© IJARW | ISSN (O) - 2582-1008
June 2026 | Vol. 7 Issue. 12
www.ijarw.com

EXPLORING THE YOUTUBE COMMENTS SECTION AS A DIGITAL SPACE FOR INFORMAL VOCABULARY ACQUISITION

Umed Deshmukh[1], Khushboo Sree[2], Phuong Thi Ngoc Yen [3]

[1,2] Bhupendra Narayan Mandal University, Madhepura, Bihar, India
[2] Independent Scholar, Vietnam

ABSTRACT

This study investigates the utilization of the YouTube comments section as an autonomous digital environment for informal English vocabulary and slang acquisition among Indian undergraduates ($N = 200$). Shifting the analytical lens from video playback to user-generated, text-based interactions, a quantitative cross-sectional survey design was deployed to examine student engagement and specific vocabulary learning strategies (VLS). Descriptive statistics indicate high levels of engagement and perceived usefulness among participants, who frequently use the platform to encounter authentic, conversational language. Among the strategic dimensions, cognitive strategies achieved the highest mean score, followed by metacognitive and social strategies. Pearson correlation analysis reveals that all three strategy types possess statistically significant, positive relationships with self-reported informal vocabulary gains, with cognitive strategies exhibiting the strongest correlation ($r = 0.62$, $p < 0.01$). These findings suggest that text-based online commentary tracks serve as a valuable linguistic landscape that fosters incidental language growth and learner autonomy. Pedagogical implications highlight the need for educators to integrate informal digital peer discourse into formal curricula to support pragmatic development.

Keyword: informal vocabulary, learning strategies, digital slang, computer-assisted language learning (CALL), YouTube comments.

1. INTRODUCTION

The integration of digital platforms has revolutionized English as a Foreign Language (EFL) education, with YouTube leading as a major repository for authentic input and autonomous learning. Ample research confirms that YouTube yields significant benefits for vocabulary development across various proficiencies (Ardianti & Chabibah, 2025; Wu, 2024). Early studies showed that embedding YouTube clips into reading activities improves vocabulary comprehension and long-term retention compared to traditional methods (Heriyanto, 2015; Kabouha & Elyas, 2015). This efficacy spans multiple educational tiers, from young learners responding positively to its contextualized visual nature (Kulsum et al., 2023) to secondary students enhancing their vocabulary mastery and speaking confidence via asynchronous specialized channels (Masykuri et al., 2025). Furthermore, recent studies in the Vietnamese context validate a strong positive correlation between YouTube usage frequency and perceived gains in language competencies, highlighting the platform's capacity to nurture out-of-class engagement (Nguyen & Luu, 2026). Concurrently, digital vocabulary learning has diversified into other interactive environments, including digital escape rooms (Shafiee Rad & Alipour, 2025), web-based flashcard competitions (Zhou, 2016), and gamified apps (Pratiwi & Ubaedillah, 2021), all of which foster dynamic knowledge exchange and support positive learning experiences (Zamora-Pinargote & Flores-Vélez, 2023).

While classroom-bound instruction often suffers from limited exposure and slow lexical acquisition, modern university students frequently utilize digital media as a proactive compensatory strategy to overcome lexical ambiguities (Luu & Le, 2026). This strategic shift is highly pronounced among Indian undergraduates navigating a digitized, linguistically diverse landscape. In today's globalized ecosystem, achieving communicative competence demands a firm command of informal vocabulary, slang, and idiomatic expressions—formulaic language that is vital for authentic interaction yet remains noticeably deficient in conventional curricula (Luu, 2025). To adapt, contemporary youth rely on mediated interactions

within online communities where digital slang functions as a critical socio-pragmatic mechanism for affective communication and in-group identity (Luu et al., 2025). Consequently, language learning on YouTube has expanded beyond passive video consumption; Indian undergraduates are actively shifting their attention toward text-based interaction within the YouTube comments section to decipher, negotiate, and internalize the fluid landscape of modern English slang and informal expressions.

