learning in the long run relies more on activating users' intrinsic motivation, which ultimately leads to better learning outcomes. Therefore, we grounded our design in Self-Determination Theory (SDT), which explains how intrinsic motivation can be cultivated by fulfilling individuals' basic needs (Deci & Ryan, 2012). In detail, intrinsic motivation flourishes when the below three psychological needs are met:

- **Autonomy:** The feeling of being in control and able to determine the outcomes of one's action.
- **Competence:** This is the need for challenge and a feeling of mastery.
- **Relatedness:** This is the feeling of connection with themselves or with the surrounding environments.

B. Cultivating Interest via the Four-Phase Model

Besides motivation, developing interests is also crucial in improving user engagements. Here, we utilize Renninger and Hidi's Four-Phase Model of Interest Development to help us understand how a learner's interest evolves from a 3-minute passion to a sustained desire to learn. As illustrated in the table below (shown in Fig.2), interest is not static, but rather a variable that develops through interaction with the environment. According to this model, interest progresses through four main phases:

- **Phase I. Triggered Situational Interest:** This is a psychological state resulting from short-term changes in cognitive and affective processing associated with a particular class of content. It is often sparked by environmental features like novelty or surprise, but can die down very fast when the learner encounters difficulties or generates negative emotions.
- **Phase II: Maintained Situational Interest:** This phase involves focused attention to content that reoccurs or persists over time. Here, the learner begins to develop knowledge of the content and a sense of its value, often supported by meaningful tasks or personal relevance.
- **Phase III: Emerging Individual Interest:** This marks the beginning of a relatively enduring predisposition to actively seek reengagement with the content. Learners in this phase have stored knowledge and value, are reflective, and likely to independently reengage with the material.
- **Phase IV: Well-Developed Individual Interest:** This phase is characterized by a relatively enduring predisposition to reengage. Learners at this stage have significant stored knowledge, self-regulate easily to reframe questions, and persevere through frustration to meet their goals.

TABLE 1.2 The Four Phases of Interest Development.

<i>Phases of interest development</i>				
<i>Less developed (earlier)</i>		<i>More developed (later)</i>		
	Phase I: Triggered situational interest	Phase 2: Maintained situational interest	Phase 3: Emerging individual interest	Phase 4: Well-developed individual interest
Definition	<ul style="list-style-type: none"> Psychological state resulting from short-term changes in cognitive and affective processing associated with a particular class of content 	<ul style="list-style-type: none"> Psychological state that involves focused attention to a particular class of content that reoccurs and/or persists over time 	<ul style="list-style-type: none"> Psychological state <i>and</i> the beginning of relatively enduring predisposition to seek reengagement with a particular class of content over time 	<ul style="list-style-type: none"> Psychological state <i>and</i> a relatively enduring predisposition to reengage a particular class of content over time
Learner characteristics	<ul style="list-style-type: none"> Attends to content, if only fleetingly May or may not be reflectively aware of the experience May need support to engage from others and through instructional design May experience either positive or negative feelings May not persevere when confronted with difficulty May simply want to be told what to do 	<ul style="list-style-type: none"> Reengages content that previously triggered attention Is developing knowledge of content Is developing a sense of the content's value Is likely to be able to be supported by others to find connections to content based on existing skills, knowledge, and/or prior experience Is likely to have positive feelings May not persevere when confronted with difficulty May want to be told what to do 	<ul style="list-style-type: none"> Is likely to independently reengage content Has stored knowledge and stored value Is reflective about the content Is focused on their own questions Has positive feelings May not persevere when confronted with difficulty May not want feedback from others 	<ul style="list-style-type: none"> Independently reengages content Has stored knowledge and value Is reflective about the content Is likely to recognize others' contributions to the discipline Self-regulates easily to reframe questions and seek answers Has positive feelings Can persevere through frustration and challenge in order to meet goals Appreciates and may actively seek feedback

Source: This is a revised version of a table originally presented in Renninger and Su 2012.

Figure 2 *The Four Phases of Interest Development by Renninger and Hidi*

Instead of merely stimulating user's interests, we aimed to cultivate their long-term desire to learn through well-structured activities—from providing surprising tasks to promoting creative expression. By doing so, users are more likely to maintain their learning behavior driven by interest and so reach a more meaningful and considerable learning outcome.

Design Overview

SnapStory is a fun English vocabulary and picture book learning application designed for children aged 6 to 12 who are learning English as a second language. To move away from rote memorization and decontextualized learning, we designed a learning process that starts with real-world exploration. Children can take three, five, or seven photos of everyday objects. These photos are automatically recognized by artificial intelligence and transformed into vocabulary cards. These cards are then further linked together to generate a personalized English picture book story with translation features. The entire learning process provides a contextualized, narrative, and generative language learning experience.

I. Design Rationale

After analyzing existing English learning products for children, we identified a common problem: the learning content is mostly preset by developers, such as vocabulary packs with fixed themes, highly repetitive mini-game sequences, or flashcards composed of simple illustrations. For example, Studycat emphasizes vocabulary input driven by themes, but the learning path remains closed; children can only click, match, and repeat within the predetermined topics. EWA Kids uses graded readers to introduce vocabulary, but the narratives are unrelated to children's own experiences. Many flashcard apps remain at the level of looking at pictures and memorizing words. While these methods are convenient, they lack true personalization and fail to create deep semantic connections for children.

Therefore, this project aims to address this gap: making children's daily lives the source of content. Children can freely choose a theme, take 3/5/7 photos of real objects they are interested in, and the system will automatically identify the objects and generate corresponding vocabulary. These images, captured by the children themselves, will then be integrated into a coherent English picture book story generated by AI, transforming vocabulary learning into a real, understandable, and narratable meaning-making process.

