Journal of Education, Linguistics and Psychology
Volume 1, Number 2 (2026) · September 2026
e-ISSN 3165-4436 · Research Center Progresif · jelp@researchcenterprogresif.org

The Use of English Songs to Improve Students' Vocabulary Mastery

Siti Radiana Sari
English Education Department, Faculty of Teacher Training and Education, Universitas Muria Kudus, Kudus, Indonesia
Corresponding author: Siti Radiana Sari, radiana0110@gmail.com

Article History — Received: 06-11-2026 · Accepted: 16-11-2026

Abstract

Background: English songs are popular in EFL classrooms and in learners' out-of-class lives, and Indonesian teachers often use them for vocabulary teaching. Yet the evidence on their effects is scattered across small, diverse studies. Aim: This review asks (1) what the evidence says about the effect of songs on vocabulary (form, meaning, collocation, retention; incidental versus intentional learning; listening versus singing), (2) what learners report about songs for vocabulary learning, and (3) what this implies for Indonesian classrooms and for a classroom study design. Method: A structured narrative review was conducted, with searches of OpenAlex, ERIC and Crossref on 6 November 2026. Every source was verified against these databases and summarised from its published abstract. Forty-five sources were included: 35 empirical studies (21 on learning outcomes, 12 on perceptions or beliefs, two corpus studies) and ten background works. Findings: Song-based input consistently produced vocabulary gains, most clearly for receptive and recognition-level knowledge. Repeated listening was associated with larger gains, explicit instruction with more durable gains, and singing tended to outperform listening or speaking alone. Learners reported enjoyment, motivation and autonomous use, but also difficulties with fast tempo, slang and figurative lyrics. Indonesian outcome studies were mostly short and single-site, with pre-experimental or quasi-experimental designs. Their abstracts rarely reported effect sizes, and none reported a delayed test. Implications: Songs work best as repeated, guided input combined with explicit vocabulary work. A quasi-experimental design with a control group, multi-dimensional vocabulary measures, a delayed post-test and a perception questionnaire is recommended.

Aim: This review asks (1) what the evidence says about the effect of songs on vocabulary (form, meaning, collocation, retention; incidental versus intentional learning; listening versus singing), (2) what learners report about songs for vocabulary learning, and (3) what this implies for Indonesian classrooms and for a classroom study design.

Method: A structured narrative review was conducted, with searches of OpenAlex, ERIC and Crossref on 6 November 2026. Every source was verified against these databases. Data were extracted from the published abstracts only. Full texts were not systematically read, and this is a principal limitation of the review. Forty-five sources

e-ISSN 3165-4436 · Open access, CC BY-SA 4.0

were included: 35 empirical studies (21 on learning outcomes, 12 on perceptions or beliefs, two corpus studies) and ten background works.

Findings: Song-based input was consistently reported to produce vocabulary gains, most clearly for receptive and recognition-level knowledge. This consistency is probably inflated by publication bias and by the reliance on abstracts. Repeated listening was associated with larger gains, explicit instruction with more durable gains, and singing tended to outperform listening or speaking alone. Learners reported enjoyment, motivation and autonomous use, but also difficulties with fast tempo, slang and figurative lyrics. Indonesian outcome studies were mostly short and single-site, with pre-experimental or quasi-experimental designs. Their abstracts rarely reported effect sizes, and none reported a delayed test.

Implications: Songs work best as repeated, guided input combined with explicit vocabulary work. A quasi-experimental design with a control group, multi-dimensional vocabulary measures, a delayed post-test and a perception questionnaire is recommended.

Keywords: English songs; vocabulary learning; incidental vocabulary learning; singing; learner perceptions; Indonesian EFL

1. Introduction

Vocabulary knowledge underpins every language skill. Learners who lack enough words struggle to understand what they hear and read and to express what they mean. Vocabulary research has long held that knowing a word involves more than linking a spoken or written form to a meaning. It also involves knowledge of the word's spoken and written forms, its grammatical behaviour, its collocations and the constraints on its use. Each of these aspects can be known receptively or productively (Nation, 2022; Schmitt & Schmitt, 2020). It is equally well established that vocabulary grows through two complementary routes: incidental learning while attention is on meaning, and deliberate, intentional learning in which words are the explicit object of study. Repeated encounters drive both routes (Webb & Nation, 2017). A meta-analysis of correlational studies found a medium overall relationship (r = .34) between the number of encounters and incidental vocabulary learning. The size of this relationship depended on learner variables such as age and prior vocabulary knowledge, and on treatment variables such as spacing, visual support and engagement (Uchihara et al., 2019). These insights give a useful lens for evaluating any classroom resource, including songs.

Songs have a long history in English language teaching. Practitioner and review literature argues that songs support listening, pronunciation, vocabulary and sentence structures. It also argues that songs carry cultural content and make lessons more enjoyable, particularly for young learners (Millington, 2011). Engh (2013) surveyed literature from cognitive science, anthropology, sociolinguistics, psycholinguistics and first and second language acquisition and argued that there is a strong relationship between music and language that can justify teachers' intuitive use of songs. Teachers appear to share this view. An international questionnaire of 568 respondents in 41 countries found that most respondents believed songs were useful for language learning. Many reported using songs for defined pedagogical

e-ISSN 3165-4436 · Open access, CC BY-SA 4.0

purposes, including vocabulary learning (Tegge, 2015). Outside the classroom, listening to songs in English is described as one of the most common out-of-class activities reported by EFL learners (Mannarelli & Serrano, 2024; Chang, 2024). Learners now meet songs through streaming and short-video platforms such as Spotify and TikTok (Butar Butar & Katemba, 2023; Isnaeni & Farida, 2026).

Despite this popularity, the empirical base has developed unevenly. Pavia et al. (2019) observed that research on incidental vocabulary learning from songs had relied mainly on self-report surveys linking listening habits to vocabulary knowledge. In their account, only one earlier experimental study had examined vocabulary learning from listening to songs, and it did not report the size of the learning gains. Since then, experimental work has grown. Studies have measured specific dimensions of word knowledge, the effect of repetition, the role of printed lyrics, and the comparison between singing and other modes of practice (e.g., Chang, 2024; Chen et al., 2026; Nie et al., 2022; Zhang et al., 2023). In parallel, many Indonesian studies have examined songs in junior and senior high schools and universities. They are often small classroom experiments or qualitative perception studies published in national journals (e.g., Gushendra, 2017; Lestari & Hardiyanti, 2020; Nasia P et al., 2026). Existing reviews either take a broad view of music and language learning (Engh, 2013) or focus on informal engagement with pop music (Ludke & Morgan, 2022). To our knowledge, no recent review brings together international experimental evidence on specific vocabulary dimensions and the Indonesian classroom literature. No review has used that synthesis to inform the design of a classroom study.

This review was prompted by an undergraduate English Language Teaching proposal titled "The Use of English Songs to Improve Students' Vocabulary Mastery". The proposal asks whether songs improve students' vocabulary mastery and how students perceive learning vocabulary through songs. Before such a study is run, it is worth asking what the existing evidence can and cannot tell us. This review therefore addresses three review questions:

> RQ1. What does the evidence from 2013 to 2026 say about the effect of English songs on vocabulary learning, in terms of form, meaning, collocation and retention, incidental versus intentional learning, and listening versus singing?
>
> RQ2. What do learners (and, secondarily, teachers) report about using songs for vocabulary learning?
>
> RQ3. What are the implications for Indonesian EFL classrooms and for the design of a classroom study on songs and vocabulary mastery?

2. Method

2.1 Review type

This article is a structured narrative review, not a systematic review or meta-analysis. The aim was to map and interpret the evidence transparently, using explicit searches and inclusion criteria, without claiming the exhaustive coverage, duplicate screening or quantitative pooling of a systematic review. No PRISMA flow diagram is reported because a full PRISMA

e-ISSN 3165-4436 · Open access, CC BY-SA 4.0

screening procedure was not carried out. The hit counts below are the raw counts returned by the databases' relevance-ranked searches, not screened pools.

2.2 Sources and searches

All searches were run on 6 November 2026. They used three openly accessible bibliographic services.