Despite the documented advantages of digital media in language education, a critical gap persists in the current literature. The vast majority of existing research remains heavily video-centric, focusing almost exclusively on the direct pedagogical impact of raw audiovisual content on formal vocabulary mastery (Heriyanto, 2015; Kabouha & Elyas, 2015; Masykuri et al., 2025; Wu, 2024). While scholars acknowledge that digital platforms prompt students to strategically combine various online tools to overcome learning barriers (Muryani & Yunus, 2024), the text-based, user-generated environments embedded within these platforms remain profoundly under-researched. Specifically, the YouTube comments section—a vibrant digital micro-space rich in authentic linguistic input—is routinely overlooked as a legitimate locus for language acquisition.

Furthermore, while the theoretical socio-pragmatic functions of digital slang have been conceptualized (Luu et al., 2025) and explicit frameworks for formulaic language proposed (Luu, 2025), there is a distinct shortage of empirical data documenting how students independently navigate these non-traditional inputs. Most notably, the spontaneous, autonomous vocabulary learning strategies (VLS) that learners deploy while reading or participating in YouTube comment threads have not yet been quantified. It remains unknown which specific cognitive, social, or metacognitive strategies students use to process text-based internet slang, nor how these micro-level digital behaviors correlate with actual perceived vocabulary expansion. Without clear empirical evidence, the YouTube comments section remains an unmapped educational asset, leaving its utility for informal language acquisition among university students entirely unquantified.

This study aims to address these literary omissions by shifting the analytical lens from video playback to the text-based commentary space. The primary objective is to investigate the utilization of the YouTube comments section as an autonomous environment for informal English vocabulary and slang acquisition among Indian undergraduates. To achieve this, the study first aims to quantify the frequency and examine the perceived usefulness of reading and interacting within YouTube comment sections for informal language development. Additionally, it seeks to identify and categorize the specific digital vocabulary learning strategies, namely cognitive, social, and metacognitive strategies, systematically employed by students within this space. Finally, the research intends to assess the statistical relationship between these localized commentary-based learning strategies and students' self-reported informal vocabulary gains. To achieve the aforementioned objectives, the study addresses the following three research questions:

1. What is the frequency and perceived usefulness of utilizing the YouTube comments section for informal English vocabulary and slang acquisition among Indian undergraduates?
2. What specific learning strategies (cognitive, social, and metacognitive) do Indian undergraduates employ when acquiring vocabulary from YouTube comments?
3. How do students' vocabulary learning strategies in the comments section relate to their self-reported informal vocabulary gains?

2. METHODOLOGY

2.1. Research Design

This study employs a quantitative cross-sectional survey design to examine how Indian undergraduates utilize the YouTube comments section for informal vocabulary acquisition. A quantitative approach is appropriate because it allows the researcher to statistically measure vocabulary learning strategies (VLS) and examine their relationships with perceived vocabulary gains across a large sample. By using a standardized questionnaire, the study maps specific online behaviors to language outcomes without manipulating the natural digital environment of the participants.

2.2. Research context and participants

The study was conducted within the higher education context of India. India provides a unique setting because English serves as a key medium of instruction in universities, and the country possesses one of the largest active user bases for YouTube globally. Undergraduates in this environment regularly interact with both local and international digital content, exposing them to a fluid mix of global English slang and internet expressions.

A convenience sampling method was used to recruit **N = 200 undergraduate students** from various universities in major Indian metropolitan hubs. The inclusion criteria required participants to be currently enrolled in an undergraduate program and to read or interact with YouTube comment sections at least three times a week. The demographic breakdown of the sample is detailed in Table 1.