This design aims to create a learning process that starts from real-life experiences, centers on story construction, and combines motivational incentives with multimedia support, allowing children to acquire English vocabulary and engage in language expression in a more natural way.

II. Design Goals

1. Building personalized language materials based on life experiences
2. Supporting meaning construction within narrative contexts
3. Enhancing sustained motivation through exploration and feedback mechanisms

III. Instructional Strategies

1. Multimedia Support

Our design is based on key theories of multimodal learning. The dual-coding theory suggests that visual and auditory information, as two complementary channels, can jointly promote knowledge processing; when images, sounds, and text are presented together, learners are more likely to form robust semantic representations (Paivio, 1991). Multimedia learning theory further emphasizes that learning involves integrating linguistic and visual information in working memory; therefore, appropriate use of images and text can effectively improve comprehension efficiency (Mayer, 2017). Related research also shows that multimedia and

interactive reading methods can increase children's engagement and promote language comprehension (Korat & Or, 2010).

In the SnapStory learning process, children continuously experience the integration of visual, auditory, and linguistic information, from photographing real objects to reading storybooks composed of these images. This allows the "image-sound-text" elements to form a natural correspondence within the narrative context, thereby deepening their understanding of language meaning.

2. Motivational Support

To maintain children's focus and sustained engagement in the learning process, this design uses novelty, interactivity, and achievement feedback as core motivational drivers, ensuring that learning remains consistently engaging. Research shows that the development of interest is not instantaneous, and novel experiences play a significant role in driving this process at every stage (Hidi & Renninger, 2006). Therefore, the system provides children with continuous novelty and achievement feedback through constantly changing shooting targets, uniquely generated story content each time, and a sticker reward mechanism. Combined with instant AI feedback, autonomous task selection, and collectible learning achievements, this allows fleeting situational interest to gradually transform into long-term and stable learning motivation.

3. Contextualized Vocabulary Learning Strategy

Research indicates that children acquire the meanings of most vocabulary words not through rote memorization, but by naturally learning them through inferring and extracting meaning from context. For example, Batterink and Neville (2011) pointed out that vocabulary acquisition primarily comes from contextual extraction, rather than simple memorization or explicit instruction.

In our design, children photograph real-world objects and then encounter these words again in AI-generated storybooks, naturally embedding the vocabulary within a narrative context. This process provides children with multiple contextual cues, allowing the meaning of words to be understood in use, rather than being presented in isolation.

IV. Tools

This system relies on model services provided by Google AI Studio and uses tools such as the Camera Input module, AI object recognition, vocabulary card generator, AI co-creation engine, and Storybook Renderer to automate the process from photos to picture books. The application runs on mobile devices with camera capabilities (phones or tablets), while on computers, photos need to be manually uploaded to enter the learning process.

V. Design Structure

SnapStory is centered around an immersive language learning cycle driven by authentic exploration, dividing the learning process into four stages, forming a complete path from perceptual experience to meaning construction, and then to application and reinforcement (*Figure 3*).

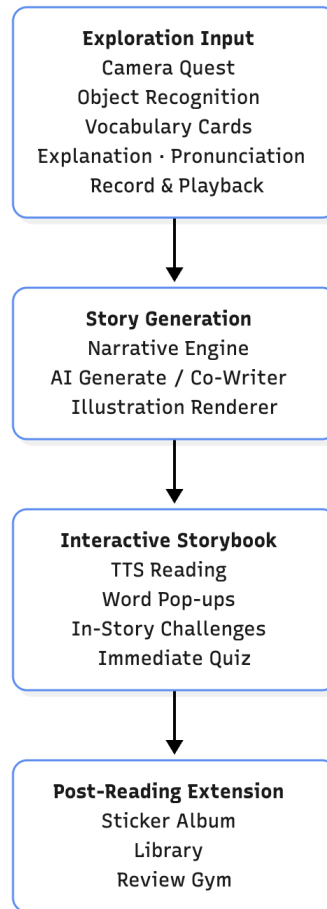


Figure 3 Design Structure Flowchart

Stage 1: Exploration Input

Children begin with real-world situations, using a camera to record everyday objects, making the real environment personalized learning material. The system instantly recognizes images and generates vocabulary cards containing pictures, spelling, definitions, and pronunciation, providing children with multi-channel, multi-modal basic input, and supporting self-monitoring and repetition practice.

Stage 2: Story Generation

After vocabulary collection, the system automatically generates coherent English stories based on the photos, or collaborates with children to brainstorm plots, embedding the exploration results into a narrative structure. This stage transforms memorized material into applicable context, prompting children to understand the meaning and function of vocabulary within a context.

Stage 3: Interactive Storybook

The story is presented as an illustrated storybook with natural TTS narration. An inline translation feature is included in the reading interface. Interactive tools such as highlighting, vocabulary cards, word-finding, and spelling tasks reinforce form–meaning–use connections, and short quizzes provide immediate feedback for consolidation.

Stage 4: Post-Reading Extension

All vocabulary words that appear in the stories are automatically added to a sticker album and personal vocabulary library, forming a visual learning record. The Gym module, based on spaced repetition theory, supports the continuous reinforcement of long-term memory through activities such as flashcards and pronunciation comparisons.

VI. User Flow

To further illustrate the actual operational path of users in SnapStory, this section organizes the children's learning experience into a linear user flow diagram (see Figure 4).