(a) OpenAlex (works endpoint). Searches were restricted to journal articles published from 1 January 2013 onwards and sorted by relevance. For each query, the 12 to 20 top-ranked records were inspected. The queries (with raw hit counts) were: "incidental vocabulary learning songs" (3,382); "songs vocabulary learning EFL" (18,386); "singing vocabulary learning foreign language" (7,968); "music lyrics vocabulary acquisition second language" (2,769); "songs young learners English vocabulary" (16,943); "lexical coverage popular songs" (2,971); "efficacy of singing foreign language learning" (2,899); "music video captions vocabulary learning" (2,359); "songs vocabulary meta-analysis music language learning" (4,295); "English songs vocabulary mastery quasi-experimental Indonesia" (919); "audio-visual input on-screen text second language vocabulary" (6,507); and "song lyrics vocabulary profile EFL" (479). The raw hit counts (479 to 18,386 per query) were far too large to screen exhaustively in a single-reviewer narrative review. The inspection therefore relied on OpenAlex's relevance ranking to bring the most pertinent records to the top. Three steps were taken to reduce the risk of missing key studies. Twelve overlapping queries targeting different aspects of the review questions were used. The ERIC search, unlike the OpenAlex searches, was inspected in full (see (b)). The reference list of the undergraduate proposal was used as a seed set (see below). Records ranked below the inspected top records were not examined. Relevant studies that OpenAlex ranked lower may therefore have been missed, and the relevance-ranking procedure itself is not fully transparent to the user. This is acknowledged as a limitation in Section 6.

(b) ERIC (IES API). The query was songs AND vocabulary AND (EFL OR ESL). It returned 39 records, and all 39 titles were inspected.

(c) Crossref (REST API). Crossref was used to verify each candidate DOI and to retrieve bibliographic details. It was also used for bibliographic queries that located DOIs for items found in ERIC, and for prefix-restricted queries that tried to locate the real records behind non-resolving DOIs.

The reference list of the undergraduate proposal served as a seed set. It contained 21 DOIs. Sixteen resolved in Crossref and OpenAlex. Three were found in OpenAlex with working publisher pages, but their DOIs were not registered (the doi.org handle service returned "not found"). Two could not be traced to any record. Of the three unregistered DOIs, two studies with confirmed publisher pages were included (Nasia P et al., 2026; Rohmah & Indah, 2021) and are cited with their publisher URLs. The third was excluded (see Section 2.3). One further included study (Karim et al., 2022) also had an unregistered DOI. It was verified through its OpenAlex record and is cited with its publisher URL. One study found through ERIC

e-ISSN 3165-4436 · Open access, CC BY-SA 4.0

(Magnussen & Sukying, 2021) has no DOI and is cited with its ERIC record. One abstract (Ludke et al., 2014) was not available in OpenAlex or Crossref. It was read in Europe PMC, and the existence of the work was confirmed in Crossref and OpenAlex.

2.3 Inclusion and exclusion criteria

Sources were included if they met all of the following criteria. (1) They were published between 2013 and 2026. One earlier practitioner article from the seed list (Millington, 2011) was retained as background and is flagged as such. (2) They examined songs, singing or song lyrics in relation to second or foreign language vocabulary, or provided directly relevant theoretical or methodological background (vocabulary knowledge, repetition, multimodal input). (3) Their existence and bibliographic details could be verified in Crossref, OpenAlex or ERIC. (4) An abstract or open-access text was available, so that what is reported here could be checked. (5) They were not published in journals of Research Center Progresif. (6) The venue did not raise obvious concerns about editorial quality.

Studies with English as the target language were prioritised. A small number of studies with other target languages (Hungarian, French) were included because they bear directly on the comparison between singing and speaking. Studies of singing for grammar or fluency alone, studies of music as background noise, and studies of first-language vocabulary were excluded. Several otherwise relevant records were excluded because no abstract could be read through the services above. These included a study of songs and stories with preschool learners in *System* (2018), a study of music as a mnemonic device (2020), and a corpus study of the lexical coverage of popular songs (2017). Two seed-list DOIs that could not be traced were also excluded. So was one seed-list item from a venue whose editorial quality could not be judged, a two-participant interview study. A lexical comparison of curriculum and YouTube songs was excluded as a precaution because of uncertainty about its publication venue.

2.4 Data extraction and synthesis

For each included empirical study, the following information was extracted from its published abstract: authors and year, country or setting, learners, design, vocabulary measures and main finding. The published abstract was the basis for every entry in Table 1 and for the descriptions of individual studies in Section 3. Criterion (4) in Section 2.3 was applied so that this abstract could be read and checked. Information not stated in the abstract is marked "not reported in abstract" in Table 1. Full texts were not systematically read. The descriptions in this review are therefore limited to what the published abstracts state, and no figures were added beyond those reported there. Means and significance values are given only when the abstract itself reported them. A consequence of this procedure is that information marked as not reported may well be reported in the full text. The absence of a detail from an abstract (for example, a control group, an effect size or a delayed post-test) should therefore not be read as its absence from the study. Studies were grouped thematically to answer the review questions, and no effect sizes were pooled. Screening and extraction were carried out by a single reviewer with the support of an AI writing assistant (see the Declaration of Generative AI Use). No dual coding or formal risk-of-bias assessment was performed. The reliance on

e-ISSN 3165-4436 · Open access, CC BY-SA 4.0

abstracts rather than full texts is a major limitation of this review and is discussed in Section 6.

2.5 Included sources

In total, 45 sources were included. Thirty-five were empirical studies: 21 measured vocabulary (or, in two cases, closely related language) outcomes of song-based treatments; 12 examined learners' or teachers' perceptions and beliefs (one of these, Tegge, 2015, also contained corpus and classroom components); and two analysed the lexical profile of song lyrics. The remaining ten were background works: three vocabulary books, five reviews or practitioner syntheses, one meta-analysis, and one empirical study of vocabulary learning from audiovisual (non-song) input (Peters, 2019). Sixteen of the 35 empirical studies were Indonesian, meaning that they were conducted in Indonesia or, where the abstract did not name the setting, were written by Indonesia-based authors. The remaining studies involved learners in or from Algeria, China, Ecuador, Hungary, Iran, Italy, Spain, Taiwan and Thailand, or did not state the setting in the abstract.

3. Findings

Table 1. Summary of included empirical studies (descriptions based on published abstracts; "Not reported in abstract" = information not stated in the abstract)

Author (year)	Country / setting	Learners	Design	Vocabulary measure	Main finding (as reported)
A. Studies of vocabulary (or closely related) learning outcomes					
Coyle & Gómez Gracia (2014)	Spain	25 five-year-old preschool EFL learners	One group; three 30-min song-based lessons; pre-, post- and 5-week delayed test	Vocabulary picture tests	Receptive gains; song input insufficient for productive knowledge in most children
Ludke et al. (2014)	Not reported in abstract	60 adults; unfamiliar L2 (Hungarian)	Random assignment to three listen-and-repeat conditions: speaking, rhythmic speaking, singing	Hungarian language tests, incl. recall and production of spoken phrases	Singing group best overall; significant on the two recall/production tests
Good et al. (2015)	Ecuador	Spanish-speaking children (n not reported in abstract)	Sung vs. spoken (oral poem) learning of an English passage over 2 weeks; 6-month delay	Verbatim recall; vowel pronunciation; translation of target terms	Sung condition outperformed spoken in all three; advantage persisted after 6 months
Leśniewska & Pichette (2016)	Not reported in abstract (French-speaking children)	24 children aged 35–59 months	57 words in songs and stories across four weekly workshops	Picture-selection recall	No difference between songs and stories; animate concepts recalled better; more encounters likely needed
Gushendra (2017)	Indonesia	40 students (level not reported in abstract)	Experimental vs. control class	Vocabulary mastery test	Experimental mean 82.25 vs. control 63.25; significant
Ludke (2018)	Not reported in abstract	Native English speakers aged	Exploratory quasi-	Curriculum-based French	Both gained; greatest