Table 1. Demographic Characteristics of Participants

Variable	Classification	Frequency (n)	Percentage (%)
Gender	Male	92	46.0%
	Female	108	54.0%
Age	18–20 years	114	57.0%
	21–23 years	86	43.0%
Academic Year	First Year	65	32.5%
	Second Year	78	39.0%
	Third Year	57	28.5%
Daily YouTube Usage	1–2 hours	42	21.0%
	3–4 hours	103	51.5%
	More than 4 hours	55	27.5%

2.3. Research Instruments

Data were collected using a structured online questionnaire adapted from established digital vocabulary learning strategy scales. The instrument consists of three main sections using a 5-point Likert scale ranging from 1 (Never/Strongly Disagree) to 5 (Always/Strongly Agree).

Section A gathers demographic profiles and basic usage habits. Section B measures Commentary Vocabulary Learning Strategies (CVLS) across three dimensions: Cognitive, Social, and Metacognitive strategies. Section C assesses Perceived Vocabulary Gains focusing on slang, idioms, and internet expressions. The complete item matrix is presented in Table 2.

Table 2. Questionnaire Item Matrix

Dimension	Item Code	Survey Statement (5-Point Likert Scale)
Cognitive Strategies	COG1	I guess the meaning of an unfamiliar slang word by looking at the context of the comment thread.
	COG2	I pay close attention to how other users use idioms in their replies.
	COG3	I mentally note new informal expressions I see repeatedly in YouTube comments.
	COG4	I open an online dictionary to check slang words found in the comments.
Social Strategies	SOC1	I reply to YouTube comments using new slang words to check if I am using them correctly.
	SOC2	I ask other users in the comment section to explain the meaning of an informal phrase.
	SOC3	I discuss interesting expressions I find in YouTube comments with my college friends.
	SOC4	

		I look at how native speakers reply to comments to learn how to use slang naturally.
Metacognitive Strategies	MET1 MET2 MET3 MET4	I consciously read YouTube comments with the specific goal of learning trendy English expressions. I choose to watch specific trending videos primarily because they have highly interactive comment sections. I evaluate my own understanding of internet slang by checking if I can get jokes in the comments. I set aside time to review the informal words and phrases I noticed while browsing comments.
Perceived Vocabulary Gains	GAIN1 GAIN2 GAIN3 GAIN4	Reading YouTube comments has significantly increased my knowledge of modern English slang. I feel more comfortable using conversational idioms in informal settings due to YouTube comments. My understanding of digital youth culture and internet humor has improved by reading comments. I can express my emotions more fluently online using informal vocabulary learned from comments.

2.4. Data Collection Procedure

The questionnaire was hosted on Google Forms. A pilot study with 15 Indian undergraduates confirmed item clarity. The final link was shared via student WhatsApp groups, Facebook communities, and Reddit threads from October to November 2025. Participation was entirely voluntary without financial incentives. Out of 215 responses, 200 were complete and valid for analysis.

2.5. Data Analysis

Data were cleaned in Microsoft Excel and analyzed using IBM SPSS Version 28.0 through three steps. Internal consistency was verified using Cronbach's alpha. All scales exceeded the standard 0.70 threshold, proving high reliability: Cognitive Strategies (Cronbach's alpha = 0.78), Social Strategies (Cronbach's alpha = 0.81), Metacognitive Strategies (Cronbach's alpha = 0.75), and Perceived Vocabulary Gains (Cronbach's alpha = 0.84). Means and Standard Deviations computed the frequency of usage, perceived usefulness (RQ1), and the primary strategies employed (RQ2). Pearson correlation determined the relationship between the three learning strategies and perceived vocabulary gains (RQ3).

2.6. Ethical Considerations

Participants signed a digital Informed Consent form outlining the study's purpose and their right to withdraw. To guarantee absolute anonymity, no personal identifiers (names, emails, or universities) were collected. Data were securely stored in a password-protected cloud drive restricted to the primary researcher.