Upon entering the application, children first go through a simple startup interface and create a personal profile. The system automatically adjusts the narrative difficulty, vocabulary density, and guidance prompts of subsequent stories based on the child's age and language level. Then, the user enters the task selection interface, where they can choose tasks from themes such as colors, shapes, and family spaces, or use the surprise mode to have the system randomly assign a target. Children need to select the number of items they wish to find (e.g., 3, 5, or 7) to determine the task difficulty.

Once the task begins, children enter the camera-based treasure hunt phase, using the device's camera to photograph real-world objects that match the theme. The AI instantly determines if the target matches and provides audio-visual feedback. Each time an object is successfully identified, the system automatically generates a visual vocabulary card and adds it to a temporary inventory. The learning cards not only include pronunciation, definitions, and audio playback functions, but also offer click-based inline translation: when children click "Translate," the corresponding Chinese explanation appears directly below the English text, helping them instantly confirm the meaning of words during their exploration without leaving the current interface.

After completing the required number of photos, children can choose between fully automatic story generation or a collaborative creation mode. In the collaborative mode, children simply input simple ideas (such as "I want to go on an adventure to the beach"), and the AI will embed the previously photographed objects into the story structure, naturally integrating the exploration results into the creative activity.

After generating the picture book, the system enters an immersive reading phase. The picture book is presented with cartoon illustrations and segmented text, accompanied by TTS (Text-to-Speech) that simulates natural reading. During reading, the system synchronously highlights the currently read text, strengthening children's understanding of the correspondence between sound and text.

While reading, children can click the "Translate" button to display inline translations. They can also click on any word they want to learn to open a learning card, which can then be saved.

In addition, the reading process includes word-finding games and spelling challenges to further reinforce vocabulary understanding. After reading, the system provides a short quiz to test short-term memory and adds the mastered vocabulary to the personal learning profile. Children can also engage in spaced repetition exercises in the learning gym, such as flashcards and read-aloud comparisons, to continuously reinforce what they have learned.

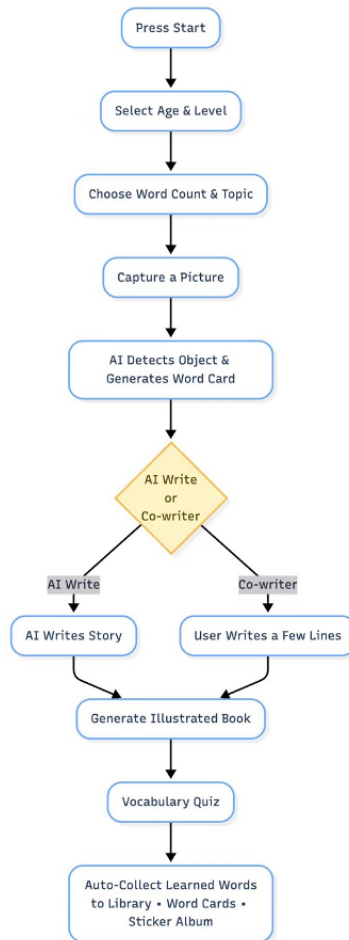


Figure 4 User Flow Flowchart

Alignment of Cognitive Theory of Multimedia Learning (CTML)

1. Multimedia Principle

SnapStory consistently adheres to the principle of combining images, text, and sound in its design. As shown in *Figure 5*, learning cards always appear as a combination of a real-life photograph, the word's spelling, and an audio pronunciation. Similarly, as shown in *Figure 6*, the picture book pages present a corresponding relationship between contextual illustrations, contextual sentences, and human voice narration. This consistent multi-channel input method helps children form strong connections between visual and linguistic information and prevents them from getting bogged down in the abstract nature of text when learning new words.

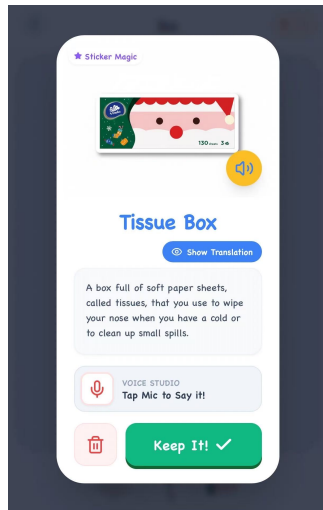


Figure 5 Vocab Learning Card

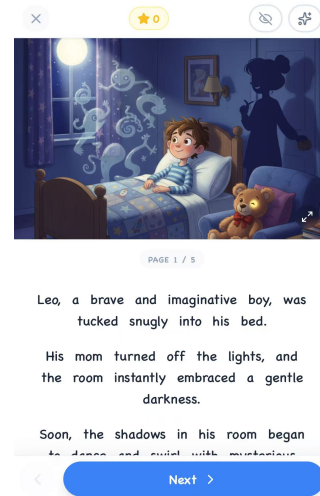


Figure 6 Story Book

2. Coherence Principle

The coherence principle emphasizes removing visual, auditory, or textual elements unrelated to the learning objectives from the teaching materials. In the design of SnapStory, this principle is reflected in the simplification of the interface structure. As shown in *Figures 7* and *8*, when children are reading interactive picture books, the system automatically hides the bottom navigation bar, the top logo, and interface components unrelated to the current task.

