

"I Learned It on TikTok": Conceptualising a Multimodal Framework for Incidental English Vocabulary Learning Through Short-Form Video Content

Jacinta Ann Xavier Belaman^{1*}, Siti Fauziana Zakaria², Diana Ann Xavier³

^{1,2}Akademi Pengajian Bahasa, Universiti Teknologi MARA, Shah Alam, Malaysia

³Sekolah Menengah Kebangsaan Gelang Patah, Gelang Patah, Johor, Malaysia

*Corresponding Author

DOI: <https://doi.org/10.47772/IJRISS.2026.102100131>

Received: 02 September 2026; Accepted: 07 September 2026; Published: 01 October 2026

ABSTRACT

Learners now come across English in many of the digital spaces they use every day. TikTok is one such space, where users may encounter English while watching videos for entertainment, information, or social interaction. Videos on the platform often combine spoken language, visuals, subtitles, on-screen text, captions, and comments. This means that an unfamiliar word may be heard, seen in written form, and understood with the help of what is happening in the video. When this happens without the user deliberately trying to study vocabulary, it can lead to incidental vocabulary learning. Research on audiovisual viewing and informal digital learning shows that vocabulary can develop through this kind of meaning-focused exposure. Not every encounter, however, leads to learning. A word may not attract the learner's attention, its meaning may remain unclear, or it may be forgotten after the video ends. Some words, however, may become more familiar when learners notice them, infer their meaning from the surrounding context, and encounter them again. This paper proposes a Multimodal Incidental Vocabulary Learning Framework to show how this process may occur during everyday TikTok use. It includes six processes: multimodal exposure, noticing, contextual meaning construction, repeated encounters and reinforcement, retention, and potential receptive or productive use. These processes are not expected to occur in the same way for every word or every learner. What happens may depend on the learner's attention and prior knowledge, the support available in the content, and whether the word appears again. The framework therefore focuses on what may happen between first encountering an unfamiliar word on TikTok and developing more stable vocabulary knowledge.

Keywords: Tiktok, Incidental Vocabulary Learning, Multimodal Input, Short-Form Video, Informal Digital Learning

INTRODUCTION

English is no longer encountered only in classrooms or formal learning materials. For many learners, some contact with English occurs through everyday online activities, including watching videos, browsing websites, playing games, using social media, and following entertainment content in English. Lee and Dressman (2018) discuss these activities through the concept of Informal Digital Learning of English (IDLE), where English-related digital activity takes place outside formal instruction. Learning outcomes may also depend on the nature of learners' online engagement. Lee (2019) found that learners who participated in a wider range of IDLE activities tended to have stronger vocabulary knowledge. This suggests that the type of digital activity may matter alongside the amount of exposure.

TikTok provides one example of this kind of informal English exposure. Language on the platform is rarely presented in only one form. A video may contain spoken English, visual cues, subtitles, on-screen text,

written captions, hashtags, comments, and replies. Users may hear a word while also seeing its written form or interpreting it through the surrounding visual context. The same word may later appear in another video, caption, trend, or comment. Much of this exposure occurs while users seek entertainment, information, humour, lifestyle content, or social interaction. English learning is not necessarily the primary purpose of platform use.

When vocabulary develops during an activity in which learning words is not the main goal, it can be understood as incidental vocabulary learning. Webb et al. (2023) describe this type of learning as vocabulary growth that occurs through meaning-focused activity rather than deliberate vocabulary study. Audiovisual research shows that some vocabulary can be learned in this way. Peters and Webb (2018), for instance, found vocabulary gains after learners watched a full-length L2 television programme. Their findings also showed that some words were more likely to be learned than others, with frequency of occurrence and learners' existing vocabulary knowledge playing a role.

Written information can add another source of support when learners watch audiovisual content. Kurokawa et al. (2025) examined this in a meta-analysis involving 49 studies and 89 effect sizes. They found that learners generally gained more vocabulary from captioned viewing than from viewing without captions. Having the spoken and written forms available together may help with processing a new word. The benefit is not always the same, however, because learners differ in what they already know and in how they respond to the material they are watching. These differences are particularly relevant outside formal learning, where users choose what to watch and how much attention to give it.

TikTok has also been examined in relation to vocabulary learning. Many of these studies focus on planned learning, where TikTok is used for teaching or where learners deliberately use the platform to improve their English. Ordinary use presents a different situation. A learner may encounter an unfamiliar English word while watching comedy, lifestyle content, reviews, stories, or other videos that have no language-learning purpose. Sely and Priyatmojo (2026) examine TikTok from this perspective and describe it as an informal digital environment where incidental English vocabulary learning may occur. They point to features such as non-educational content, repeated exposure, code-switching, and interaction through comments as possible sources of vocabulary encounters.

