



BEYOND RECOGNITION: TOWARD A TASK-BASED FRAMEWORK FOR DEVELOPING PRODUCTIVE VOCABULARY IN MOBILE- ASSISTED LANGUAGE LEARNING

Muborak Azimova Nodir qizi

Uzbekistan State World Languages University, Master student, Tashkent,

Uzbekistan E-mail: azimovamuborak66@gmail.com

ORCID: 0009-0001-8706-6250

Abstract

Mobile-assisted language learning (MALL) has become one of the most extensively researched areas in vocabulary pedagogy, with recent systematic reviews and meta-analyses converging on a consistent conclusion: mobile tools reliably support vocabulary gains. Yet a closer reading of this literature reveals a persistent asymmetry. Most MALL research, and most MALL tools themselves, are built around tasks that develop receptive knowledge — recognizing a word's form and meaning — while productive knowledge, the ability to retrieve and correctly use a word in speech or writing, develops more slowly and is measured far less often. This asymmetry is particularly consequential in English for Specific Purposes (ESP) contexts, where the communicative goal of instruction — writing a professional email, negotiating a deal, delivering a presentation — is inherently productive. This paper synthesizes findings from recent systematic reviews, meta-analyses, and empirical studies on mobile-assisted vocabulary learning to examine why this receptive bias persists, and proposes a conceptual Recognition-to-Production (R2P) framework for auditing and redesigning MALL vocabulary tasks. The framework organizes mobile vocabulary tasks into three tiers — recognition, controlled production, and free production — and maps common mobile affordances onto each tier. Pedagogical implications are discussed with particular attention to ESP



instruction in higher education, and directions for empirical validation of the framework are outlined.

Keywords: mobile-assisted language learning; productive vocabulary knowledge; receptive vocabulary knowledge; task-based language teaching; English for Specific Purposes; vocabulary pedagogy

1. Introduction

Vocabulary occupies a peculiar position in language pedagogy: everyone agrees it matters, yet it is rarely taught with the same systematic attention given to grammar or the four skills. Mobile-assisted language learning (MALL) has partly filled this gap. Over the past decade, mobile vocabulary tools — from simple flashcard applications to gamified platforms and AI-driven chatbots — have generated a substantial and increasingly well-synthesized body of research. A 2026 systematic review of 36 empirical studies conducted between 2019 and 2025 confirmed that mobile-assisted vocabulary learning (MAVL) has become one of the most active strands within MALL research, with game-based and self-directed applications receiving particular attention (Duan et al., 2026). Earlier meta-analyses reached similarly positive conclusions about the overall effectiveness of mobile tools for vocabulary gains (Burston, 2015; Mahdi, 2018).

What these reviews also converge on, however, is a quieter and more consequential finding: the vocabulary knowledge that MALL tools typically build is disproportionately receptive rather than productive. Nation's (2001, 2013) influential framework distinguishes receptive knowledge — recognizing a word's form and retrieving its meaning while listening or reading — from productive knowledge, which requires retrieving and correctly producing a word's spoken or written form to express an intended meaning. The two are related but develop asynchronously, and the gap between them can be substantial (Laufer & Nation, 1995). Several recent studies confirm that mobile-assisted interventions tend to widen rather than close this gap: comparative work with Chinese EFL learners found



that mobile applications produced stronger gains in receptive than in productive vocabulary knowledge (Li & Hafner, 2022), and a mixed-methods study of a mobile vocabulary application similarly found that its input-centric design left learners better prepared to recognize words than to use them in speaking and writing (Xu et al., 2025). A systematic review and meta-analysis of ESL/EFL mobile vocabulary learning reached a comparable conclusion at the level of the field as a whole, noting that productive vocabulary knowledge remained comparatively neglected relative to receptive gains (Lin & Lin, 2019; Mahdi, 2018).

This asymmetry is not a minor methodological footnote. For general English learners, receptive vocabulary growth is valuable in its own right. But for learners of English for Specific Purposes (ESP) — including the business, economics, and technical students who make up a large share of MALL research populations — the communicative demands of the target context are overwhelmingly productive: drafting a report, negotiating terms, presenting findings, or writing a professional email all require learners to retrieve and deploy vocabulary actively, not merely recognize it on a screen. If MALL tools and the curricula built around them systematically under-develop productive knowledge, they may be solving only half of the vocabulary problem that ESP instruction exists to address.