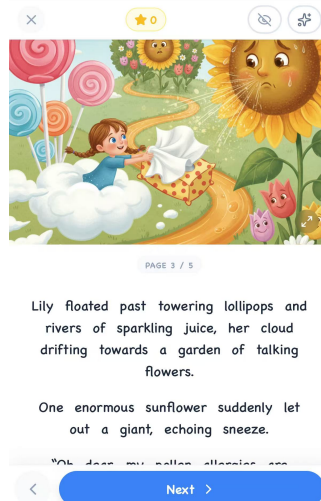


Figure 7 Hidden Bottom Navigation Bar, Top Logo, and Interface Components

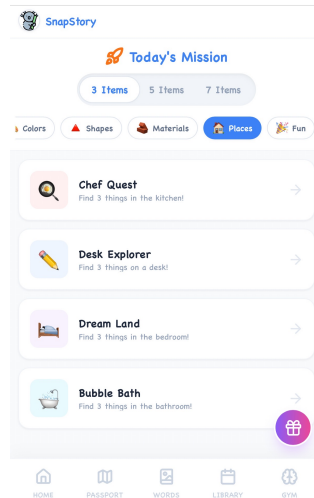


Figure 8 Origin Page With Bottom Navigation Bar, Top Logo, and Interface Components

3. Signaling Principle

To guide children's attention to the learning focus, SnapStory incorporates suggestive visual signals in multiple interfaces. As shown in *Figure 9*, during the story reading stage, the current word is highlighted synchronously when the TTS is reading, allowing children to correspond the sound with the text. As shown in *Figure 10*, a “find on this page” section also appears at the bottom of the story page, clearly indicating the target words on that page, providing a clear focus for the learning task.



Figure 9 Highlighted Book Text While TTS Is Reading

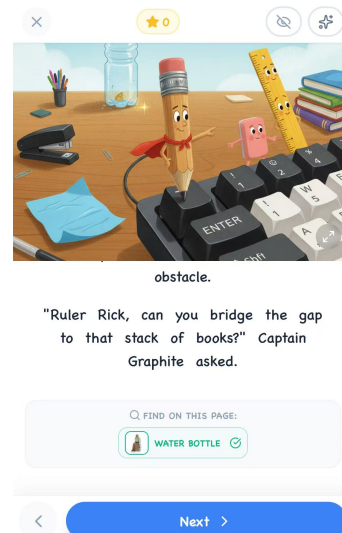


Figure 10 Vocabulary To Be Found Section

4. Redundancy Principle

To prevent users from being overly distracted during the learning process, SnapStory deliberately reduces interface elements unrelated to the learning task. As shown in *Figure 11*, when generating picture book images, we removed the text from the picture book.



Figure 11 Picture Book Image

5. Spatial Contiguity Principle

In all SnapStory interfaces, the text content related to vocabulary and the images are always presented closely together. As shown in Figure 12, when a child clicks on an unfamiliar word while reading a picture book, the vocabulary explanation appears directly on the same page in a pop-up window, instead of jumping to a new dictionary interface. This design allows children to see both the story image and the vocabulary definition simultaneously, without having to switch between different pages.

Similarly, the system's inline translation feature further reinforces spatial contiguity (Figure 12). As shown in Figure 13, when a child taps the "Translation" button, the corresponding Chinese translation is displayed directly beneath the English sentence on the same page, instead of in a separate panel or overlay. By keeping the translated text physically adjacent to the original sentence, the interface minimizes split attention and supports smoother comprehension during reading.

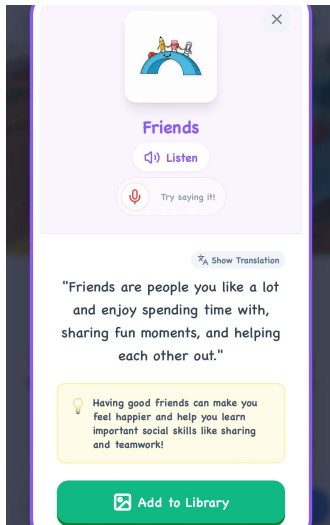


Figure 12 A Pop-Up Learning Card



Figure 13 Translation

6. Temporal Contiguity Principle

SnapStory particularly emphasizes the synchronous appearance of images and narration in all voice presentation scenarios. At the moment the picture book page is turned, the TTS

immediately begins reading while the illustration is loading, avoiding the situation of audio and visual desynchronization.

7. Segmenting Principle

To allow children to digest the content, SnapStory divides the stories and interactive activities into manageable small segments based on the user's age and English proficiency. Typically, a picture book story has around 5 pages, preventing children from seeing too much content at once.

8. Pre-training Principle

The learning path of SnapStory itself is an architecture of pre-training followed by integrated application. Before children begin reading the story, they must first complete a camera scavenger hunt, actively searching for real-world objects containing the target vocabulary words. When these objects are transformed into learning cards and appear in the story, the children already have prior experience with the words.

9. Voice Principle

In story narration and task guidance, SnapStory uses the Kore voice from Gemini TTS, which closely resembles a natural human voice with slight intonation variations, resulting in a soft and friendly listening experience.

10. Image Principle

Snapstory does not feature virtual narrator avatars occupying screen space. Instead, all visual focus is directed towards the story illustrations, photographed objects, vocabulary cards, and other core learning materials. This approach avoids additional visual competition, allowing children to concentrate on the picture book content and task objectives.

Alignment of Interest, Motivation & Engagement

Self-Determination Theory

1. Autonomy

Our “Snap Story” makes children feel they are the ones exploring and deciding what to learn on their own, rather than completing assigned homework. As Self-Determination Theory emphasizes, when learning activities offer choices and allow learners to express their interests and ideas, their sense of autonomy increases, and their motivation becomes more intrinsic.