e-ISSN 3165-4436 · Open access, CC BY-SA 4.0

Author (year)	Country / setting	Learners	Design	Vocabulary measure	Main finding (as reported)
		12–13, beginning French	experiment; singing/song vs. visual art/drama supplements, 8.5 h over 6 weeks	tests (overall, not vocabulary-specific)	improvement in singing group
Pavia et al. (2019)	Not reported in abstract	Not reported in abstract	Incidental listening to two songs; one, three or five listenings	Spoken-form recognition; form–meaning connection; collocation recognition	Songs contributed to learning; repetition and frequency of exposure positive
Magnussen & Sukying (2021)	Thailand	72 preschoolers aged 4–5	Three classes: songs, TPR, songs + TPR; 6 weeks; 12 target words; mixed methods	Two tests (pre/post); video recordings	All methods improved acquisition; combination best
Rohmah & Indah (2021)	Indonesia	40 junior high school students	Experimental (Modified English Song) vs. control	Vocabulary test	Post-test means 86.80 vs. 82.80; significant t-test
Nie et al. (2022)	Chinese college students (setting not reported in abstract)	114 college students	Incidental listening; one, three or five listenings; prior knowledge examined; 4-week delay	Word recognition; meaning association; grammar identification	Gains esp. word recognition, retained after 4 weeks; three listenings best
Tilwani et al. (2022)	Iran	150 female learners aged 11–15	Four experimental and two control groups; pre-test presence varied; five 90-min sessions; 3-week delayed test	Spoken-form recognition; form–meaning connection; collocation recognition	Experimental > control; pre-tested vs. non-pre-tested experimental groups differed
Sarnou (2023)	Algeria	51 third-year elementary pupils (aged 8–9)	Experimental (26; nursery rhymes) vs. control (25; textbook only); 3 weeks	Vocabulary scores	Experimental scores considerably higher
Zhang et al. (2023)	Chinese learners (setting not reported in abstract)	95 adolescent ESL learners (M = 14.04 years)	Three sessions; singing vs. reciting lyrics to familiar melodies	Picture naming; meaning recall; rated pronunciation	Both improved; singing group outperformed speech group
Chang (2024)	Not reported in abstract	87 university EFL learners	Counterbalanced: listening + singing, listening + oral reading, listening only; 1- and 8-week delayed tests	Aural and written meaning recognition; written form production	Higher scores for singing mode and written meaning recognition; song and time effects complex
Huang & Chen (2024)	Taiwan	34 university freshmen	Mixed methods; LyricsTraining app; pre/post (no control reported in abstract)	Listening and vocabulary tests; survey	Significant improvement; positive perceptions
Mannarelli & Serrano (2024)	Italy	Two groups of grade 8 EFL learners	Short treatment with pop songs; explicit vs. implicit focus	Not reported in abstract	Both improved; explicit approach more durable
Pravitasari &	Indonesia	39 eighth graders,	One-group	Written	Means 48.71 to

e-ISSN 3165-4436 · Open access, CC BY-SA 4.0

Author (year)	Country / setting	Learners	Design	Vocabulary measure	Main finding (as reported)
Rahmah (2024)		Islamic junior high school	pre/post; songs + vocabulary self-collection	vocabulary test	80.89; significant
Chen et al. (2026)	Not reported in abstract	114 participants	Three groups: no intervention, listening with lyrics, listening without lyrics	Spoken-form recognition; meaning recognition; meaning recall	Both listening groups > control; lyrics gave no significant added gain
Isnaeni & Farida (2026)	Indonesia	Junior high school students (n not reported in abstract)	Pre/post (described as quasi-experimental); five TikTok songs; paired t-test	Descriptive vocabulary test	Mean increase of 10.07 points; significant
Nasia P et al. (2026)	Indonesia	41 second-year vocational (SMK) students	Non-equivalent control group (21 vs. 20)	35-item multiple-choice test (alpha = 0.931)	Experimental 71.76 to 86.86; control 74.85 to 78.60; significant
Sunra et al. (2026)	Indonesia	25 seventh graders	One-group pre/post; five sessions with modified songs; interviews (n = 3)	Vocabulary achievement test; interviews	Means 69.64 to 89.12 (N-gain 64.47%); more engagement and enjoyment
B. Studies of perceptions and beliefs					
Tegge (2015)	International survey (41 countries); NZ thesis	568 teacher respondents; corpus of 635 songs; classroom study (n not reported in abstract)	Questionnaire; corpus analysis; classroom comparison of song, poem and prose lessons	Teacher beliefs; lexical coverage; verbatim recognition and cued recall	Most teachers believe songs useful; 4,000 families give 98% coverage; song and poem lessons > prose
Lestari & Hardiyanti (2020)	Indonesia	4 second-year English department students	Qualitative; in-depth interviews	Self-report	Prefer slow tempo; songs give authentic lexis, emotion, enjoyment, retention; cognitive and metacognitive strategies
Suciati & Zarkasih (2021)	Indonesia	6 grade X students	Qualitative; interviews	Self-report	Songs relaxing and enjoyable; enrich vocabulary and pronunciation
Amien et al. (2022)	Indonesia	20 university students (Jakarta)	Descriptive qualitative; questionnaire and interview	Self-report	Pop songs seen as promising for vocabulary and other skills
Karim et al. (2022)	Indonesia	56 university students (6 interviewed)	Online survey (4-point Likert) and interviews	Self-report	Most agreed songs facilitate learning and prompt looking up meanings; two interviewees mixed
Wardiman et al. (2022)	Indonesia	30 university English students	Questionnaire (Google Form)	Self-report	Songs perceived to help listening, pronunciation, vocabulary and grammar
Butar Butar &	Indonesia	97 secondary and	Descriptive	Self-report	Positive attitudes

e-ISSN 3165-4436 · Open access, CC BY-SA 4.0

Author (year)	Country / setting	Learners	Design	Vocabulary measure	Main finding (as reported)
Katemba (2023)		tertiary students	quantitative; questionnaire		toward vocabulary learning via Spotify songs
Fadli et al. (2023)	Not reported in abstract	10 participants	Qualitative; WhatsApp chat interviews; thematic analysis	Self-report	Favourable views; motivation, engagement, retention, confidence
Basori & Setyowati (2024)	Not reported in abstract	Not reported in abstract	Qualitative; questionnaires and interviews	Self-report	Perceived gains in vocabulary and pronunciation; challenges; two coping strategies
Rahman et al. (2024)	Indonesia	2 university students	Qualitative case study (self-directed learning); interviews	Self-report	Motivation guides song choice; repetition; looking up words; fast lyrics and slang difficult
Ghidhaoui (2025)	Hungary	5 university EFL instructors	Exploratory interview study	Teacher beliefs	Songs motivational and mnemonic; barriers: copyright, age, proficiency, lyric appropriateness
Putri et al. (2026)	Indonesia	8 seventh-semester English education students	Qualitative descriptive; semi-structured interviews	Self-report	Fun and motivating; challenges: tempo, accents, slang, metaphor, distraction
C. Lexical-profile (corpus) studies					
Romanko (2017)	Not reported in abstract	Corpus of 2,175 songs (678,309 tokens)	Lexical coverage analysis	Coverage at 95% and 98%	2,000 families + PN/MW = 96.05%; 5,000 + PN/MW = 98%
Yuanjai & Gampper (2025)	Not reported in abstract	Taylor Swift pop song corpus	Lexical profiling (AntConc) against Oxford 3000/5000 and CEFR	List and CEFR coverage	18.08% Oxford 3000; 1.99% Oxford 5000; A1 most frequent level (41.44%)

Note. PN/MW = proper nouns and marginal words; TPR = total physical response; SMK = Indonesian vocational high school. Means and significance results are reproduced only as reported in the abstracts. Peters (2019), a study of audiovisual (non-song) input, is discussed as background and is not listed here.