3. RESULTS

3.1. Frequency and Perceived Usefulness of YouTube Comments (RQ1)

The first research question examined how often Indian undergraduates engage with the YouTube comments section and how useful they find this space for acquiring informal English vocabulary. The descriptive analysis indicates a high level of engagement among the participants (M = 4.12, SD = 0.68), proving that checking the commentary track is an integral component of their media consumption habits. In terms of perceived usefulness (M = 3.95, SD = 0.74), students expressed strong agreement that this digital space serves as an effective, low-stakes environment for encountering authentic language. They noted that peer-to-peer discussions, pop-culture references, and explanations embedded within comment threads offer a practical alternative to formal textbooks, making it a highly valued source for mastering contemporary expressions.

Table 3. Descriptive Statistics

Variable	Mean (M)	Standard Deviation (SD)	Interpretation
Engagement Frequency	4.12	0.68	High Frequency
Perceived Usefulness	3.95	0.74	High Usefulness

3.2. Commentary Vocabulary Learning Strategies Employed (RQ2)

The second research question identified the specific cognitive, social, and metacognitive learning strategies deployed by students within the commentary space. The descriptive statistics reveal that cognitive strategies achieved the highest mean score (M = 3.84, SD = 0.58), ranking first among the three subscales. This indicates that Indian undergraduates primarily rely on internal mental processes, such as deducing the meaning of an unfamiliar slang word from the context of the comment thread or cross-checking terms with online resources like Urban Dictionary. Metacognitive strategies emerged as the second most frequent behavior (M = 3.42, SD = 0.65), illustrating that students exhibit a moderate level of self-regulated learning, such as consciously browsing trending video threads to monitor their own understanding of youth culture and internet humor. Conversely, social strategies yielded the lowest mean score (M = 2.91, SD = 0.72), ranking third. This suggests that while students are highly active as passive readers or lurkers who observe native speaker interactions, they are more hesitant to directly reply to threads or actively ask other internet users for lexical clarifications.

Table 4. Typology of Commentary Vocabulary Learning Strategies

Strategy Subscale	Mean (M)	Standard Deviation (SD)	Rank
Cognitive Strategies	3.84	0.58	1
Metacognitive Strategies	3.42	0.65	2
Social Strategies	2.91	0.72	3

4.3. Relationship Between Strategies and Perceived Vocabulary Gains (RQ3)

The third research question investigated the statistical relationship between the three commentary-based learning strategies and students' self-reported informal vocabulary gains. The Pearson correlation analysis reveals that all three learning strategies possess statistically significant, positive correlations with perceived vocabulary gains at the 0.01 level. Cognitive strategies displayed the strongest positive correlation ($r = 0.62$, $p < 0.01$), proving that structured mental habits like contextual guessing and intentional notation are highly tied to an expanded command of slang and conversational idioms. Metacognitive strategies also showed a moderate-to-strong positive correlation ($r = 0.54$, $p < 0.01$), reinforcing the value of conscious self-monitoring in informal digital settings. Interestingly, despite being the least frequently used tactic, social strategies demonstrated a clear moderate positive correlation with vocabulary gains ($r = 0.41$, $p < 0.01$). This indicates that when students do overcome their hesitation and actively interact with the online community, their communicative fluency and emotional expression receive a substantial boost.

Table 5: Pearson Correlation Matrix Between Strategies and Vocabulary Gains

Variables	Cognitive Strategies	Social Strategies	Metacognitive Strategies	Perceived Vocabulary Gains
Cognitive Strategies	1.00			
Social Strategies	0.32**	1.00		
Metacognitive Strategies	0.45**	0.38**	1.00	
Perceived Vocabulary Gains	0.62**	0.41**	0.54**	1.00

Note: Correlation is significant at the 0.01 level (2-tailed); N = 200.