Recognising that TikTok can provide these encounters is only part of the issue. Coming across a word is not the same as learning it. A learner may hear or see an unfamiliar word and pay little attention to it. Another word may be understood from the visual or linguistic context but forgotten soon afterwards. Some words, however, may become more familiar when learners notice them, infer their meaning from the surrounding context, and encounter them again. Over time, some of this knowledge may remain available when the word appears in another situation or when the learner wants to use it.

This leaves an important question about what happens between first exposure and later vocabulary knowledge. Existing work helps explain why TikTok may provide opportunities for incidental learning, but less is known about how a particular lexical encounter might develop after it occurs. The issue is therefore not only whether learners meet English words while using TikTok. It is also whether those words are noticed, understood, encountered again, remembered, and eventually recognised or used.

This paper proposes a Multimodal Incidental Vocabulary Learning Framework to organise these possibilities. The framework includes six processes: multimodal exposure, noticing, contextual meaning construction, repeated encounters and reinforcement, retention, and potential receptive or productive use. These processes are not expected to occur in the same way for every word or every learner, nor do they necessarily form a fixed sequence. A lexical encounter may stop at any point.

What happens may depend on factors such as the learner's attention and prior vocabulary knowledge, the support available in the content, and whether the word is encountered again. By focusing on these processes, the framework considers how an unfamiliar word encountered during everyday TikTok use may, under some conditions, develop into more stable vocabulary knowledge.

LITERATURE REVIEW

Incidental Vocabulary Learning

Learners do not always need to sit down and study a list of words for vocabulary learning to occur. New words can also be picked up while attention is focused on something else. This is the idea behind incidental vocabulary learning, which refers broadly to vocabulary development that takes place through meaning-focused exposure rather than deliberate study (Webb et al., 2023). A learner may come across an unfamiliar word while reading, listening to a conversation, or watching a video. The same can happen on social media, where users may meet new English words while concentrating mainly on the content they are watching or reading.

Audiovisual input provides one setting where this kind of learning can take place. Peters and Webb (2018) examined learners who watched a full-length L2 television programme and found gains in knowledge of unfamiliar words. The gains, however, were not the same for all words or all learners. Words that appeared more frequently were more likely to be learned, while learners with stronger existing vocabulary knowledge tended to benefit more from the viewing experience. Simply seeing or hearing a word does not mean that it will be learned. How often it appears and what the learner already knows may make a difference.

Montero Perez (2020) also found vocabulary gains after audiovisual viewing. Learners became better at recognising unfamiliar word forms and meanings even though they did not know beforehand that their vocabulary would later be tested. Their main task was to watch the material, not to study a list of words. Vocabulary learning happened alongside that activity. A similar situation may occur online when users encounter English while watching something for entertainment or looking for information.

Learning a word usually takes time. A first encounter may make the word seem familiar without giving the learner complete knowledge of it. Schmitt (2008) explains that different parts of word knowledge can develop at different rates. Webb (2007) found that learners gained knowledge of unfamiliar words when they encountered them repeatedly, although they did not fully learn every target word. Uchihara et al. (2019) found a similar pattern across 26 studies involving 1,918 participants, with a medium positive relationship between encounter frequency and incidental vocabulary learning ($r = .34$). Meeting the same word again can help, but the number of encounters alone does not determine whether it will be learned.

Attention matters as well. Incidental learning does not mean that a learner can completely ignore a word and still learn it. Schmidt (1990) argues that linguistic information needs to be noticed before it can be processed further. Written information in audiovisual material may help some words stand out. Montero Perez et al. (2014), for example, found differences between captioned and uncaptioned viewing on some vocabulary measures. They also compared different ways of presenting captions, including full captions, keywords, and highlighted keywords. How a word appears on screen may therefore affect whether learners notice its written form. If they meet it again later, they have another opportunity to pay attention to it and process it further (Uchihara et al., 2019).

Not every word develops in the same way. A learner may see one word several times and still forget it, while another becomes familiar after it is noticed, understood from context, and encountered again. What the learner already knows may also help when trying to make sense of a new word. Vocabulary learning from incidental exposure therefore develops differently from one encounter to another. This is important on social media, where a word may appear only briefly and the user may have no intention of studying it.

Informal Digital Learning of English

Learners may come into contact with English while doing many ordinary things online. They might watch videos, play games, communicate with others, use social media, or search for information. These activities can form part of Informal Digital Learning of English (IDLE) when they take place outside formal instruction. Lee and Dressman (2018) include activities such as entertainment, online communication, social

networking, and gaming within IDLE. Learners are not necessarily taking part in a lesson when they do these things. They encounter English because it is already present in the activity.

This can happen quite naturally on social media. Someone may hear English in a funny video, read it in a discussion, or come across it while searching for information or interacting with other users. Their attention is usually on the content rather than on learning the language. Even so, an unfamiliar word or expression may sometimes become familiar along the way.

Time spent online does not necessarily show how much useful English a learner encounters. Lee (2019) found stronger vocabulary knowledge among learners who participated in a wider range of informal digital English activities. This suggests that the variety of activities may matter alongside the amount of exposure. Two learners may spend the same amount of time on TikTok but encounter very different kinds of English.