This paper does not attempt another empirical comparison of app effectiveness; the meta-analytic case for MALL's overall effectiveness is already well established (Burston, 2015; Duan et al., 2026; Mahdi, 2018; Okumuş Dağdeler, 2023). Instead, it takes a conceptual approach: synthesizing the receptive-productive asymmetry documented across this literature, examining the design and assessment mechanisms that appear to sustain it, and proposing a framework — the Recognition-to-Production (R2P) continuum — for evaluating and redesigning MALL vocabulary tasks so that they deliberately progress learners toward productive use. The framework is developed with ESP instruction as its primary



illustrative context, though its logic extends to general English teaching wherever productive vocabulary use is a genuine instructional goal.

2. Literature Review

2.1 Vocabulary Knowledge as a Multi-Dimensional Construct

Vocabulary knowledge is often treated informally as a single quantity — the number of words a learner "knows" — but foundational work in second language vocabulary studies has long argued against this simplification. Nation's (2001, 2013) framework decomposes word knowledge into three components (form, meaning, and use), each with both a receptive and a productive dimension, yielding a multi-layered picture of what "knowing" a word actually involves. Receptive knowledge allows a learner to recognize a word's spoken or written form and retrieve its meaning; productive knowledge requires the reverse process — retrieving the appropriate form to express a meaning the learner already has in mind. Laufer and Nation's (1995) work on lexical richness in written production offered one of the first systematic ways to measure this productive dimension directly, rather than inferring it from receptive test scores, and demonstrated that the two dimensions do not move in lockstep.

This distinction matters pedagogically because most classroom and app-based vocabulary assessment defaults to receptive measures: multiple-choice items, matching tasks, and translation-recognition formats are easy to design, administer, and auto-grade, particularly within a mobile interface. Productive measures — sentence writing, controlled production tasks, oral elicitation — are comparatively costly to design and score at scale. This paper argues that this asymmetry in what is easy to *measure* has shaped what is easy to *teach*, a pattern visible throughout the MALL literature reviewed below.

2.2 The Receptive Bias in Mobile-Assisted Vocabulary Learning

Multiple independent strands of evidence point to the same pattern. Mahdi's (2018) meta-analysis of mobile-assisted vocabulary learning found that the



substantial majority of studies in the field measured and reported receptive rather than productive outcomes. Lin and Lin's (2019) systematic review and meta-analysis of ESL/EFL mobile vocabulary learning reached a comparable conclusion, noting that productive vocabulary knowledge remained comparatively underexamined relative to receptive gains. More recent empirical work confirms that this is not simply a measurement artifact but a genuine developmental asymmetry: Li and Hafner (2022) compared mobile-based and paper-based word cards among Chinese university EFL learners and found that while both formats improved receptive knowledge of form-meaning connections, gains in productive knowledge of collocations were comparatively modest. A large mixed-methods study of a mobile vocabulary application similarly found that its predominantly input-centric task design — built around exposure to words in context rather than tasks requiring active use — left learners under-prepared for productive application in speaking and writing (Xu et al., 2025). Even interventions explicitly designed to target both dimensions have found asymmetric outcomes: research on mobile-assisted technical vocabulary instruction using digital flashcards found that productive gains, while significant, were consistently smaller than receptive gains within the same intervention (Koleini et al., 2024).

The most recent systematic synthesis of the field makes this point explicitly at the level of the literature as a whole rather than any single study. Reviewing 36 empirical studies published between 2019 and 2025, Duan et al. (2026) found that while mobile platforms reliably supported word retention and recognition, deeper aspects of vocabulary knowledge — including collocational and contextual, productive use — remained comparatively underexplored, and called for applications and research designs that move beyond surface-level gains. This is consistent with earlier observations that mobile-assisted vocabulary tools, whatever their format, tend to privilege the kinds of tasks that are cheapest to build and easiest



to auto-grade: flashcards, matching exercises, and multiple-choice recognition (Godwin-Jones, 2011; Stockwell, 2010).

2.3 Self-Regulation and Engagement: A Necessary but Insufficient Condition

A separate and substantial strand of MALL research has focused not on task type but on learner self-regulation — the capacity to plan, monitor, and sustain one's own vocabulary study, which mobile devices are well positioned to support through portability and just-in-time access. Studies applying this lens have found that mobile-assisted vocabulary interventions can meaningfully strengthen learners' self-regulatory capacity alongside vocabulary gains (Boroughani, Behshad, & Xodabande, 2023; Boroughani, Xodabande, & Karimpour, 2023), that mobile learning attitudes and self-regulation tend to improve together over the course of an intervention (Lei et al., 2022), and that self-directed use of mobile applications can support meaningful academic vocabulary growth outside formal instruction (Xodabande & Atai, 2020). Work incorporating gamification and social features has extended this line further, showing that such design elements can strengthen learners' persistence with mobile vocabulary study over time (Xu & Richardson, 2024).