- 1) Our user can choose among mission topics (*Figure 14*), number of items (3/5/7 items) (*Figure 14*) they would like to learn, and story modes (*Magic Story* which AI generates story directly vs. *Co-Writer* that user can co-create story with AI) (*Figure 15*), giving them meaningful control over how they participate.

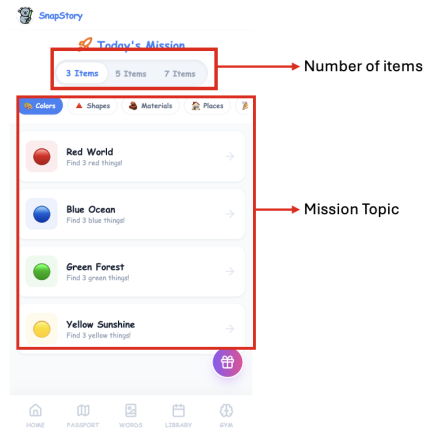


Figure 14 Different choice of mission

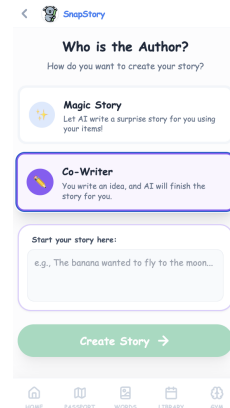


Figure 15 Different choice of creating story modes

- 2) Mission gives the user a topic direction (e.g. Red World: Find 3 red things!) but leaves the specific objects up to the user, supporting self-directed exploration in their own environment. Children can find, for example, what is in red by themselves in the real world. Through taking a photo of the object (*Figure 16*), they can learn the vocab without just reading the vocab book passively.

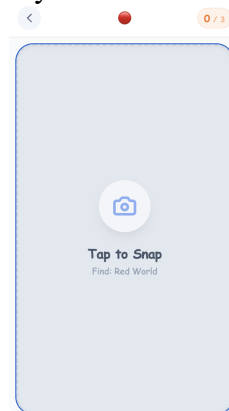


Figure 16 Snap the object to generate a vocab card

- 3) The *Explorer Passport* (*Figure 17*), *Sticker vocab card* (*Figure 18*), and personal *Story Library* (*Figure 19*) turn each child's photos into a unique learning archive. They would feel a sense of ownership over their learning.

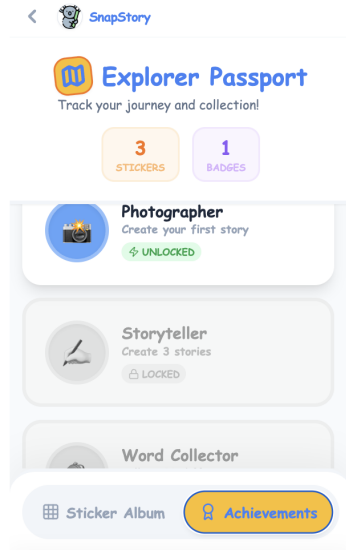


Figure 17 Explorer Passport

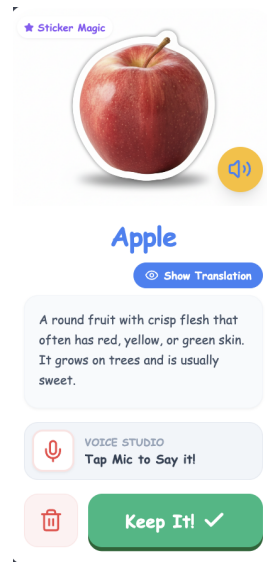


Figure 18 Sticker vocab card

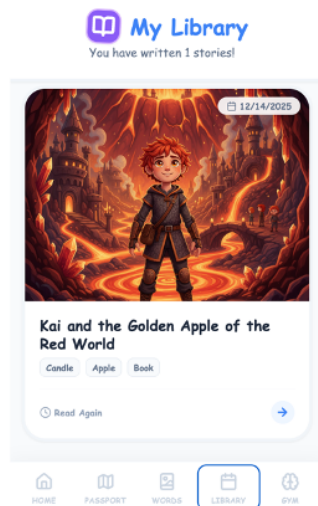


Figure 19 personal Story Library

2. Competence

Learners need to continually experience a sense of competence – “I can do this”, “I’m getting better” – to motivate them to learn and feel engaged. We reinforce this feeling consistently through progressive challenges, immediate positive feedback, and visual growth tracking, thereby supporting sustained motivation to participate.

- 1) The learning progress has layered practice for the same words: *vocab card* during photo recognition, *story “find the word,” Spell It game, reading aloud, end-of-story quiz*, and *vocab gym* where students can review vocabs provide multiple scaffolded opportunities for users to succeed with the same vocabulary.
- 2) The *end-of-story quiz* (Figure 20) helps users assess their mastery of the words, checking whether they have truly learned the words or need more practice.

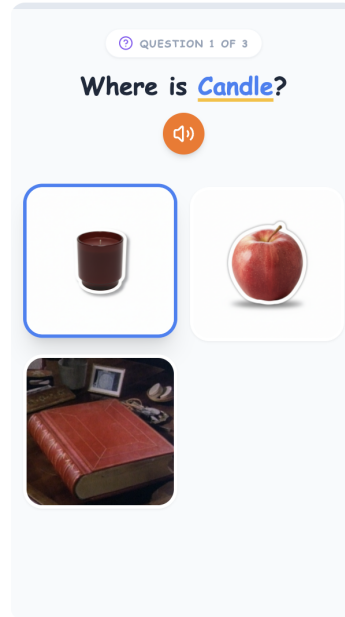


Figure 20 End-of-Story Quiz

- 3) In the SPELL IT game and end-of-story quiz, encouraging messages (“Awesome!”) (Figure 21), stars, and the check mark (Figure 22) offer frequent signals of progress and strengthen learners’ confidence.