One may regularly watch English-language videos, read captions and comments, and interact with the content, while another may come across very little English during the same amount of time.

TikTok is one of the digital spaces where this informal contact with English can happen. Users may hear English in a video, read it in subtitles or comments, or come across a new expression while watching something that is unrelated to language learning. A word may catch their attention because it appears several times or because the surrounding video makes its meaning easier to understand. Another word may pass unnoticed. These everyday encounters create possibilities for vocabulary learning, but they do not mean that learning has already occurred.

Multimodal Input in Audiovisual Language Learning

Audiovisual materials present language through more than one form at the same time. Learners may hear speech, read written text, and use images, actions, gestures, or other visual information to understand what is happening. Mayer's Cognitive Theory of Multimedia Learning helps explain how this kind of input may be processed. According to Mayer (2021), information is processed through visual/pictorial and auditory/verbal channels, and both channels have limited capacity.

Learning therefore depends on how learners select, organise, and connect the information they receive. In language learning, this is relevant because an unfamiliar word may be heard in speech, seen in captions, and understood with the help of the surrounding visual context.

Research on captioned audiovisual material shows that written support can contribute to vocabulary learning. Peters et al. (2016), for example, examined the effects of L1 subtitles and L2 captions on different aspects of vocabulary knowledge. They found that captions were particularly useful for learning written word forms, although the effects were not the same for every aspect of word knowledge. Kurokawa et al. (2025) found a similar advantage in a meta-analysis of 49 primary studies and 89 effect sizes. Captioned viewing showed a significant medium-sized advantage over uncaptioned viewing for incidental L2 vocabulary acquisition ($g = .56$). When learners can both hear a word and see its written form, they may have more than one source of information to help them process it.

Having several modes available, however, does not mean that learners will automatically learn the vocabulary they encounter. Mayer (2021) points out that processing capacity is limited. Learners cannot attend equally to every spoken word, caption, image, and visual detail at the same time. Some information may attract attention while other information is missed. This issue becomes more noticeable in short-form social media, where one video may end within seconds and another begins immediately. A learner may therefore have access to several kinds of support without having enough time or attention to process all of them.

TikTok provides a clear example of this type of multimodal input. A user may hear an English word in a video, see the same word in subtitles or on-screen text, and use images or gestures to infer what it means. Written language can also appear outside the video through captions, hashtags, and comments (Kang & Lou,

2022). These features give users several possible sources of information when they encounter unfamiliar vocabulary.

The conditions on TikTok, however, are not the same as those in controlled audiovisual studies. Much of the content is short, user-generated, and continuously replaced by new videos in the For You feed (Schellewald, 2021; Kang & Lou, 2022). Users also decide for themselves whether to keep watching, replay a video, read the comments, or move on. Vocabulary exposure on the platform can therefore be brief and unpredictable. For this reason, it is not enough to consider only which modes are available. It is also necessary to consider which parts of the input users actually notice and engage with during ordinary TikTok use.

TikTok and Vocabulary Learning

TikTok has received growing attention in language-learning research because of its popularity and its short-form audiovisual format. Studies suggest that the platform can support vocabulary learning, but much of this research has focused on situations where TikTok is deliberately used for educational purposes. This is different from ordinary social media use, where learners may encounter English without intending to study it.

Shanthi et al. (2025), for example, examined the use of TikTok videos for vocabulary acquisition among university students in Malaysia. Their mixed-methods study found generally positive perceptions of TikTok as an engaging way to learn vocabulary, and students also reported better retention of newly encountered English words. The authors therefore suggested that TikTok could be used as a supplementary resource in formal language learning.

This type of use is not the main focus of the present paper. A learner who deliberately follows English-learning accounts or searches for vocabulary videos already has a learning purpose. The experience is different when the same learner hears an unfamiliar word in a comedy clip, lifestyle video, review, or other content that was not created to teach English. Their attention may be on the message or entertainment rather than the language. If they pick up vocabulary during that experience, the learning is incidental.

Sely and Priyatmojo (2026) consider this more informal type of TikTok use. They discuss how learners may encounter English vocabulary through ordinary non-educational content rather than planned lessons. Their analysis draws attention to multimodal input, repeated exposure, code-switching, and interaction in comment sections. Each of these features can create additional contact with words while users are engaged in everyday activity on the platform.

Their work helps explain why TikTok may provide opportunities for incidental vocabulary learning, but it leaves another question open. What happens after a learner first encounters an unfamiliar word? A word may be noticed and understood from the surrounding context, appear again in later content, become more familiar, and eventually be remembered. Other words may disappear after the first encounter. Sely and Priyatmojo (2026) identify several conditions that may support incidental learning, but they do not organise these conditions into a process showing how vocabulary knowledge may develop over time.