3.1 Overall direction of the evidence (RQ1)

Across the 21 outcome studies, the direction of the findings was consistent. Every study reported that learners who worked with songs gained vocabulary, or that a song condition performed at least as well as its comparison condition. No included abstract reported that songs harmed vocabulary learning. This consistency should be read with caution, however. The studies varied widely in rigour, from one-group pre-test–post-test designs to randomised and counterbalanced experiments. Many compared a song treatment with "conventional"

e-ISSN 3165-4436 · Open access, CC BY-SA 4.0

teaching that was not described in the abstract. In addition, the evidence base is likely to be affected by publication bias. Positive results are more likely to be published than null results. Apart from one doctoral dissertation (Tegge, 2015), the included empirical studies were published journal articles, and unpublished theses and other grey literature were not searched. The few results that did not favour songs are therefore particularly informative. These are: no difference between words learned from songs and from stories (Leśniewska & Pichette, 2016); no significant added gain from printed lyrics (Chen et al., 2026); and song input that was insufficient for productive knowledge in most children (Coyle & Gómez Gracia, 2014). Because no effect sizes were pooled, publication bias could not be assessed formally. The uniformly positive direction of the findings should therefore be read as a feature of the published, abstract-level record, not as proof that songs always improve vocabulary learning.

The strongest designs came mostly from outside Indonesia. Pavia et al. (2019) examined incidental learning from listening to two songs. They measured three dimensions of vocabulary knowledge (spoken-form recognition, form–meaning connection and collocation recognition) and found that listening to songs contributed to vocabulary learning. Nie et al. (2022) examined 114 Chinese college students. They found that listening to songs improved vocabulary knowledge, particularly word recognition, which was retained four weeks later. Tilwani et al. (2022) assigned 150 Iranian female learners aged 11 to 15 to four experimental and two control groups. The control groups did not listen to songs. The experimental groups outperformed the control groups, and the groups differed in spoken-form recognition, form–meaning connection and collocation recognition. Chen et al. (2026) compared 114 participants in three groups: no intervention, listening to a pop song with lyrics, and listening without lyrics. Both listening groups outperformed the control group on spoken-form recognition, meaning recognition and meaning recall.

3.2 Which aspects of word knowledge benefit?

The studies that separated dimensions of word knowledge suggest that songs mainly build recognition-level and receptive knowledge. Nie et al. (2022) reported the clearest gains in word recognition. For low-, intermediate- and high-level learners alike, they found immediate positive effects on word recognition and meaning connection that were likely to be preserved after four weeks. In preschool settings, Coyle and Gómez Gracia (2014) taught 25 five-year-old Spanish EFL learners through three 30-minute lessons built around a well-known children's song. They found evidence of growth in receptive vocabulary. However, exposure to the song input was not enough to develop productive knowledge in most of the children. Chang (2024) measured aural meaning recognition, written meaning recognition and written form production. Scores were highest for written meaning recognition. This again points to recognition of meaning as the dimension most readily supported.

Collocation was measured as a separate dimension in two studies. Pavia et al. (2019) and Tilwani et al. (2022) both included collocation recognition among the dimensions measured. Both reported that songs contributed to vocabulary learning overall, but their abstracts do not report collocation results separately. This matters because songs present words in recurring multi-word sequences such as refrains and repeated lines. Tegge's (2015) classroom study

e-ISSN 3165-4436 · Open access, CC BY-SA 4.0

found that song-based (and poem-based) lessons led to higher recognition and cued recall of verbatim text than a prose-based lesson. Tegge argued that songs may be especially useful for entrenching already "(half-)familiar" words together with their phraseological patterning. Productive knowledge of form, such as spelling a word or using it in writing, was assessed less often. Isnaeni and Farida (2026) reported that, after five TikTok songs, students' descriptive vocabulary improved and their writing became "longer, more detailed, and clearer". Their abstract does not describe how this was measured.

3.3 Repetition and retention

Repetition is central to song-based learning, and the evidence supports its importance but also its limits. Pavia et al. (2019) found that repeated listening to a single song (one, three or five times) had a positive effect on vocabulary gains, and that the frequency with which target words occurred in the songs was positively related to learning. Nie et al. (2022) found that listening three times, with exposure frequencies of three to nine, produced better results than listening once or five times. These results are consistent with the meta-analytic finding of a medium frequency–learning relationship (Uchihara et al., 2019). They also suggest that more listening is not always better, perhaps because of fatigue or loss of attention. With very young learners, Leśniewska and Pichette (2016) concluded that children aged 35 to 59 months may need more encounters than adults to achieve significant recall.

Retention was examined with delayed post-tests in several studies. Good et al. (2015) found that a song advantage among Spanish-speaking Ecuadorian children persisted after a six-month delay. Coyle and Gómez Gracia (2014) used a delayed test five weeks after teaching. Tilwani et al. (2022) used one three weeks after the treatment, and Nie et al. (2022) one four weeks after. Chang (2024) used one-week and eight-week delayed tests and reported that the songs themselves and the timing of tests played "an important but complicated role" in retention. Mannarelli and Serrano (2024) compared explicit and implicit instructional approaches to pop songs with grade 8 Italian learners. Vocabulary improved in both conditions, but the explicit approach led to more durable gains. In contrast, none of the Indonesian outcome abstracts reviewed here reported a delayed post-test.

3.4 Incidental versus intentional learning

The evidence suggests that songs can support incidental learning, and that adding deliberate attention to words strengthens and prolongs the gains. Incidental learning was demonstrated when learners simply listened (Chen et al., 2026; Nie et al., 2022; Pavia et al., 2019). However, the best-documented comparison of instructional focus (Mannarelli & Serrano, 2024) favoured the explicit approach for retention. Classroom studies in Indonesia that combined songs with explicit vocabulary procedures also reported gains. Pravitasari and Rahmah (2024) combined English songs with a vocabulary self-collection strategy for 39 eighth-grade students at an Islamic junior high school. They reported mean scores rising from 48.71 to 80.89. Sunra et al. (2026) integrated modified songs with targeted vocabulary instruction and interactive activities for 25 seventh graders over five sessions. They reported means rising from 69.64 to 89.12 (N-gain 64.47%). Because both studies used one-group

e-ISSN 3165-4436 · Open access, CC BY-SA 4.0

designs, however, they cannot separate the contribution of songs from that of the accompanying instruction or from ordinary learning over time.

Printed lyrics are a further way to direct attention to word forms. Chen et al. (2026) found that listening with lyrics, a dual-modality condition, did not lead to significantly greater gains than listening without lyrics on any of the three dimensions measured. This contrasts with research on audiovisual input. Peters (2019) found that learners who watched a documentary with captions made the largest vocabulary gains, and that on-screen imagery near the spoken word was positively related to learning. Montero Perez (2022) reviews this wider literature on audio-visual input and on-screen text. Huang and Chen (2024) used the LyricsTraining application, which requires learners to engage with lyrics while listening. They reported significant improvement in listening and vocabulary among 34 Taiwanese university students, together with positive perceptions. As far as the abstract indicates, however, there was no control group. Whether written lyrics help vocabulary learning from songs therefore remains an open question. The answer may depend on whether learners actively process the text rather than merely see it.

3.5 Listening versus singing

A consistent pattern across studies with different learners and target languages is that singing outperforms speaking, reciting or listening alone. Ludke et al. (2014) randomly assigned 60 adults to one of three "listen-and-repeat" conditions (speaking, rhythmic speaking or singing) for learning Hungarian phrases. The singing group performed best overall. The difference was significant on the two tests that required recall and production of spoken Hungarian phrases, and it was not explained by age, gender, mood, phonological working memory, or musical ability and training. Good et al. (2015) found that children who learned an English passage as a song outperformed those who learned it as an oral poem in verbatim recall, vowel pronunciation and translation of target terms. Zhang et al. (2023) trained 95 adolescent Chinese ESL learners (mean age 14.04) over three sessions to learn lyrics set to familiar melodies, either by singing or by reciting. Both groups improved in vocabulary (picture naming and meaning recall) and in pronunciation, but the singing group outperformed the speech group. Chang (2024) used a counterbalanced design with 87 university learners, each practising songs in three modes: listening plus singing, listening plus reading lyrics aloud, and listening only. Students scored higher overall in the singing mode.

Evidence beyond vocabulary points in the same direction. Ludke (2018) compared two groups of 12- to 13-year-old beginners of French whose lessons were supplemented either with singing and song activities or with visual art and drama for 8.5 hours over six weeks. Both groups gained, but the singing group showed the greatest percentage improvement on curriculum-based French tests. Two studies with young children show that the benefit depends on the comparison. Magnussen and Sukying (2021) found that songs, total physical response (TPR) and a combination of both all improved Thai preschoolers' acquisition of 12 target words, and that the combination worked best. Leśniewska and Pichette (2016), however, found no difference in recall between words presented in songs and words presented

e-ISSN 3165-4436 · Open access, CC BY-SA 4.0

in stories. Overall, singing appears to add value over passive listening or plain speech. Songs do not appear to be uniquely superior to other rich, engaging input such as stories.