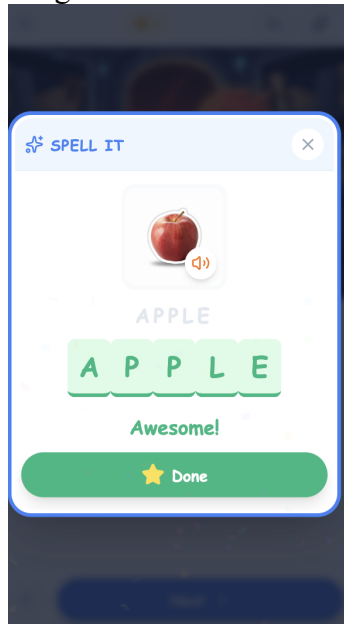


Figure 21 SPELL IT game



Figure 22 After Finding the word and Spelling it correct

- 4) Explorer Passport displays the number of stickers (Figure 23) and badges (Figure 24) the user has collected, allowing children to see how many vocabulary items they have learned and which achievements they have unlocked, strengthening their perceived competence.

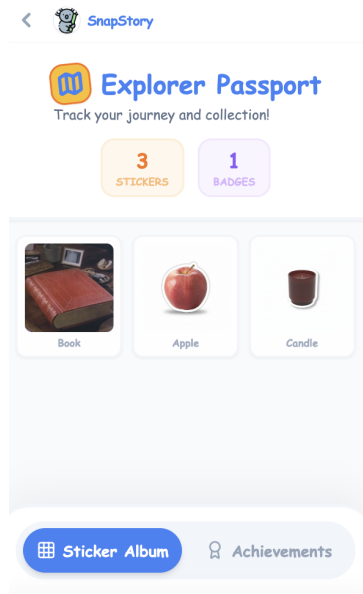


Figure 23 Stickers Album

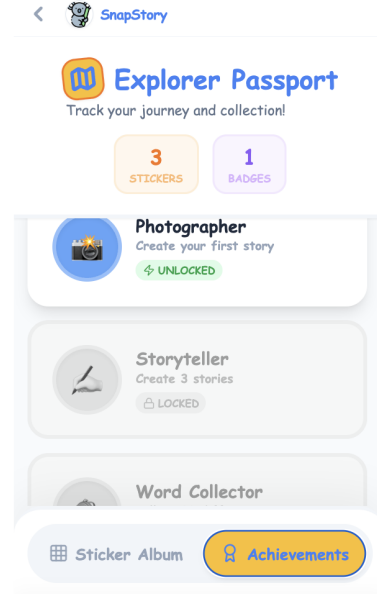


Figure 24 Badges of Achievement

3. Relatedness

In the co-write mode, it creates a shared narrative space where children can create their own story with AI, supporting a sense of connection to the learning environment. Also, users can use our app with parents or friends to explore the world and collect vocab together. They would feel the connection of the learning process and exploring the real-world around them when they do missions.

Four-Phase Model of Interest Development

1. Triggered Situational Interest

Our product combines photo snapping, AI recognition, instant word cards and stickers, and a one-of-a-kind mini story, using a sense of playful novelty to attract elementary school learners.

- 1) With “Tap to Snap,” users can immediately use the device to photograph objects around them, creating a direct link between the real world and the screen. The experience of seeing AI instantly recognize what they just captured (*Figure 25*) feels novel and fun for kids.

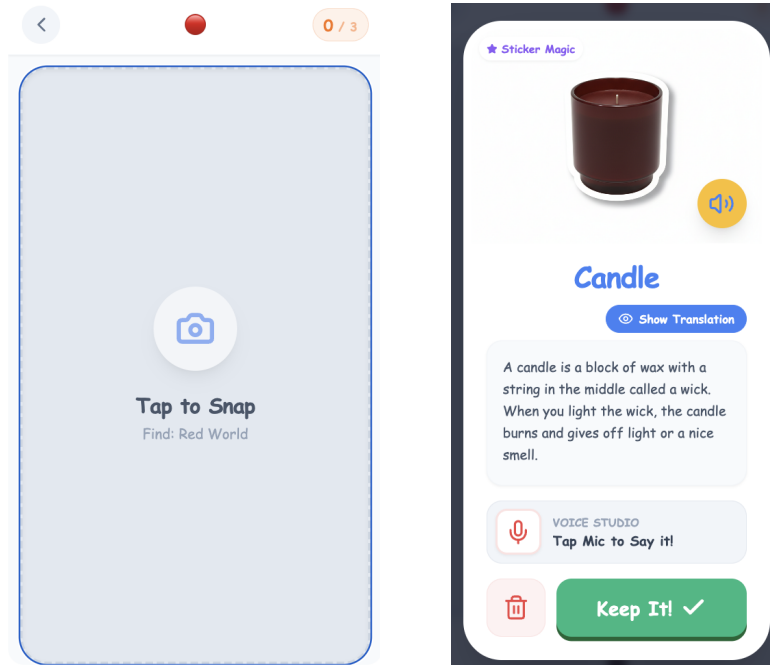


Figure 25 Tap to Snap → AI recognizes item and generates a word card

- 2) Instant rewards make the experience feel playful and motivating: as soon as a user snaps an item, it turns into a sticker (Figure 26), and learning activities can immediately earn stars and other rewards. This quick feedback builds both fun and a sense of achievement, encouraging the user to keep exploring and learning.