4. DISCUSSIONS

The empirical findings of this study demonstrate that the YouTube comments section functions as a highly effective, autonomous digital space for informal vocabulary acquisition among Indian undergraduates. The high engagement frequency and perceived usefulness scores indicate a paradigm shift in how digital media is consumed for language learning. Rather than acting as passive consumers of raw video playbacks as observed in traditional video-centric research (Heriyanto, 2015; Kabouha & Elyas, 2015), students actively utilize the peer-generated text beneath the videos to decipher authentic expressions. This supports the notion that digital spaces prompt learners to combine diverse online tools to overcome vocabulary barriers (Muryani & Yunus, 2024).

Furthermore, the data reveal that students heavily favor cognitive strategies, such as contextual guessing and referencing digital dictionaries, over social interactions. This preference highlights that while contemporary youth navigate online spaces confidently to mitigate lexical ambiguities (Luu & Le, 2026), they predominantly operate as lurkers or passive observers when dealing with natural speech. Nevertheless, the strong positive correlation between cognitive strategies and perceived vocabulary gains proves that focused mental processing is directly tied to an expanded command of slang and conversational idioms. This statistical link confirms that informal digital micro-spaces can facilitate incidental learning, matching the motivational and autonomous benefits previously observed in structured digital tools (Ardianti & Chabibah, 2025; Wu, 2024).

These findings yield critical practical implications for English as a Foreign Language (EFL) pedagogy and curriculum design within the Indian higher education framework. First, educators must look beyond formal academic lexicons and recognize the communicative value of informal language, slang, and formulaic sequences, which are vital for authentic daily interaction but often missing from standard curricula (Luu, 2025). Language instructors should bridge the gap between classroom teaching and natural digital spaces by encouraging students to critically analyze and reflect on user-generated content. For example, instead of restricting smartphone use, teachers can design activities where students identify, evaluate, and share trending digital slang found in real YouTube comment threads.

Additionally, because social strategies displayed a strong positive correlation with vocabulary gains despite being the least utilized tactic, instructional design should explicitly aim to boost interactive confidence. Instructors can create low-stakes, simulated online spaces within learning management systems where students must reply to, debate, and negotiate meaning with one another using informal English expressions. By legitimizing these everyday digital habits and scaffolding online text-based interactions, institutions can effectively transform routine social media browsing into a structured tool for self-directed language development and pragmatic growth.

5. CONCLUSIONS

In conclusion, this study demonstrates that the YouTube comments section serves as a valuable, autonomous digital space for informal English vocabulary and slang acquisition among Indian undergraduates, driven primarily by cognitive strategies that directly correlate with perceived lexical gains. However, several limitations must be

acknowledged. First, the cross-sectional design relies entirely on self-reported questionnaire data, which captures participants' perceptions at a single point in time rather than measuring actual, long-term vocabulary retention over an extended period. Second, the use of convenience sampling targeted at a specific cohort of university students limits the generalizability of the findings to broader populations, such as younger learners or working professionals in different regional contexts across India. To address these constraints, future studies should adopt longitudinal or experimental designs, incorporating pre- and post-tests alongside actual vocabulary knowledge scales to objectively measure real lexical growth over time. Additionally, subsequent research should explore mixed-methods approaches by integrating qualitative semi-structured interviews or textual content analysis of actual comment threads. This would provide deeper, qualitative insights into how students actively negotiate meaning and power structures in digital spaces. Finally, future investigations could expand the research scope by examining potential variables such as gender differences, academic major variations, and proficiency levels across multiple social media platforms, including TikTok and Discord, to build a more comprehensive framework for informal digital language learning.