3.6 Young learners

Studies with young learners mostly report positive outcomes, but they also show the limits of song input alone. Sarnou (2023) taught number, colour and family vocabulary through nursery rhymes to 26 Algerian third-year elementary pupils for three weeks. A control group of 25 pupils used only the textbook. The song group's average vocabulary scores improved considerably compared with the control group. Magnussen and Sukying (2021) and Coyle and Gómez Gracia (2014) reported gains among preschoolers, although in the latter study the gains were receptive only. Leśniewska and Pichette (2016) highlighted the role of word characteristics: animate concepts were recalled better than inanimate ones. Butler's (2019) synthesis of research on teaching vocabulary to young learners recommends frequent and repeated exposure to target words, explicit definitions and meanings in context, opportunities for discussion and interaction around words, and multimodal approaches. Songs combined with movement, pictures and talk about the words fit these recommendations better than songs played on their own.

3.7 Indonesian classroom studies

Six Indonesian outcome studies were included. Gushendra (2017) compared an experimental and a control class (40 students). The abstract reports a mean vocabulary score of 82.25 in the experimental class and 63.25 in the control class, with a statistically significant difference. Rohmah and Indah (2021) compared 40 junior high school students taught with or without "Modified English Song" as a medium. They reported post-test means of 86.80 and 82.80 and a significant t-test. Nasia P et al. (2026) used a non-equivalent control group design with 41 second-year vocational students. They used a 35-item multiple-choice test with reported reliability (Cronbach's alpha = 0.931). Means rose from 71.76 to 86.86 in the experimental class and from 74.85 to 78.60 in the control class. Pravitasari and Rahmah (2024), Isnaeni and Farida (2026) and Sunra et al. (2026) used single-group pre-test–post-test designs, sometimes described as quasi-experimental.

Together, these studies show that Indonesian teachers and researchers have tried songs across junior high, senior high, vocational and university settings, and that they report consistent short-term gains. Their abstracts, however, share several features that limit the strength of the conclusions. Where reported, sample sizes were small (25 to 41). Half of the studies had no control group, and the comparison groups in the others are not described as randomly assigned. Where duration was reported, interventions were short (for example, five sessions in Sunra et al., 2026). Results were reported as a single overall "vocabulary mastery" score rather than as separate dimensions of word knowledge. Results were reported as means and significance tests, and none of the abstracts reported an effect size other than the N-gain in Sunra et al. (2026). None reported a delayed post-test. These are limits of the reporting and design visible in the abstracts, not judgments of the full studies, which were not systematically read.

e-ISSN 3165-4436 · Open access, CC BY-SA 4.0

3.8 The lexical demands of songs

Three corpus-based sources help explain why songs may suit some vocabulary goals better than others. Romanko (2017) analysed 2,175 popular songs (678,309 tokens). Knowledge of the most frequent 2,000 word families, plus proper nouns and marginal words, gave 96.05% coverage. Knowledge of the most frequent 5,000 word families, plus proper nouns and marginal words, gave 98% coverage. Tegge (2015) analysed 635 songs used for ESL/EFL purposes (177,384 tokens). Teacher-selected songs were short, repetitive and relatively undemanding in their vocabulary compared with other authentic genres. Knowledge of the 4,000 most frequent word families gave 98% coverage, with little difference between songs intended for beginners and those intended for advanced learners. Yuanjai and Gampper (2025) profiled Taylor Swift's lyrics against the Oxford 3000 and 5000 lists and CEFR levels. The most frequent CEFR level in the songs was A1, with a lexical coverage of 41.44%. Taken together, these findings suggest that popular songs mostly recycle high-frequency vocabulary. They may therefore be better suited to consolidating and deepening knowledge of familiar and mid-frequency words, including their collocations, than to introducing many new low-frequency words.

3.9 Learners' and teachers' perceptions (RQ2)

The perception studies, most of them Indonesian, show a highly consistent picture of positive attitudes alongside specific difficulties. Four themes recur.

Enjoyment, motivation and lower pressure. Learners describe songs as fun, relaxing and motivating (Fadli et al., 2023; Putri et al., 2026; Suciati & Zarkasih, 2021). In a questionnaire study of 97 secondary and tertiary students, respondents expressed positive attitudes, feelings and perceptions toward learning vocabulary by listening to English songs on Spotify (Butar Butar & Katemba, 2023). Karim et al. (2022) surveyed 56 university students and interviewed six. Most respondents agreed that songs facilitated learning and were an interesting medium, although two interviewees saw both advantages and disadvantages. Outcome studies that collected interview data also reported higher engagement, enjoyment and confidence (Sunra et al., 2026) or positive perceptions of a lyrics application (Huang & Chen, 2024).

Perceived gains beyond vocabulary. Learners frequently attribute improvements in pronunciation and listening to songs, as well as in vocabulary (Basori & Setyowati, 2024; Suciati & Zarkasih, 2021; Wardiman et al., 2022). Amien et al. (2022) reported that 20 Jakarta university students saw listening to English pop songs as a promising activity for augmenting vocabulary and other skills.

Autonomy and strategies. Several studies describe learners using songs on their own outside the classroom and applying deliberate strategies. Lestari and Hardiyanti (2020) interviewed four second-year English department students. These students chose songs for their authentic and contextual vocabulary, their emotional expressiveness, their enjoyable atmosphere and their help with retention, and they used cognitive and metacognitive strategies while listening. Rahman et al. (2024) described two university students engaged in self-directed learning. Personal motivation guided their song choice, repetition was vital for retention, and they

e-ISSN 3165-4436 · Open access, CC BY-SA 4.0

looked up unfamiliar words online. Karim et al. (2022) reported that most respondents agreed they would look up the meanings of lyrics.

Difficulties. Learners consistently report problems with fast tempo and unclear pronunciation, unfamiliar accents, slang and metaphor, and being distracted by the music rather than the words (Putri et al., 2026; Rahman et al., 2024). Learners in Lestari and Hardiyanti (2020) preferred songs with a slow tempo so that the words were clear. Basori and Setyowati (2024) identified two ways in which students coped with such challenges and argued that teachers need to take them into account.

Teachers' views add a practical dimension. Most teachers in Tegge's (2015) international survey believed songs were useful for vocabulary learning. Five Hungarian university EFL instructors interviewed by Ghidhaoui (2025) acknowledged songs' motivational and mnemonic value for vocabulary retention. They also named obstacles to classroom use: copyright, learners' age and proficiency, and the appropriateness of lyrics. Ludke and Morgan's (2022) systematic analysis of qualitative studies on informal engagement with foreign-language pop music identified six themes. The authors called for more research on learner perspectives and on how learners engage with L2 pop music autonomously.

Perceptions are not evidence of learning. Most perception studies used small purposive samples or self-selected questionnaire respondents, and several asked students who were already enthusiastic about songs. Their value lies in identifying the conditions under which learners find songs useful or difficult, which can inform the design of instruction and of research instruments.

4. Discussion

This review set out to establish what recent evidence says about English songs and vocabulary learning. Four broad conclusions emerge.

First, songs are a genuine source of vocabulary learning, not merely a motivational extra. Controlled studies show gains from listening alone (Chen et al., 2026; Nie et al., 2022; Pavia et al., 2019). They also show gains from song-based teaching compared with no-song or textbook-only conditions (Sarnou, 2023; Tilwani et al., 2022). These findings fit the broader account of incidental vocabulary learning, in which meaning-focused input with repeated encounters produces gradual, incremental growth (Uchihara et al., 2019; Webb & Nation, 2017). The gains appear strongest at the level of recognition: recognising a word's spoken form, matching a form to its meaning, or recognising a collocation. Productive knowledge appears to need more than exposure (Coyle & Gómez Gracia, 2014). This is what a word-knowledge framework would predict (Nation, 2022; Schmitt & Schmitt, 2020). Receptive knowledge develops earlier and from less input than productive knowledge.