TikTok's multimodal environment makes this question particularly relevant. A user may first hear an unfamiliar expression in a video and then see the same expression in subtitles or on-screen text. Images, actions, or gestures may help the user infer what it means. The expression may appear again in another video, while comments may show how other users understand or use it. None of these encounters means that the word has necessarily been learned. They do, however, show how different forms of input and later encounters may gradually contribute to stronger vocabulary knowledge.

Research from several areas helps explain parts of this process. Audiovisual studies show that incidental vocabulary learning can occur while learners are mainly focused on meaning (Montero Perez, 2020; Peters & Webb, 2018). Repeated encounters have also been associated with stronger vocabulary learning (Uchiyama et al., 2019), while captioned viewing can provide written support alongside spoken input (Kurokawa et al.,

2025). Research on IDLE further shows that English learning opportunities can arise through everyday digital activities outside formal instruction (Lee, 2019; Lee & Dressman, 2018).

TikTok brings many of these conditions together in one environment. The Multimodal Incidental Vocabulary Learning Framework proposed in this paper uses these areas of research to consider what may happen between the first encounter with an unfamiliar word and the development of more stable vocabulary knowledge.

DISCUSSION

TikTok as a Multimodal Environment for Incidental Vocabulary Learning

TikTok gives users several ways to encounter English within the same piece of content. A short video may include spoken language, images, subtitles, on-screen text, captions, hashtags, comments, and replies. A learner may therefore hear a word, see its written form, and use what is happening in the video to help understand it. This is similar to other audiovisual settings where learners process spoken, written, and visual information while paying attention mainly to meaning. Research on second-language audiovisual learning shows that vocabulary can develop through this kind of exposure, although the outcome is not the same for every learner or every word (Montero Perez, 2020; Peters & Webb, 2018).

Written support may make some of these encounters easier to process. Peters et al. (2016) found that captions supported certain aspects of vocabulary learning, particularly written word-form knowledge, while Kurokawa et al. (2025) similarly reported an advantage for captioned over uncaptioned viewing. When learners can hear a word and see its written form simultaneously, the written representation may provide additional support for identifying the form and connecting it with meaning.

A vocabulary encounter on TikTok can also extend beyond one video. A word heard in the spoken content may appear again in the subtitles, the caption, or a comment below the video. If users continue watching material on the same topic or trend, they may meet the expression again later. Sely and Priyatmojo (2026) also point to repeated exposure, code-switching, non-educational content, and comment interaction as possible parts of incidental learning on TikTok. In this way, contact with a word can occur across different parts of the platform rather than in one isolated viewing.

What users actually do on TikTok can also change the kind of English they encounter. Zainuddin and Yunus (2022) show that social networking sites can support both formal and informal English learning, while Lai and Wang (2025a, 2025b) suggest that the purposes behind informal digital media use may be related to different aspects of vocabulary knowledge.

This is useful when considering TikTok because users do not all use the platform in the same way. Two people may spend the same amount of time scrolling, but one may regularly watch English-language content, read captions and comments, replay videos, or interact with posts, while another may do very little of this. Their opportunities to encounter and process English vocabulary may therefore be quite different.

Although TikTok provides multiple opportunities for incidental vocabulary exposure, these encounters do not automatically result in vocabulary learning. Their contribution depends on whether unfamiliar words are noticed, meaning is constructed, and emerging knowledge is subsequently retained.

Proposed Multimodal Incidental Vocabulary Learning Framework

The research discussed so far points to several processes that may be involved when vocabulary is encountered incidentally through TikTok. These include exposure to multimodal input, noticing, understanding meaning from context, meeting the same word again, retaining what has been learned, and possibly recognising or using the word later.

To organise these processes, this paper proposes a Multimodal Incidental Vocabulary Learning Framework for considering how an unfamiliar word encountered during everyday TikTok use may gradually develop into more stable vocabulary knowledge. Figure 1 shows how these processes are connected.

Figure 1 Proposed Multimodal Incidental Vocabulary Learning Framework Through TikTok