This body of work is valuable, but it answers a different question than the one this paper is concerned with. Self-regulation research asks whether learners engage with vocabulary study consistently; it does not, by itself, ask what kind of cognitive task learners are engaging with once they open the app. A highly self-regulated learner who completes daily flashcard sessions with excellent consistency is still, on the evidence reviewed above, likely to be building receptive rather than productive knowledge — because the task itself, however consistently performed, remains a recognition task. Autonomy over when and how often to study is not the same as exposure to the kind of task that builds productive competence. Closing the



receptive-productive gap therefore requires attention to task design specifically, not only to engagement and consistency.

2.4 Task-Based Language Teaching as a Design Resource

Task-based language teaching (TBLT) offers a well-established pedagogical logic for addressing this gap, built on the principle that language is best learned through meaning-focused tasks that require learners to use the target language to achieve a communicative outcome, rather than through decontextualized practice of forms. Laurillard's (2007) Conversational Framework, developed specifically for mobile and technology-mediated learning contexts, similarly emphasizes cycles of learner action, feedback, and revision as the mechanism through which technology-mediated learning becomes productive rather than merely receptive. Recent reviews of the MALL vocabulary literature have begun to gesture toward this direction, noting the growing use of interactive and gamified task formats as a promising development (Teymouri, 2024) and calling explicitly for research designs and applications with genuinely context-responsive, multi-dimensional task features rather than uniform recognition drills (Duan et al., 2026; Okumuş Dağdeler, 2023).

What is currently missing from this literature, however, is a systematic way of describing where any given MALL task or tool sits along the path from recognition to production, and therefore what a learner using it repeatedly is actually likely to develop. The framework proposed in the next section is intended to fill that gap.

3. A Proposed Framework: The Recognition-to-Production (R2P) Continuum

Building on the distinctions reviewed above, this paper proposes organizing mobile-assisted vocabulary tasks along a three-tier continuum, moving from recognition through controlled production to free production. The purpose of the continuum is diagnostic and generative: it offers vocabulary teachers, materials



designers, and researchers a shared vocabulary for auditing existing MALL activities and for deliberately designing new ones that do not stall at the recognition tier.

Tier 1 — Recognition tasks. These are tasks in which the learner selects, matches, or identifies a correct form or meaning from options already provided: digital flashcards, multiple-choice items, matching exercises, and true/false comprehension checks. Cognitively, these tasks require only that the learner recognize a previously encountered form-meaning pairing; no retrieval of form from meaning is required. Tier 1 tasks are inexpensive to build, trivially auto-gradable, and well suited to spaced-repetition scheduling, which explains their dominance in commercial MALL applications and their prominence in the research literature reviewed above.

Tier 2 — Controlled production tasks. These require the learner to retrieve and supply a target form, but within a constrained context that limits the range of acceptable answers: gap-fill sentences, sentence completion using a target collocation, guided paraphrase, or short-answer definitions supplied from memory. Controlled production narrows the cognitive distance between recognition and free use without yet requiring the learner to generate original context, and remains largely auto-gradable with well-designed answer keys.

Tier 3 — Free production tasks. These require the learner to generate original spoken or written language that deploys the target vocabulary to achieve a genuine communicative purpose: a short voice-recorded response to a prompt, a brief written email using a set of target terms, a role-play scenario conducted via a messaging app, or a recorded summary of a short reading. These tasks are the most cognitively demanding, the most difficult to auto-grade, and the most directly aligned with the productive knowledge that ESP instruction ultimately aims to build. They are also, per the evidence reviewed in Section 2, the tier that current MALL tools and curricula most consistently under-supply.



Mobile devices possess specific affordances that make all three tiers technically feasible at scale, even though current tools cluster heavily around Tier 1. Push notifications and spaced-repetition scheduling support Tier 1 recognition practice efficiently. Adaptive branching and short-answer input fields support Tier 2 controlled production. Voice recording, asynchronous messaging, and peer- or instructor-reviewed submission features — already built into widely available messaging platforms — support Tier 3 free production without requiring specialized software, a point of particular relevance in contexts where dedicated MALL applications may be less accessible than the general-purpose messaging apps students already use daily.