Second, how songs are used matters at least as much as whether they are used. Mannarelli and Serrano (2024) found that explicit attention led to more durable gains. Tegge (2015) argued that the mnemonic potential of songs is realised through the way teachers exploit them in class. Indonesian studies that combined songs with structured vocabulary procedures also reported gains (Pravitasari & Rahmah, 2024; Sunra et al., 2026), although their designs cannot

e-ISSN 3165-4436 · Open access, CC BY-SA 4.0

isolate the song effect. The evidence on repetition (Nie et al., 2022; Pavia et al., 2019) suggests that a moderate number of listenings, around three in Nie et al.'s study, may be more efficient than either a single hearing or many hearings. The finding that printed lyrics did not add significantly to learning (Chen et al., 2026) differs from the advantage reported for captions in video (Peters, 2019). One explanation is that seeing lyrics does not guarantee processing them. Tasks that require learners to work with the text, such as those in Huang and Chen's (2024) application-based study, may be needed. This interpretation is tentative and should be tested directly.

Third, singing appears to add value over listening or speaking. The singing advantage appeared with adults learning an unfamiliar language (Ludke et al., 2014), with children (Good et al., 2015), with adolescents (Zhang et al., 2023) and with university EFL learners (Chang, 2024). In several of these studies, however, the outcome measures targeted verbatim memory for phrases, pronunciation or meaning recall of words in taught lyrics, not broader vocabulary growth. The advantage may come from several sources at once: melody and rhythm supporting memory, the active production involved in singing, and greater enjoyment and attention. The included abstracts do not allow these explanations to be separated. Songs also do not seem to be uniquely effective. Words learned from stories were recalled as well as words learned from songs (Leśniewska & Pichette, 2016), and combining songs with physical response was better than either alone (Magnussen & Sukying, 2021).

Fourth, learners' perceptions are strongly positive but realistic. Learners value songs for enjoyment, authenticity and memorability. They also identify concrete obstacles, including fast tempo, slang, figurative language and divided attention (Putri et al., 2026; Rahman et al., 2024). These obstacles match the corpus evidence. Most song vocabulary is high-frequency (Romanko, 2017; Yuanjai & Gampper, 2025), so the difficulty of songs often lies less in rare words than in rapid, connected and sometimes non-standard delivery. Teachers therefore have a selection task: songs need to be clear enough to be decoded and to contain target words worth learning. Teachers also face practical constraints such as copyright and lyric appropriateness (Ghidhaoui, 2025).

The Indonesian literature reviewed here is extensive and closely connected to classroom practice. It also reveals a methodological gap. Short single-site interventions, one-group designs, single-score "vocabulary mastery" results and the absence of delayed post-tests and effect sizes mean that these studies show only that students' scores rose after song-based lessons. They cannot show that songs caused the improvement or that the improvement lasted. The international studies that measure dimensions of word knowledge, repetition, retention and mode of practice provide a model that Indonesian classroom research could adapt. Tilwani et al. (2022) also found a difference between pre-tested and non-pre-tested experimental groups. Researchers should therefore remember that a pre-test can itself alert learners to target words and influence subsequent learning.

e-ISSN 3165-4436 · Open access, CC BY-SA 4.0

5. Implications for Practice and for Future Research

5.1 Implications for Indonesian classrooms (RQ3)

The evidence supports several practical principles for teachers. They are drawn from the studies reviewed above and should be adapted to local conditions.

(1) Select songs deliberately. Choose songs with clear articulation and a moderate tempo, age-appropriate content, and recurring target words or phrases. Teachers can check lyrics against frequency lists or CEFR-based lists, as Yuanjai and Gampper (2025) did, to identify words worth teaching. Avoid songs dominated by slang or dense metaphor for lower-level learners (Lestari & Hardiyanti, 2020; Putri et al., 2026).

(2) Combine listening with explicit vocabulary work. Introduce or highlight target words before or after listening, and use tasks that make learners notice and use them. Explicit focus produced more durable gains than implicit exposure (Mannarelli & Serrano, 2024).

(3) Plan for repetition, but not endless repetition. Listening to a song about three times within a lesson, and returning to it across lessons, is consistent with the evidence (Nie et al., 2022; Pavia et al., 2019).

(4) Have learners sing. Singing along, or practising lyrics through singing, outperformed listening or reciting in several studies (Chang, 2024; Good et al., 2015; Zhang et al., 2023).

(5) Use multimodal support with young learners. Combine songs with pictures, gestures and TPR, and talk about the words (Butler, 2019; Magnussen & Sukying, 2021). Do not expect productive use from song exposure alone (Coyle & Gómez Gracia, 2014).

(6) Connect classroom and out-of-class listening. Learners already use Spotify, TikTok and other platforms (Butar Butar & Katemba, 2023; Isnaeni & Farida, 2026). Teachers can guide this use with strategy training, such as keeping a vocabulary notebook, looking up meanings and choosing songs at an appropriate level (Lestari & Hardiyanti, 2020; Rahman et al., 2024).

(7) Check whether words have been retained. Revisit target words one or more weeks later. Short-term gains do not guarantee retention (Chang, 2024).

5.2 Recommended design for the planned classroom study

The undergraduate proposal that prompted this review asks whether English songs improve students' vocabulary mastery and how students perceive learning vocabulary through songs. Based on the evidence and gaps identified above, the following design is recommended. It is a recommendation only. No data have been collected, and no results are implied.

Design. A quasi-experimental, non-equivalent control group design with a pre-test, an immediate post-test and a delayed post-test, using two intact classes at the same grade level in the same school. Classes should be assigned to conditions by a coin toss or similar procedure where the school allows. If a third class is available, a second experimental condition

e-ISSN 3165-4436 · Open access, CC BY-SA 4.0

(listening plus singing versus listening only) would allow the study to test the singing advantage reported internationally.

Target words. Select about 20 to 30 target words and phrases, including some collocations, that occur in the chosen songs. They should be relevant to the curriculum and unlikely to be known already. Confirm this with a pilot of the pre-test with a similar class.

Treatment. The experimental class learns the target words through a structured song-based cycle over about six to eight meetings. The cycle consists of pre-listening attention to selected words, about three listenings, singing along, and a short explicit vocabulary task. The control class learns the same target words for the same amount of time through the school's usual procedures, without songs. Keeping the target words and time equal is essential so that any difference can be attributed to the use of songs rather than to extra exposure. A simple lesson log should record what was actually done in each meeting (implementation fidelity).

Vocabulary measures. Instead of a single "vocabulary mastery" score, use a test with separate sections for (a) spoken-form recognition (hearing a word and identifying it), (b) form–meaning connection (meaning recognition and, if feasible, meaning recall), (c) collocation recognition, and (d) written-form production. This follows the dimensions measured by Pavia et al. (2019), Tilwani et al. (2022) and Chang (2024). Include some non-target control items. Report the reliability of each section from the pilot.

Timing. Give the pre-test one week before the treatment, the immediate post-test at the end, and the delayed post-test two to four weeks later. Because a pre-test can sensitise learners to target words (Tilwani et al., 2022), its possible influence should be acknowledged and, where possible, reduced, for example by mixing target items with distractors.

Analysis. Compare groups on post-test and delayed post-test scores while controlling for pre-test scores (for example, with an analysis of covariance), or use a mixed-design analysis of variance with time as a within-subject factor. Report effect sizes with confidence intervals, not only p-values. Report each vocabulary dimension separately. Check the statistical assumptions and state the limitations that come from using intact classes. An a priori power analysis will show what size of effect the available sample can detect.

Perception questionnaire. After the immediate post-test, the experimental class completes a short Likert-scale questionnaire. It should cover enjoyment and motivation, perceived usefulness for learning word forms, meanings and collocations, perceived difficulties (tempo, accent, slang, figurative language, distraction), self-reported out-of-class use of songs, and preferences about lyrics and singing. A small number of semi-structured interviews can follow. The items can be adapted from the themes identified in Section 3.9, and they should be piloted and checked for internal consistency.