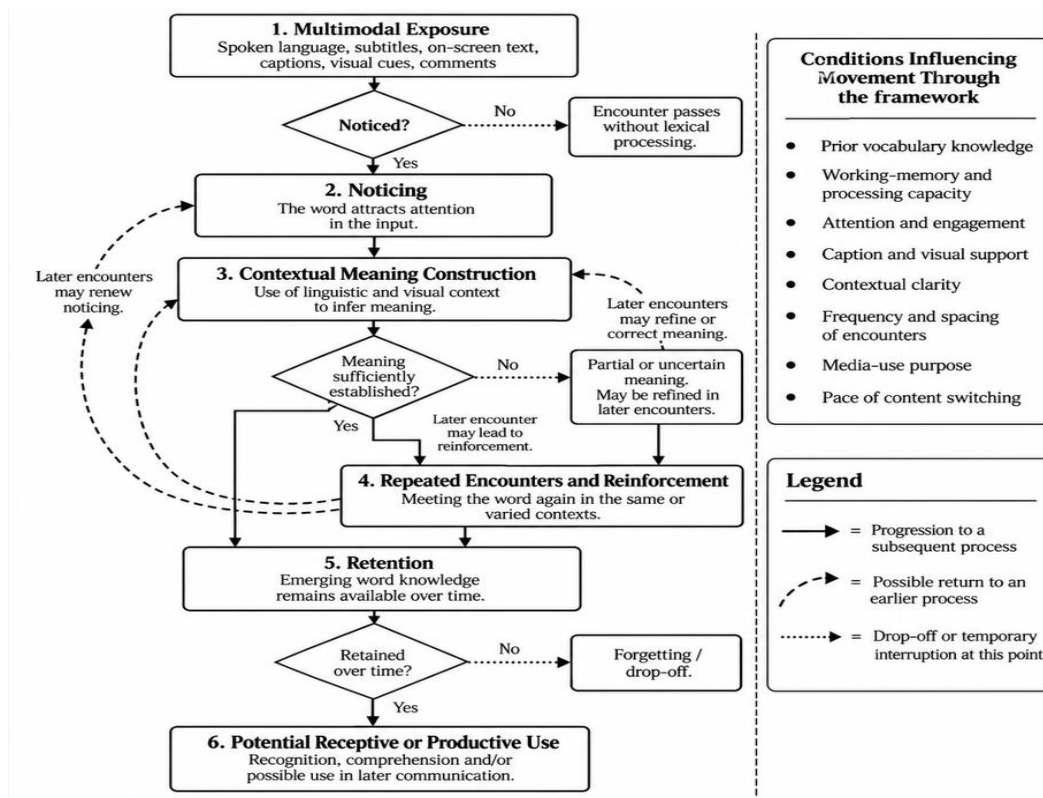


Figure 1 presents the proposed Multimodal Incidental Vocabulary Learning Framework as a flexible process rather than a fixed sequence. It begins with **multimodal exposure**, where learners may encounter an unfamiliar English word through spoken language, subtitles, on-screen text, captions, visual cues, or comments during everyday TikTok use. Exposure, however, does not mean that the word will necessarily be processed. The word must first attract the learner's attention before further processing can occur. This distinction reflects Schmidt's (1990) account of noticing, which places attention to linguistic information as an important part of subsequent processing. In a multimodal environment, the availability of spoken, written, and visual information may provide several possible sources of support, although learners cannot attend equally to all of them because processing capacity is limited (Mayer, 2021).

Once a word is noticed, the learner may begin contextual meaning construction by using the information available around it. Speech, subtitles, images, actions, captions, or other contextual cues may help the learner form an initial understanding of the word. This meaning does not have to be complete or fully accurate. Vocabulary knowledge often develops gradually, and different aspects of a word may become established at different rates (Schmitt, 2008; Webb, 2007). For this reason, the framework allows a word to remain at the level of partial or uncertain meaning. A later encounter may then provide clearer contextual information and help the learner refine or correct the earlier interpretation.

Repeated encounters and reinforcement are therefore represented as a possible route rather than a compulsory stage. If an initial form–meaning connection is sufficiently strong, some lexical knowledge may remain available even without an immediate further encounter. At the same time, meeting the word again may strengthen that emerging knowledge. Webb (2007) found that repeated encounters contributed to gains across several aspects of vocabulary knowledge, while Uchihara et al. (2019) reported a positive relationship between encounter frequency and incidental vocabulary learning. Later encounters may also return the learner to earlier processes. The same word may attract renewed attention or appear in a clearer context that

changes the learner's understanding. The dashed feedback arrows in Figure 1 represent this possibility and show why vocabulary development should not be treated as strictly linear.

The next process is retention, which refers to whether emerging word knowledge remains available over time. Research on audiovisual learning provides evidence that vocabulary gained through meaning-focused exposure can sometimes be retained beyond the immediate viewing experience (Feng & Webb, 2020; Peters & Webb, 2018). Retention is still not guaranteed. Some knowledge may weaken or be forgotten, while other words may remain accessible and later become available for recognition, comprehension, or use. Nation (2013) distinguishes between receptive and productive dimensions of vocabulary knowledge, with productive use generally requiring learners to retrieve and use a word independently. The final process is therefore described as potential receptive or productive use rather than an outcome that every lexical encounter is expected to reach.

Movement through the framework may also be shaped by several learner- and input-related conditions. Prior vocabulary knowledge can influence how easily unfamiliar words are processed, while the frequency and spacing of encounters may affect the extent to which emerging knowledge is reinforced (Peters & Webb, 2018; Uchihara et al., 2019). Attention, engagement, contextual clarity, and access to captions or visual support may also affect what learners notice and understand. The purpose and pattern of digital media use are relevant as well, since IDLE research suggests that different forms and purposes of informal digital engagement may be associated with different vocabulary outcomes (Lai & Wang, 2025a, 2025b; Lee, 2019). The pace of short-form content switching may create an additional condition because attention has to be redirected rapidly from one video to another. Luo et al. (2025) did not examine vocabulary learning directly, but their findings on reduced cue-based preparation following short-form video exposure raise a relevant question about whether rapid switching may sometimes limit the time available for sustained lexical processing. These conditions are not separate stages of the framework; instead, they may support, limit, or interrupt what happens at different points in the development of lexical knowledge.