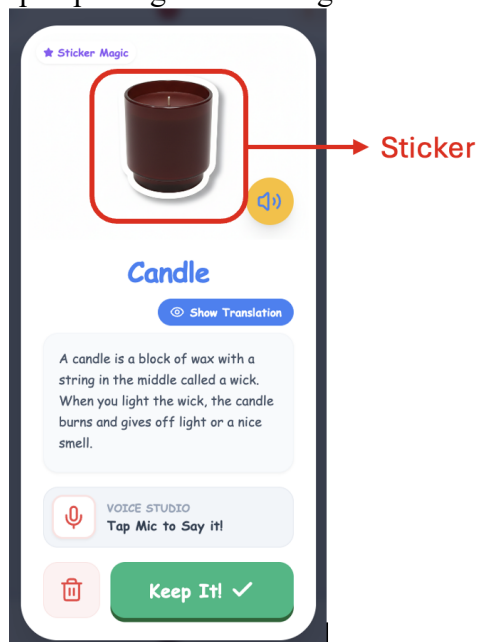


Figure 26 Vocab Sticker

2. Maintained Situational Interest

Once interest is sparked, learners need structured activities and meaningful challenges to keep their attention over time. In our app, we use a consistent learning flow to turn a one-time “snap-and-learn” moment into a sustained learning journey, so engagement doesn’t fade after the

initial novelty. Specifically, our design follows a fixed gamified loop: snap to identify a word → find the word in the story → spell and read it aloud → complete a quiz. The same set of vocabulary words reappears across different contexts and activities, which keeps practice varied rather than repetitive. This repeated, contextual exposure makes learning feel more playful while also strengthening understanding and long-term memory of the words.

3. Emerging Individual Interest

When users photograph objects they personally want to learn, it already reflects emerging individual interest—they're not just completing a task, but following their own curiosity. Our design also includes features that help this personal interest grow further. For example, instead of only reading stories fully generated by AI, users can write a few lines first and then have AI continue the story based on their ideas (*Figure 27*). This allows children to actively express themselves and bring their own imagination into the English learning process.



Figure 27 Co-Writing story with AI

4. Well-Developed Individual Interest

To encourage users to actively engage with our app for long-term learning, we designed several features that help children move from short-term curiosity to a more sustained interest in English learning and creative storytelling.

- 1) Endlessly expandable missions: Our themes will continue to update over time, so learners can keep photographing new objects and generating non-repetitive stories. This supports a growing personal vocabulary album and an evolving collection of storybooks.
- 2) As users learn more, they earn more stickers, stars, and badges. These visible achievements create a strong sense of progress and motivate users to return to the app and keep leveling up.
- 3) Repeatable practice for personal goals: Users can revisit activities based on their needs—either to review vocabulary or to earn more stars. For example, for learners who want extra practice reviewing vocabs, the *Word Gym* (*Figure 28*) offers a structured way to practice words again and again, supporting a more self-directed learning habit.

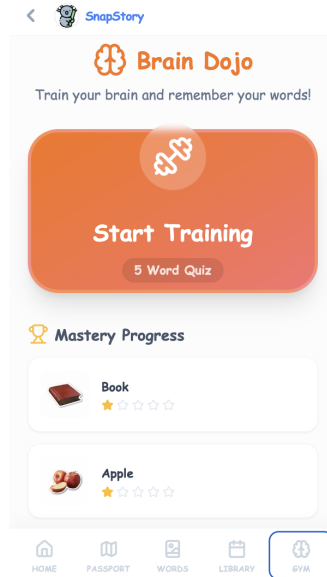


Figure 28 Word Gym

Limitation

While our app offers an engaging way to connect real-world exploration with English learning, it also has several limitations. First, the design currently targets elementary-aged learners who already have some basic English reading ability. Many activities, such as following mission instructions, reading story texts, and completing spelling tasks, assume that children can decode simple words and sentences. Very early readers or complete beginners may feel frustrated without substantial adult support.

Second, the learning focus is strongest on concrete nouns and object-based vocabulary that can be photographed (e.g., “apple,” “book,” “candle”). Abstract words, grammar structures, and more complex language use are only indirectly addressed through the stories. As a result, the app is better suited as a supplement for vocabulary enrichment rather than a comprehensive English curriculum.

Third, the experience relies on consistent access to a mobile device with a camera and stable internet for AI processing. This may limit use in low-resource settings and can raise concerns around screen time and dependency on technology for learning. In addition, because the system stores photos of real environments and belongings, data privacy and parental consent must be carefully managed.

Finally, although there are stories and badges, the app currently offers limited peer or family collaboration features. Learning remains largely individual, and opportunities for joint meaning-making—such as co-reading with parents or sharing stories with classmates—are not yet fully designed. These limitations point to important directions for future iterations, including stronger support for beginners, richer language targets, and more social, low-tech-friendly use cases.