Ethics. Obtain permission from the school, informed consent from participants (and parental consent for minors), and keep data anonymous. Make sure the control class also receives the song-based materials after the study, so that it is not disadvantaged.

e-ISSN 3165-4436 · Open access, CC BY-SA 4.0

5.3 Directions for future research

Future research should: (a) use control groups and delayed post-tests in Indonesian classrooms; (b) measure separate dimensions of word knowledge, including collocations and productive knowledge; (c) report effect sizes so that later syntheses can pool results; (d) test the singing advantage and the role of printed lyrics in authentic EFL classrooms; (e) compare explicit and implicit approaches to songs over longer periods; (f) follow learners' out-of-class song use longitudinally; (g) study young learners in Indonesian primary schools; and (h) examine Indonesian teachers' beliefs and constraints regarding songs, including copyright and lyric appropriateness. In addition, (i) a systematic review should extract data from full texts rather than abstracts and screen all retrieved records rather than only the top-ranked ones. It should include grey literature such as Indonesian theses and formally assess risk of bias and publication bias. Such a review could also check the abstract-based descriptions given here against the full studies. Pre-registration of hypotheses and analysis plans would further strengthen the evidence base.

6. Limitations

This review has several limitations. First, and most importantly, the data were extracted from published abstracts, not from full texts. Full texts were not systematically read. Details of participants, procedures, instruments, comparison conditions and results may therefore be incomplete or, in places, imprecise. Several features criticised in the Indonesian studies (for example, missing effect sizes or delayed post-tests) are features of the abstracts and may not apply to the full studies. Where an abstract was silent, Table 1 says so instead of inferring the missing information. The conclusions of this review should be regarded as provisional until they are checked against the full texts. Second, it is a structured narrative review, not a systematic review. For each OpenAlex query, only the 12 to 20 top-ranked records were inspected, out of raw hit counts of 479 to 18,386. The selection therefore depended on OpenAlex's relevance ranking. Screening was done by a single reviewer, no formal risk-of-bias assessment was performed, and no meta-analysis was attempted. Relevant studies have therefore almost certainly been missed. Third, several relevant works were excluded because their abstracts could not be read through the services used. Studies with non-registered DOIs were included only when a publisher page could be confirmed. Fourth, the searches were limited to English-language records and to three bibliographic services. Indonesian-language theses and journals not indexed in OpenAlex, ERIC or Crossref are under-represented. Fifth, publication bias probably inflates the consistency of positive findings. The included sources almost exclusively report positive results, grey literature was not searched, and publication bias could not be assessed formally because effect sizes were not pooled. Finally, the review was drafted with the help of an AI writing assistant. All sources were verified against bibliographic databases, and the content was reviewed by the author, but readers should check the original studies before relying on specific details.

e-ISSN 3165-4436 · Open access, CC BY-SA 4.0

7. Conclusion

English songs can help EFL and ESL learners learn vocabulary. The evidence from 2013 to 2026, as reported in the published abstracts, is consistent in direction, although publication bias probably inflates this consistency. It points to several conditions under which songs work best. Gains are most reliable for recognising word forms, form–meaning connections and collocations, and less reliable for productive knowledge. Repetition helps up to a point. Explicit attention to words makes the gains last longer. Singing tends to add value over listening or speaking alone. Learners, including many Indonesian students, value songs for enjoyment, authenticity and autonomy. They also report difficulties with tempo, accent, slang and figurative lyrics that teachers can anticipate. The Indonesian classroom literature is rich in practice but limited in design. For an undergraduate study such as the one that prompted this review, the evidence recommends treating songs as structured, repeated input combined with explicit vocabulary work. It also recommends evaluating them against an equally resourced control class with multi-dimensional and delayed vocabulary measures, and complementing test scores with a carefully designed perception questionnaire. Because the synthesis rests on abstracts and on a relevance-ranked search, these conclusions should be confirmed by a full-text systematic review.

Declaration of Generative AI Use

The draft of this article, including the literature search, the verification of sources against bibliographic databases and the initial writing, was prepared with the assistance of an AI writing assistant (Claude, Anthropic). The author reviewed the draft and takes full responsibility for the content of the published article.

Author Contributions (CRediT)

CRediT) Siti Radiana Sari: Conceptualization, Initial literature search. Final responsibility for the manuscript rests with the author after review and approval.

Conflict of Interest

None declared. Acknowledgements and AI declaration. An initial draft of this manuscript was prepared with generative-AI assistance by Research Center Progresif (Pusat Kajian Progresif) from publicly available data. The author(s) revised it (additional data/cases, re-checked and extended analysis, rewritten discussion, expanded literature and interpretation) and take full responsibility for the data, analysis and conclusions of the final version.

Acknowledgements and AI declaration. An initial draft of this manuscript was prepared with generative-AI assistance by Research Center Progresif (Pusat Kajian Progresif) from publicly available data. The author(s) revised it (additional data/cases, re-checked and extended analysis, rewritten discussion, expanded literature and interpretation) and take full responsibility for the data, analysis and conclusions of the final version.

Declaration of Generative AI Use

During the revision of this manuscript, an AI writing assistant (Claude, Anthropic) provided by the editorial office was used to draft changes responding to the reviewers' comments. The

e-ISSN 3165-4436 · Open access, CC BY-SA 4.0

author(s) reviewed and approved every change and take full responsibility for the content of the article.

References

Amien, J. A. K., Prasetyo, N. A., & Febriyanti, R. H. (2022). Investigating listening to English pop song in augmenting vocabulary mastery: A case study of Indonesian undergraduate students' views. *JEdu: Journal of English Education, 2*(3), 247–254. https://doi.org/10.30998/jedu.v2i3.8099

Basori, B., & Setyowati, I. S. C. (2024). A case study of EFL students' perceptions on the use of English songs in learning English vocabulary. *Journal of English Language Teaching and Learning (JETLE), 5*(2), 107–117. https://doi.org/10.18860/jetle.v5i2.25908

Butar Butar, E. L., & Katemba, C. V. (2023). EFL students' perceptions toward vocabulary learning by listening to English songs on Spotify. *Journey: Journal of English Language and Pedagogy, 6*(1), 194–205. https://doi.org/10.33503/journey.v6i1.2685

Butler, Y. G. (2019). Teaching vocabulary to young second- or foreign-language learners. *Language Teaching for Young Learners*, 4–33. https://doi.org/10.1075/ltyl.00003.but

Chang, A. C.-S. (2024). The effects of varying practice modes on L2 learners' vocabulary retention through songs: Listening, singing, and oral reading. *Teaching English as a Second or Foreign Language–TESL-EJ, 28*(3). https://doi.org/10.55593/ej.28111a9

Chen, S., Weng, F., Wang, L., & Ye, S. X. (2026). The effect of input modality on incidental vocabulary learning through pop songs. *Humanities and Social Sciences Communications, 13*, Article 1868. https://doi.org/10.1057/s41599-026-08087-x

Coyle, Y., & Gómez Gracia, R. (2014). Using songs to enhance L2 vocabulary acquisition in preschool children. *ELT Journal, 68*(3), 276–285. https://doi.org/10.1093/elt/ccu015

Engh, D. (2013). Why use music in English language learning? A survey of the literature. *English Language Teaching, 6*(2). https://doi.org/10.5539/elt.v6n2p113

Fadli, Z., Sembiring, H. M., Isdianto, M. E., & Andayani, W. (2023). Learning while relaxing: Physical students' experiences of learning English vocabulary by songs. *Wacana: Jurnal Penelitian Bahasa, Sastra dan Pengajaran, 21*(2), 137–147. https://doi.org/10.33369/jwacana.v21i2.31178

Ghidhaoui, R. (2025). Music to their ears? Exploring Hungarian EFL teachers' beliefs about the educational potentials of songs. *Hungarian Educational Research Journal, 15*(1), 128–144. https://doi.org/10.1556/063.2024.00314

Good, A. J., Russo, F. A., & Sullivan, J. (2015). The efficacy of singing in foreign-language learning. *Psychology of Music, 43*(5), 627–640. https://doi.org/10.1177/0305735614528833