The central claim of the R2P framework is that the effectiveness of a MALL vocabulary tool or curriculum should not be evaluated by retention rate alone, but by **tier coverage**: whether learners who use it consistently are ever required to progress beyond recognition into controlled and free production. A tool or course that achieves excellent Tier 1 retention scores while never engaging learners in Tier 2 or Tier 3 tasks is, on the framework's terms, addressing only part of the vocabulary knowledge that instruction is meant to build — a pattern consistent with what the reviewed literature describes.

4. Pedagogical Implications for ESP and ELT Practice

For ESP instruction in particular, the R2P framework has direct practical implications. Business, economics, and technical English learners are typically preparing for communicative situations — writing a proposal, participating in a negotiation, presenting a report — that are Tier 3 by definition. A curriculum that relies solely on Tier 1 mobile practice for vocabulary acquisition, however well it performs on retention measures, is unlikely by itself to prepare learners for these demands.

Several practical directions follow from this. First, materials designers and instructors can use the three tiers as an audit tool, reviewing existing MALL



activities to identify where a course currently clusters — most will cluster heavily at Tier 1 — and deliberately adding Tier 2 and Tier 3 tasks rather than adding more Tier 1 volume. Second, blended designs that pair app-based Tier 1 practice (for efficient initial form-meaning encoding) with classroom- or mobile-mediated Tier 2 and Tier 3 tasks (for consolidation and use) are likely to be more effective at building productive competence than either mode alone, consistent with calls in the literature for hybrid approaches that combine digital efficiency with more demanding production work. Third, general-purpose mobile messaging platforms — already the subject of considerable MALL research in their own right — offer a low-cost route to Tier 3 tasks: voice notes responding to a prompt, short written summaries submitted to a class channel, or peer-to-peer role-plays conducted asynchronously via chat, all of which require minimal specialized infrastructure. This is a particularly relevant consideration in contexts where institutional access to commercial, feature-rich MALL platforms may be limited, but general messaging-app usage among students is close to universal.

Finally, it is worth noting a geographic pattern in the literature reviewed for this paper: the empirical base underlying current understanding of the receptive-productive gap in MALL is drawn overwhelmingly from East Asian, Middle Eastern, and, to a lesser extent, South and Southeast Asian contexts. Comparatively little of this research has been conducted in Central Asian higher education settings, despite the region's rapidly growing investment in English-medium and ESP instruction. This is not a claim that the underlying mechanisms differ by region — there is no evidence to suggest they do — but it does mean that the practical calibration of the R2P framework, including which tasks and platforms are most feasible for a given student population, would benefit from context-specific validation in currently underrepresented settings.

5. Limitations and Directions for Future Research



As a conceptual paper, the R2P framework synthesizes existing findings rather than reporting new empirical data, and its central claims should be understood accordingly: the receptive-productive asymmetry is empirically well supported by the studies reviewed here, but the three-tier framework itself, and its proposed diagnostic value, awaits direct empirical testing. Several directions for such testing follow naturally. Researchers could operationalize "tier coverage" as a measurable property of a MALL curriculum or tool and test whether curricula with broader tier coverage produce correspondingly stronger productive vocabulary outcomes than Tier-1-dominant curricula, holding overall study time constant. Longitudinal designs would be particularly valuable, since several of the studies reviewed above suggest that productive gains may consolidate more slowly than receptive gains, and short intervention periods may therefore understate them. Finally, given the geographic gap noted above, studies situated in Central Asian ESP contexts — testing, for example, whether messaging-app-mediated Tier 3 tasks are a feasible and effective supplement to app-based Tier 1 practice for business or economics students — would extend both the empirical and geographic reach of this literature.

6. Conclusion

The extensive body of MALL vocabulary research produced over the past decade has convincingly answered one question: mobile tools support vocabulary learning. It has, this paper has argued, left a second and arguably more consequential question comparatively underexamined: what kind of vocabulary knowledge do these tools build, and is it the kind that learners actually need? For ESP learners in particular, whose target language use is overwhelmingly productive, an instructional ecosystem that excels at building recognition while under-supplying production is solving only part of the problem. The Recognition-to-Production framework proposed here offers one way of naming and addressing that gap — not by abandoning the recognition-based tools that have proven their value, but by treating them as the first tier of a more complete design, one that deliberately carries learners



through controlled production and into the free, communicative use of vocabulary that ESP instruction ultimately exists to develop.