References

- Barahona, M. (2016). Challenges and accomplishments of ELT in Chile: Towards a local perspective. *Colombian Applied Linguistics Journal*, 18(2), 249-260.
- Batterink, L., & Neville, H. (2011). Implicit and explicit mechanisms of word learning in a narrative context: an event-related potential study. *Journal of cognitive neuroscience*, 23(11), 3181–3196. https://doi.org/10.1162/jocn_a_00013
- Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11(6), 671-684.
- Deci, E. L., & Ryan, R. M. (2012). *Self-determination theory*. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (pp. 416–436). Sage Publications Ltd. <https://doi.org/10.4135/9781446249215.n21>
- Harnad, S. (1990). The symbol grounding problem. *Physica D: Nonlinear Phenomena*, 42(1-3), 335-346. [https://doi.org/10.1016/0167-2789\(90\)90087-6](https://doi.org/10.1016/0167-2789(90)90087-6)
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*. Paul H. Brookes Publishing.
- James, K. H. (2010). Sensori-motor experience leads to changes in visual processing in the developing brain. *Developmental science*, 13(2), 279–288. <https://doi.org/10.1111/j.1467-7687.2009.00883.x>
- Korat, O., & Or, T. (2010). How new technology influences parent–child interaction: The case of e-book reading. *First Language*, 30, 139-154. <https://doi.org/10.1177/0142723709359242>
- Lightbown, P. M., & Spada, N. (2013). *How languages are learned* (4th ed.). Oxford University Press.
- Mayer, R. E. (Ed.). (2014). *The Cambridge handbook of multimedia learning* (2nd ed.). Cambridge University Press.
- Mayer, R. E. (2017). Using multimedia for e-learning. *Journal of Computer Assisted Learning*, 33(5), 403–423. <https://doi.org/10.1111/jcal.12197>
- McCrindle, M., & Fell, A. (2020). *Understanding Generation Alpha*. McCrindle Research.
- National Literacy Trust. (2023). *Children and young people's access to books and educational devices at home during the cost-of-living crisis*. <https://literacytrust.org.uk/research-services/research-reports/children-young-people-book-access-cost-of-living/>
- Neuman, S. B., & Celano, D. (2001). Access to print in low-income and middle-income communities: An ecological study of four neighborhoods. *Reading Research Quarterly*, 36(1), 8-26.
- Paivio, A. (1991). *Dual coding theory: Retrospect and current status*. *Canadian Journal of Psychology / Revue canadienne de psychologie*, 45(3), 255–287. <https://doi.org/10.1037/h0084295>
- Renninger, K. A., & Hidi, S. (2015). *The power of interest for motivation and engagement* (1st ed.). Routledge. <https://doi.org/10.4324/9781315771045>
- Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology*, 25(1), 54-67.
- Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research*, 56(1), 72-110.

<u>Application of Learning Sciences Topics</u>	<u>Needs Further Articulation (B-/B)</u>	<u>Adequately Articulated (B+)</u>	<u>Well-Articulated (A-/A)</u>	<u>Comments</u>
<u>Analysis Phase: Problem, Learner, and Context</u> <u>The paper provides documentation of the analysis phase, including a clear statement of the learning need or problem, learner characteristics, and context grounded in real-world conditions.</u>			_____√	<u>See comments below</u>
<u>Design Phase: Goals, Strategies, and Structure</u> <u>The paper presents clear learning goals, instructional strategies, tools, and overall structure aligned with learning needs and theory.</u>			_____√+	<u>See comments below</u>
<u>Demonstration Phase: Prototype Materials</u> <u>The paper includes visuals or materials (e.g., flowcharts, screenshots) that show how the learning environment functions.</u>			_____√	<u>See comments below</u>
<u>Coherence and Professional Presentation</u> <u>The paper is well-organized, clearly written, and professionally presented with appropriate visuals and references.</u>			_____√	<u>See comments below</u>

Analysis Phase: Problem, Learner, and Context. You clearly identify a meaningful and well-supported learning problem in children’s EFL education, including the transfer gap caused by decontextualized vocabulary instruction, passive learning experiences, and unequal access to learning resources. The analysis is grounded in both empirical literature and realistic educational contexts. Your description of the target learners (children aged 6-12, Generation Alpha EFL learners) is thoughtful and detailed. You clearly explain how learners’ developmental characteristics, preferences for visual and interactive media, and embodied learning needs shape the design of SnapStory. This section is very strong but quite long. You could improve clarity by summarizing the core learning needs in a short synthesis paragraph or bullet list before transitioning into the design section.

Design Phase: Goals, Strategies, and Structure.

Your design phase is a major strength of this project. You articulate clear design goals and present a coherent “photo-to-story” learning pipeline that logically follows from your analysis. Instructional strategies are well-aligned with learning theories, including Cognitive Theory of Multimedia Learning, Self-Determination Theory, and the Four-Phase Model of Interest Development. You do an especially strong job explicitly mapping multimedia principles (coherence, signaling, contiguity, segmenting, pre-training, voice, and image principles) to concrete interface and interaction decisions. The staged learning structure (exploration, story generation, interactive reading, and post-reading extension) is clear and pedagogically sound. Given the richness of this section, adding a brief table that maps learning goals → features → theories would help readers quickly see alignment at a glance.

Demonstration Phase: Prototype Materials. You provide extensive and high-quality demonstration materials, including flowcharts, interface mockups, user flow diagrams, vocabulary cards, storybook screens, and gamified feedback elements. These visuals clearly show how learners move through the system and how learning activities build on one another over time. The user flow diagram and staged walkthrough (from snapping photos to story reading and reinforcement activities) are particularly effective and make the learning environment easy to imagine as a functional product. You might further strengthen this section by briefly narrating a single concrete learner scenario (for example, one child completing a full mission from start to finish) to complement the detailed visuals.

Coherence and Professional Presentation. The paper is well-organized, professionally presented, and clearly demonstrates deep engagement with the course material. Figures are well-integrated and consistently referenced, and citations are appropriate and thorough.

Well done!

In addition, your project has been selected as the runner-up for the Best Design Phase. Congratulations!

Final Project Grade: A+