Gushendra, R. (2017). An experimental study: Improving students' vocabulary mastery by using English songs. *Indonesian Journal of Integrated English Language Teaching, 3*(1). https://doi.org/10.24014/ijielt.v3i1.3968

e-ISSN 3165-4436 · Open access, CC BY-SA 4.0

Huang, Y.-C., & Chen, K. T.-C. (2024). Supporting EFL listening ability and vocabulary acquisition with Lyrics Training app. *International Journal of Instruction, 17*(2), 539–560. https://doi.org/10.29333/iji.2024.17230a

Isnaeni, R. N., & Farida, A. N. (2026). The effect of using English TikTok songs on students' descriptive vocabulary mastery. *Journey: Journal of English Language and Pedagogy, 9*(1), 23–35. https://doi.org/10.33503/journey.v9i1.2925

Karim, S. A., Sudiro, S., Annisa, D. R., Khairunnisa, H. I., & Rahmawati, D. A. (2022). Enhancing vocabulary mastery through English songs: In the eyes of EFL students. *Ethical Lingua: Journal of Language Teaching and Literature*. https://www.ethicallingua.org/index.php/index/article/view/439

Leśniewska, J., & Pichette, F. (2016). Songs vs. stories: Impact of input sources on ESL vocabulary acquisition by preliterate children. *International Journal of Bilingual Education and Bilingualism, 19*(1), 18–34. https://doi.org/10.1080/13670050.2014.960360

Lestari, I. W., & Hardiyanti, N. (2020). Vocabulary learning autonomy through incorporation of English songs: Indonesian EFL students' perspectives. *3L: The Southeast Asian Journal of English Language Studies, 26*(2), 94–104. https://doi.org/10.17576/3l-2020-2602-07

Ludke, K. M. (2018). Singing and arts activities in support of foreign language learning: An exploratory study. *Innovation in Language Learning and Teaching, 12*(4), 371–386. https://doi.org/10.1080/17501229.2016.1253700

Ludke, K. M., Ferreira, F., & Overy, K. (2014). Singing can facilitate foreign language learning. *Memory & Cognition, 42*(1), 41–52. https://doi.org/10.3758/s13421-013-0342-5

Ludke, K. M., & Morgan, K. A. (2022). Pop music in informal foreign language learning. *ITL – International Journal of Applied Linguistics, 173*(2), 251–285. https://doi.org/10.1075/itl.21009.lud

Magnussen, E. S., & Sukying, A. (2021). The impact of songs and TPR on Thai preschoolers' vocabulary acquisition. *THAITESOL Journal, 34*(1), 71–95. https://eric.ed.gov/?id=EJ1304648

Mannarelli, P., & Serrano, R. (2024). 'Thank you for the music': Examining how songs can promote vocabulary learning in an EFL class. *The Language Learning Journal, 52*(1), 1–15. https://doi.org/10.1080/09571736.2022.2092198

Millington, N. T. (2011). Using songs effectively to teach English to young learners. *Language Education in Asia, 2*(1), 134–141. https://doi.org/10.5746/LEiA/11/V2/I1/A11/Millington

Montero Perez, M. (2022). Second or foreign language learning through watching audio-visual input and the role of on-screen text. *Language Teaching, 55*(2), 163–192. https://doi.org/10.1017/S0261444821000501

e-ISSN 3165-4436 · Open access, CC BY-SA 4.0

Nasia P, R. F., Azhar, F., & Prawati, A. (2026). The effect of listening to English songs on vocabulary development of the second year students of SMK Telkom Pekanbaru. *Pendas: Jurnal Ilmiah Pendidikan Dasar, 11*(02), 640–652. https://journal.unpas.ac.id/index.php/pendas/article/view/45314

Nation, I. S. P. (2022). *Learning vocabulary in another language* (3rd ed.). Cambridge University Press. https://doi.org/10.1017/9781009093873

Nie, K., Fu, J., Rehman, H., & Zaigham, G. H. K. (2022). An empirical study of the effects of incidental vocabulary learning through listening to songs. *Frontiers in Psychology, 13*, Article 891146. https://doi.org/10.3389/fpsyg.2022.891146

Pavia, N., Webb, S., & Faez, F. (2019). Incidental vocabulary learning through listening to songs. *Studies in Second Language Acquisition, 41*(4), 745–768. https://doi.org/10.1017/S0272263119000020

Peters, E. (2019). The effect of imagery and on-screen text on foreign language vocabulary learning from audiovisual input. *TESOL Quarterly, 53*(4), 1008–1032. https://doi.org/10.1002/tesq.531

Pravitasari, H., & Rahmah, A. Y. (2024). Impact of English song in vocabulary learning: A pre-experimental study at a junior Islamic boarding school. *Wiralodra English Journal, 8*(1), 54–67. https://doi.org/10.31943/wej.v8i1.262

Putri, A., Prasetyo, Y., & Aisyah, S. (2026). Exploring the potentials and barriers of using English pop songs for vocabulary acquisition among EFL students. *Journey: Journal of English Language and Pedagogy, 9*(1), 12–22. https://doi.org/10.33503/journey.v9i1.2903

Rahman, R. W., Syarif, A. R., & Tohamba, C. P. P. (2024). Enhancing vocabulary through English songs: A self-directed learning experienced for EFL students. *KLASIKAL: Journal of Education, Language Teaching and Science, 6*(3), 864–882. https://doi.org/10.52208/klasikal.v6i3.1236

Rohmah, N., & Indah, R. N. (2021). The use of modified English song to improve vocabulary. *ELTIN Journal: Journal of English Language Teaching in Indonesia, 9*(2), 121–129. https://e-journal.stkipsiliwangi.ac.id/0-last/index.php/eltin/article/view/2537

Romanko, R. (2017). Measuring the vocabulary burden of popular English songs. *Vocabulary Learning and Instruction, 6*(2), 71–78. https://doi.org/10.7820/vli.v06.2.romanko

Sarnou, D. (2023). Teaching Algerian third-year elementary-school pupils English vocabulary through songs: An effective instructional tool to enliven English classes. *Multidisciplinary Journal of School Education, 12*(2[24]), 277–296. https://doi.org/10.35765/mjse.2023.1224.14

Schmitt, N., & Schmitt, D. (2020). *Vocabulary in language teaching* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/9781108569057

Suciati, K., & Zarkasih, E. (2021). Analysis of listening English songs on student's vocabulary learning. *Lingua, 17*(1), 100–114. https://doi.org/10.34005/lingua.v17i1.1370

e-ISSN 3165-4436 · Open access, CC BY-SA 4.0

Sunra, L., Farahdiba, S., & Melinda. (2026). Examining the effects of modified songs on vocabulary mastery and learning motivation: A mixed-methods study of Indonesian EFL learners. *Research and Innovation in Applied Linguistics, 4*(1), 182–200. https://doi.org/10.31963/rial.v4i1.6232

Tegge, F. A. G. (2015). *Investigating song-based language teaching and its effect on lexical learning* [Doctoral dissertation, Victoria University of Wellington]. https://doi.org/10.26686/wgtn.17011901

Tilwani, S. A., Amini MosaAbadi, F., Shafiee, S., & Azizi, Z. (2022). Effects of songs on implicit vocabulary learning: Spoken-form recognition, form-meaning connection, and collocation recognition of Iranian English as a foreign language learners. *Frontiers in Education, 7*, Article 797344. https://doi.org/10.3389/feduc.2022.797344

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

Wardiman, M., Aminah, A., & Dewi, A. K. (2022). The use of English songs to improve students' vocabulary mastery. *e-Journal of ELTS (English Language Teaching Society), 10*(1), 12–21. https://doi.org/10.22487/elts.v10i1.2224

Webb, S., & Nation, P. (2017). *How vocabulary is learned*. Oxford University Press.

Yuanjai, W., & Gampper, C. (2025). A study of Taylor Swift's pop songs for EFL learners: Categorizing words based on the Oxford 3000 and 5000 word lists and CEFR levels. *LEARN Journal: Language Education and Acquisition Research Network, 18*(1), 212–227. https://doi.org/10.70730/txbb1055

Zhang, Y., Baills, F., & Prieto, P. (2023). Singing songs facilitates L2 pronunciation and vocabulary learning: A study with Chinese adolescent ESL learners. *Languages, 8*(3), Article 219. https://doi.org/10.3390/languages8030219

e-ISSN 3165-4436 · Open access, CC BY-SA 4.0