References

Boroughani, T., Behshad, N., & Xodabande, I. (2023). Mobile-assisted academic vocabulary learning with digital flashcards: Exploring the impacts on university students' self-regulatory capacity. *Frontiers in Psychology, 14*, Article 1112429. <https://doi.org/10.3389/fpsyg.2023.1112429>

Boroughani, T., Xodabande, I., & Karimpour, S. (2023). Self-regulated learning with mobile devices for university students: Exploring the impacts on academic vocabulary development. *Discover Education, 2*(1), Article 5. <https://doi.org/10.1007/s44217-023-00028-z>

Burston, J. (2015). Twenty years of MALL project implementation: A meta-analysis of learning outcomes. *ReCALL, 27*(1), 4–20.

Duan, H., Yunus, M. M., & Ismail, H. H. (2026). Mobile-assisted English vocabulary learning: A systematic review of platforms, learning outcomes, and implications (2019-2025). *International Journal of Learning, Teaching and Educational Research, 25*(2), 65–90. <https://doi.org/10.26803/ijlter.25.2.4>

Godwin-Jones, R. (2011). Emerging technologies: Mobile apps for language learning. *Language Learning & Technology, 15*(2), 2–11.

Koleini, N., Boroughani, T., Eslami, Z. R., & Xodabande, I. (2024). Exploring the impacts of mobile-assisted learning on university students' technical vocabulary knowledge. *International Journal of Educational Research Open, 7*, Article 100344. <https://doi.org/10.1016/j.ijedro.2024.100344>

Laufer, B., & Nation, P. (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics, 16*(3), 307–322.

Laurillard, D. (2007). Pedagogical forms for mobile learning: Framing research questions. In N. Pachler (Ed.), *Mobile learning: Towards a research agenda* (pp. 153–175). WLE Centre, IOE.



Lei, X., Fathi, J., Noorbakhsh, S., & Rahimi, M. (2022). The impact of mobile-assisted language learning on English as a foreign language learners' vocabulary learning attitudes and self-regulatory capacity. *Frontiers in Psychology, 13*, Article 872922. <https://doi.org/10.3389/fpsyg.2022.872922>

Li, Y., & Hafner, C. A. (2022). Mobile-assisted vocabulary learning: Investigating receptive and productive vocabulary knowledge of Chinese EFL learners. *ReCALL, 34*(1), 66–80.

Lin, J. J., & Lin, H. (2019). Mobile-assisted ESL/EFL vocabulary learning: A systematic review and meta-analysis. *Computer Assisted Language Learning, 32*(8), 878–919. <https://doi.org/10.1080/09588221.2018.1541359>

Mahdi, H. S. (2018). Effectiveness of mobile devices on vocabulary learning: A meta-analysis. *Journal of Educational Computing Research, 56*(1), 134–154.

Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge University Press.

Nation, I. S. P. (2013). *Learning vocabulary in another language* (2nd ed.). Cambridge University Press.

Okumuş Dağdeler, K. (2023). A systematic review of Mobile-Assisted Vocabulary Learning research. *Smart Learning Environments, 10*, Article 19. <https://doi.org/10.1186/s40561-023-00235-z>

Stockwell, G. (2010). Using mobile phones for vocabulary activities: Examining the effect of platform. *Language Learning & Technology, 14*(2), 95–110.

Teymouri, R. (2024). Recent developments in mobile-assisted vocabulary learning: A mini review of published studies focusing on digital flashcards. *Frontiers in Education, 9*, Article 1496578. <https://doi.org/10.3389/feduc.2024.1496578>

Viberg, O., & Grönlund, Å. (2012). Mobile-assisted language learning: A literature review. *Educational Technology & Society, 15*(3), 1–11.



Xodabande, I., & Atai, M. R. (2020). Using mobile applications for self-directed learning of academic vocabulary among university students. *Open Learning: The Journal of Open, Distance and e-Learning*. <https://doi.org/10.1080/02680513.2020.1847061>

Xu, Q., & Richardson, J. C. (2024). The impact of social media and gamification of a mobile vocabulary learning app: Self-regulation and learning persistence. *Online Learning*, 28(4). <https://doi.org/10.24059/olj.v28i4.4592>

Xu, Q., Richardson, J., Zhang, Z., Cheng, Z., & Guo, F. (2025). Using a mobile vocabulary application to enhance L2 learners' vocabulary acquisition: Possibilities and challenges. *Online Learning*, 29(3), 66–102. <https://doi.org/10.24059/olj.v29i3.4918>