Incidental Exposure Does Not Necessarily Constitute Learning

Seeing English on TikTok does not mean that vocabulary learning has taken place. A learner may spend a great deal of time watching English-language content and still pay very little attention to unfamiliar words. Research on informal digital learning shows that outcomes depend on more than the time learners spend online. The range of activities they take part in and the reasons for using digital media may also affect what they learn (Lai & Wang, 2025a, 2025b; Lee, 2019). Time spent in an English-rich environment therefore does not necessarily translate into lasting vocabulary knowledge.

Research on social networking sites shows a similar pattern. Zainuddin and Yunus (2022), in their systematic review, found that social networking platforms have been used in both formal and informal learning and have been associated with language skills, interaction, motivation, and flexibility. Their review also shows that different platforms and learning contexts need to be considered separately. Social media does not provide the same kind of learning experience in every situation. On TikTok, for example, simply coming across English in a video says very little about whether a user actually notices a particular word, understands it, or remembers it later.

The short-form nature of TikTok may make this issue more complicated. Users often move quickly from one brief video to another, with each video presenting a different topic, visual style, language, or purpose. Attention is therefore constantly being redirected. Luo et al. (2025) found preliminary experimental evidence that exposure to TikTok-style short-form videos was associated with reduced cue-based preparation in a later task-switching activity. The study did not examine vocabulary learning, so its findings cannot be used to claim that TikTok reduces vocabulary acquisition. Nevertheless, the findings raise a useful question about whether rapid changes in short-form content may sometimes make sustained attention more difficult.

This issue may be particularly relevant during contextual meaning construction. To infer the meaning of an unfamiliar word, a learner may need to keep the lexical form in mind while connecting it with spoken language, subtitles, visual actions, or other contextual information. Mayer's (2021) limited-capacity

principle suggests that learners cannot process unlimited amounts of visual and auditory information at the same time. When one short video is quickly replaced by another, attention has to shift to a new set of linguistic and visual cues. This may interrupt the processing of the earlier word before a sufficiently stable form–meaning connection has developed. Rapid content switching may therefore place additional demands on working memory and processing capacity during contextual meaning construction

This matters because having many sources of information does not always mean that a learner will process a word deeply. A user may hear an unfamiliar word, see it in subtitles, and understand something about its meaning from the visual context, but move on before giving it much attention. TikTok may therefore provide many encounters with English while some of those encounters remain very brief. The platform can increase exposure, but the same rapid and changing content may also limit how much attention is given to individual words.

Learning can stop at several points in the framework. A word might be present in a video without attracting any attention. It might be noticed but remain unclear, or the learner may understand it briefly and then forget it because it never appears again. Even knowledge that begins to develop may fade over time. For this reason, exposure on TikTok should not be treated by itself as evidence that vocabulary has been learned.

The framework therefore does not assume a direct relationship such as:

TikTok use → vocabulary learning

TikTok provides the opportunity to encounter vocabulary. What the learner does with that encounter is another matter. Some words may be noticed and understood, while others are not. Later encounters may strengthen what has already been learned, but this will also depend on the learner and on the conditions in which the word appears.

Conceptual Contribution

Sely and Priyatmojo (2026) show how ordinary TikTok use can expose learners to vocabulary through non-educational content, code-switching, repeated encounters, and interaction in comments. Their work establishes TikTok as a possible setting for incidental vocabulary learning.

The framework proposed here focuses on the vocabulary encounter itself and what may follow it. Research has already examined audiovisual vocabulary learning, noticing, repetition, and informal digital learning separately. The present framework connects these areas by following a word from its first appearance through possible noticing, meaning construction, later encounters, retention, and eventual recognition or use.

This changes the question slightly. Instead of asking only whether TikTok can help learners gain vocabulary, the framework asks what happens to a particular word after it appears during ordinary use. One learner may notice it, make sense of it, meet it again, and remember it. Another may hear the same word without giving it much attention. These different experiences can help explain why similar amounts of TikTok use do not necessarily produce similar vocabulary outcomes.

The contribution, then, is not a new set of vocabulary-learning processes. It is the way those processes are connected to explain possible vocabulary development during everyday TikTok use. The framework also makes room for learning to stop at different points rather than assuming that every encounter ends in retention or use.

Implications for Future Research

Future research can test different parts of the framework separately. One possibility is to examine how the forms of input available on TikTok contribute to vocabulary learning. A learner may hear a word, read it in subtitles, see it on screen, or use visual information to understand it. Researchers could compare these different sources of information and examine what happens when several are available together. Since

captioned viewing has been associated with vocabulary learning (Kurokawa et al., 2025; Peters et al., 2016), studies could also test whether hearing and seeing the same word helps learners notice or understand it more easily.