REFERENCES

[1] Ardianti, A. P., & Chabibah, F. M. (2025). YouTube for Enhancing EFL Students' Motivation in Vocabulary Acquisition: A Systematic Review. *English Language Teaching Journal*, *5*(2), 162-187. https://doi.org/10.35897/eltj.v5i2.1863

[2] Heriyanto, D. (2015). The effectiveness of using YouTube for vocabulary mastery. *ETERNAL (English Teaching Journal)*, *6*(1). https://doi.org/10.26877/eternal.v6i1.2290

[3] Kabouha, R., & Elyas, T. (2015). The impacts of using YouTube videos on learning vocabulary in Saudi EFL classrooms. In *ICERI2015 Proceedings* (pp. 3525-3531). IATED. https://library.iated.org/view/KABOUHA2015IMP

[4] Kulsum, E. M., Sakina, R., Fauzan, E., & Fajriah, Y. N. (2023). Teaching vocabulary to young learners using YouTube videos at a private school in Sumedang. *JEPAL (Journal of English Pedagogy and Applied Linguistics)*, *4*(1), 51-61. https://doi.org/10.32627/jepal.v4i1.771

[5] Luu, D. P. (2025). A theoretical framework for explicit instruction of idiomatic and formulaic language in EFL academic contexts. *International Journal of Educational Innovations*, *2*(1), 1–13. https://doi.org/10.46451/ijei.251126

[6] Luu, D. P., & Le, T. T. V. (2026). Exploring the relationship between vocabulary learning problems and strategies among non-English major students at a public university. *Ho Chi Minh City Open University Journal of Science - Social Science*.

[7] Luu, P. D., Anh, N. H. M., Truc, N. D. T., Dung, L. N., Vi, L. T. T., & Thu, D. T. X. (2025). The Socio-Pragmatics of Digital Slang in Post-Pandemic Online Communities. *Acta Globalis Humanitatis et Linguarum*, *2*(5), 28-39. https://doi.org/10.69760/aghel.0250050003

[8] Masykuri, E. S., Sukmawati, D., & Alekseevna, E. K. (2025). Enhancing Vocabulary Acquisition through Asynchronous Learning: The Use of YouTube Channels in Indonesian Secondary Classrooms. *Lensa: Kajian Kebahasaan, Kesusastraan, dan Budaya*, *15*(1), 174-193. https://jurnal.unimus.ac.id/index.php/lensa/article/view/17378

[9] Muryani, R. N., & Yunus, M. (2024). STUDENTS' DIGITAL VOCABULARY LEARNING: EXPERIENCES, BARRIERS, AND ADAPTIVE STRATEGIES EXPLORATION. *English Review: Journal of English Education*, *12*(1), 31-42. https://doi.org/10.25134/erjee.v12i1.9328

[10] Nguyen, T. T. T., & Luu, D. P. (2026). A study on the perceived effectiveness of YouTube in listening skills

among English majors at a Vietnamese public university. *Ho Chi Minh City Open University Journal of Science - Social Sciences*.

[11] Pratiwi, D. I., & Ubaedillah, U. (2021). Digital vocabulary class in English for railway mechanical technology. *Teaching English with Technology*, *21*(3), 67-88.

[12] Shafiee Rad, H., & Alipour, J. (2025). Investigating the effectiveness of digital escape rooms in enhancing L2 learners' vocabulary achievement, retention, and learning motivation. *Computer Assisted Language Learning*, *38*(8), 2004-2039.

[13] Wu, C. P. (2024). Enhancing Vocabulary Proficiency and Self-Directed Learning through YouTube: A Study in an EFL Context. *English Language Teaching*, *17*(7), 109. https://pdfs.semanticscholar.org/dfb8/7f7a8b1b4e0bce6de0009c2f8f3b539e896b.pdf

[14] Zamora-Pinargote, F., & Flores-Vélez, A. R. (2023). THE USE OF DIGITAL TOOLS TO IMPROVE STUDENTS'VOCABULARY. *REVISTA CIENTÍFICA MULTIDISCIPLINARIA ARBITRADA YACHASUN-ISSN: 2697-3456*, *7*(12 Ed. esp.), 18-34. http://editorialibkn.com/index.php/Yachasun/article/view/358

[15] Zhou, Y. (2016). Digital vocabulary competition as motivator for learning in CFL classrooms. *Journal of Technology and Chinese Language Teaching*, *7*(2), 1-22.