Repeated encounters could be examined in a similar way. Uchiyama et al. (2019) found a positive relationship between frequency and incidental vocabulary learning, but repetition on TikTok happens under much less controlled conditions. Users may meet a word again because they watch another video on the same topic, follow the same creator, or continue engaging with a trend. Researchers could follow selected words across several videos and test what learners remember afterwards. The timing of the encounters and the contexts in which the word reappears could also be compared.

Another issue is the difference between what learners think they have learned and what they can actually demonstrate. Self-report data can show whether users feel that TikTok helps them learn new vocabulary, but this does not necessarily mean that the words have been retained. Future research could therefore include direct measures of vocabulary knowledge, such as recognition of word forms and meanings, recall, and productive use. These measures would make it possible to see whether a word that was first encountered incidentally remains available after the original viewing experience.

Potential Hypotheses for Empirical Testing

The relationships proposed in the framework can also be examined through specific testable hypotheses. For example, lexical attention could be measured through eye-tracking indicators such as fixation duration and fixation count, while vocabulary development could be assessed using immediate and delayed measures of form recognition, meaning recognition, recall, and productive use. These measures would allow different stages of the framework to be examined separately rather than treating vocabulary learning as a single outcome.

- **H1:** Learners who encounter unfamiliar target words through both spoken input and on-screen captions will achieve higher subsequent form-recognition scores than learners who encounter the words through spoken input alone. Within the captioned condition, longer fixation duration and more frequent fixations on target words will be associated with higher form-recognition scores.
- **H2:** Learners who encounter unfamiliar words with clear visual and linguistic contextual support will achieve higher meaning-recognition scores than learners who encounter the same words with limited contextual support.
- **H3:** Learners who encounter target words repeatedly across several short-form videos will achieve higher immediate and delayed recognition and recall scores than learners who encounter the words only once.
- **H4:** Learners exposed to rapid switching between short-form videos will achieve lower subsequent meaning-recognition and delayed-retention scores than learners viewing the same target-word content under conditions that allow more sustained processing.

How users interact with TikTok may also influence vocabulary outcomes. Lai and Wang (2025a, 2025b) show that different purposes for informal digital engagement may be related differently to receptive and productive vocabulary knowledge. Similar questions could be examined on TikTok. Researchers might compare the kinds of content users watch, whether they read comments or captions, whether they replay videos, and how much English appears in the content they regularly engage with. This may help explain why two users who spend a similar amount of time on TikTok can have very different opportunities to encounter and process English vocabulary.

The framework could also be examined on other short-form social media platforms. Instagram Reels and YouTube Shorts, for example, also combine brief audiovisual content with written text, social interaction, and recurring exposure. Future comparative research could examine whether the same processes of

exposure, noticing, meaning construction, repeated encounters, and retention appear across these platforms. This would help show which parts of the framework may apply more broadly to short-form social media and which are more closely related to the way TikTok is designed and used.

CONCLUSION

English appears regularly in the short-form digital content that people use in everyday life. On TikTok, the same word can be heard in speech, seen in subtitles or on-screen text, read again in a caption or comment, and possibly encountered later in another video. These different forms of contact create opportunities for incidental vocabulary learning, although simply being exposed to them does not mean that learning has occurred.

The Multimodal Incidental Vocabulary Learning Framework proposed in this paper describes six processes that may be involved: multimodal exposure, noticing, contextual meaning construction, repeated encounters and reinforcement, retention, and potential receptive or productive use. A word does not have to pass through all six. One may be noticed and then forgotten, while another may become clearer and more familiar after several encounters. What happens can differ according to the learner, the available contextual support, and the way later encounters occur.

The framework therefore focuses on the space between encountering a word and eventually knowing it. Some encounters may end almost immediately. Others may give learners enough information to form a meaning, strengthen that understanding when the word appears again, and remember it later. Looking at these points separately may give a clearer picture of how vocabulary develops during ordinary short-form social media use.

The proposed framework still needs to be examined empirically as a complete framework. Research could follow words that learners encounter naturally on TikTok and investigate what happens to them over time. This would make it possible to see which words are noticed, which are remembered, where learning tends to stop, and what conditions appear to support further development. The same approach could later be tested on other short-form social media platforms.

ACKNOWLEDGEMENT

The authors gratefully acknowledge the financial support provided by Akademi Pengajian Bahasa (APB), Universiti Teknologi MARA (UiTM) through the GIA Grant, which supported the publication of this paper.

REFERENCES

1. Feng, Y., & Webb, S. (2020). Learning vocabulary through reading, listening, and viewing: Which mode of input is most effective? *Studies in Second Language Acquisition*, 42(3), 499–523. <https://doi.org/10.1017/S0272263119000494>
2. Kang, H., & Lou, C. (2022). AI agency vs. human agency: Understanding human–AI interactions on TikTok and their implications for user engagement. *Journal of Computer-Mediated Communication*, 27(5), zmac014. <https://doi.org/10.1093/jcmc/zmac014>
3. Kurokawa, S., Hein, A. M., & Uchihara, T. (2025). Incidental vocabulary acquisition through captioned viewing: A meta-analysis. *Language Learning*, 75(4), 939–987. <https://doi.org/10.1111/lang.12697>
4. Lai, C., & Wang, Q. (2025a). Online informal learning of English and receptive vocabulary knowledge: Purpose matters. *ReCALL*, 37(1), 62–78. <https://doi.org/10.1017/S095834402400017X>
5. Lai, C., & Wang, Q. (2025b). Profiles of media use purpose in informal digital learning of English and their association with vocabulary knowledge. *Computer Assisted Language Learning*. Advance online publication. <https://doi.org/10.1080/09588221.2025.2501705>

6. Lee, J. S. (2019). Informal digital learning of English and second language vocabulary outcomes: Can quantity conquer quality? *British Journal of Educational Technology*, 50(2), 767–778. <https://doi.org/10.1111/bjet.12599>
7. Lee, J. S., & Dressman, M. (2018). When IDLE hands make an English workshop: Informal digital learning of English and language proficiency. *TESOL Quarterly*, 52(2), 435–445. <https://doi.org/10.1002/tesq.422>
8. Liu, G. L., Soyooof, A., Lee, J. S., & Zhang, L. J. (2025). Informal digital learning of English in Asian English as a foreign language contexts: A thematic review. *RELC Journal*. Advance online publication. <https://doi.org/10.1177/00336882251332309>
9. Luo, W., Zhao, X., Jiang, B., Fu, Q., & Zheng, J. (2025). Swiping disrupts switching: Preliminary evidence for reduced cue-based preparation following short-form video exposure. *Behavioral Sciences*, 15(8), 1070. <https://doi.org/10.3390/bs15081070>
10. Mayer, R. E. (2021). *Multimedia learning* (3rd ed.). Cambridge University Press. <https://doi.org/10.1017/9781316941355>
11. Montero Perez, M. (2020). Incidental vocabulary learning through viewing video: The role of vocabulary knowledge and working memory. *Studies in Second Language Acquisition*, 42(4), 749–773. <https://doi.org/10.1017/S0272263119000706>
12. Montero Perez, M., Peters, E., Clarebout, G., & Desmet, P. (2014). Effects of captioning on video comprehension and incidental vocabulary learning. *Language Learning & Technology*, 18(1), 118–141. <https://doi.org/10.64152/10125/44357>
13. Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139858656>
14. Peters, E., & Webb, S. (2018). Incidental vocabulary acquisition through viewing L2 television and factors that affect learning. *Studies in Second Language Acquisition*, 40(3), 551–577. <https://doi.org/10.1017/S0272263117000407>
15. Peters, E., Heynen, E., & Puimège, E. (2016). Learning vocabulary through audiovisual input: The differential effect of L1 subtitles and captions. *System*, 63, 134–148. <https://doi.org/10.1016/j.system.2016.10.002>
16. Schellewald, A. (2021). Communicative forms on TikTok: Perspectives from digital ethnography. *International Journal of Communication*, 15, 1437–1457. https://ijoc.org/index.php/ijoc/article/view/16414?utm_source=.com
17. Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158. <https://doi.org/10.1093/applin/11.2.129>
18. Schmitt, N. (2008). Review article: Instructed second language vocabulary learning. *Language Teaching Research*, 12(3), 329–363. <https://doi.org/10.1177/1362168808089921>
19. Sely, G. R. A., & Priyatmojo, A. S. (2026). Incidental English vocabulary learning through TikTok: A conceptual analysis. *Journal of English Language and Education*, 11(1), 661–668. <https://doi.org/10.31004/jele.v11i1.2064>
20. Shanthi, A., Paramasivam, S., Chong, G. L., & Teoh, J. T. (2025). TikTok videos as a catalyst for vocabulary acquisition in tertiary education. *International Journal of Social Science Research*, 13(1), 35–53. <https://doi.org/10.5296/ijssr.v13i1.22287>
21. Uchihara, T., Webb, S., & Yanagisawa, A. (2019). The effects of repetition on incidental vocabulary learning: A meta-analysis of correlational studies. *Language Learning*, 69(3), 559–599. <https://doi.org/10.1111/lang.12343>
22. Webb, S. (2007). The effects of repetition on vocabulary knowledge. *Applied Linguistics*, 28(1), 46–65. <https://doi.org/10.1093/applin/aml048>
23. Webb, S., Uchihara, T., & Yanagisawa, A. (2023). How effective is second language incidental vocabulary learning? A meta-analysis. *Language Teaching*, 56(2), 161–180. <https://doi.org/10.1017/S0261444822000507>
24. Zainuddin, F. N., & Yunus, M. M. (2022). Sustaining formal and informal English language learning through social networking sites (SNS): A systematic review (2018–2022). *Sustainability*, 14(17), 10852. <https://doi.org/10.3390/su141710852>