



From the

AERA Online Paper Repository

<http://www.aera.net/repository>

Paper Title The Effectiveness of Using Songs for Teaching Foreign Languages to Young Learners: A Systematic Review

Author(s) Catherine Hamilton, University of Oxford; Johannes Schulz, University of Oxford; Hamish Chalmers, University of Oxford; Victoria Murphy, University of Oxford

Session Title Critical Elements in Language Teaching and Learning (Table 17)

Session Type Roundtable Presentation

Presentation Date 4/14/2023

Presentation Location Chicago, IL

Descriptors ESL/ENL, Meta-Analysis, Music Education

Methodology Mixed Method

Unit SIG-Second Language Research

DOI <https://doi.org/10.3102/2007160>

Each presenter retains copyright on the full-text paper. Repository users should follow legal and ethical practices in their use of repository material; permission to reuse material must be sought from the presenter, who owns copyright. Users should be aware of the [AERA Code of Ethics](#).

Citation of a paper in the repository should take the following form: [Authors.] ([Year, Date of Presentation]). [Paper Title.] Paper presented at the [Year] annual meeting of the American Educational Research Association. Retrieved [Retrieval Date], from the AERA Online Paper Repository.

The Effectiveness of Using Songs for Teaching Foreign Languages to Young Learners: A Systematic Review.

Purpose

With the growing worldwide trend for teaching young children foreign languages (Graddol, 2006; Garton, Copland & Burns, 2011; Holmes & Myles, 2019; Murphy, 2014), teachers of young language learners (YLL) seek effective resources that are age- and stage-appropriate and create a positive classroom atmosphere (Garton et al., 2011). Songs are popular resources with YLL teachers (Garton et al., 2011; Linse, 2006; Şevik, 2011) who express a strong intuition that songs fulfil a multitude of pedagogical purposes including memorisation of vocabulary, improving pronunciation, laying the foundations of grammatical knowledge, supporting classroom routines and behaviour management, and motivating learners (Davanellos, 1999; Forster, 2006; Hamilton & Murphy, 2023; Paquette & Rieg, 2008; Saricoban & Metin, 2000; Schoepp, 2001; Walker, 2006). There is, however, little empirical substance to the justification for using songs with YLL according to recent, non-systematic theoretical and evidence reviews (Davis, 2017; Degrave, 2019; Engh, 2013; Sposet, 2008; Werner, 2020). To map and understand the current research landscape, this systematic review synthesises and critically interprets the empirical research investigating the influence of using songs as pedagogical tools on linguistic outcomes of children (from 2–18 years) learning non-native languages in formal education. We present a narrative synthesis of the 61 studies that investigate linguistic outcomes (namely vocabulary, grammar, speaking, listening, reading and writing), offering an overview of what has been learned thus far and suggesting future areas for exploration.

Theoretical framework

Teachers require little theoretical justification for using songs with YLL because they are ubiquitous in language classrooms (Garton et al., 2011). However, besides the musical and social benefits of singing together, claims are made about the putative linguistic benefits of singing in YLL classrooms (Davanellos, 1999; Fonseca Mora, 2000; Forster, 2006; Paquette & Rieg, 2008) that lean on unsteady theoretical foundations mischaracterised as strong evidence of songs' effectiveness. Recent reviews found scant theoretical grounding for practitioners' belief in songs' effectiveness (Degrave, 2019; Engh, 2013). It is incumbent on researchers to frame theoretically motivated research questions, rather than setting out to prove that songs 'work' based on intuition. We propose that two popular theoretical refrains – (1) the mnemonic 'din' (Krashen, 1983) and 'song stuck in my head' (SSIMH; Murphey,

1990) hypotheses; and (2) aligning pedagogy with students' learning styles and musical intelligence (Fonseca Mora, 2000; Ávila, 2016) – lack substantial and reliable empirical evidence (Coffield, Moseley, Hall & Ecclestone, 2004; Schulkind, 2009; Waterhouse, 2006) and that researchers should seek alternative theoretical foundations.

More promisingly, the prosodic bootstrapping hypothesis, an L1 acquisition theory, presents empirically supported ground for future research into how songs influence L2 acquisition. Prosody represents infants' first encounter with linguistic input *in utero* (see review of evidence in Gervain, Christophe, & Mazuka, 2020). According to the prosodic bootstrapping hypothesis (Gleitman, 1990; Morgan & Demuth, 1996), prosody scaffolds language learning, helping children parse input through the prosody-based prominence of content words (nouns/verbs) and mapping these onto salient visual objects, assisting acquisition of L1 lexical and morphosyntactic features. There is evidence that prosodic bootstrapping plays a role in L2 word-order acquisition with children in classrooms (Campfield & Murphy, 2013), and adults in lab-based studies (Saksida, Flo, Guedes, Nespor & Garay, 2021). Teaching L2 prosody and suprasegmental features explicitly may improve fluency and comprehensibility of L2 learners' speech (Gordon & Darcy, 2016). If, as proposed by Campfield and Murphy (2013), the hypothesis transfers to taught YLL contexts, this would provide a much stronger theoretical motivation for research investigating songs' influence on language learning.

Methods

This review followed PRISMA guidelines (Page et al., 2021) in addressing the research question:

What is the extent and nature of research that has been conducted to explore the effectiveness of using songs for teaching second or foreign languages in compulsory formal education (i.e. with preschool, primary and secondary school-aged learners)?

To be included, papers needed to meet the eligibility criteria shown in Table 1. Journal articles and theses published in any language were included to seek all available evidence that deals with using songs to teach typically developing language learners in primary and secondary school contexts worldwide.

English, French, German and Spanish databases of education, linguistics, psychology, and multidisciplinary research that run Boolean strings were consulted (listed in Table 2). The search string was devised in English to include participant, intervention, comparison, outcome, setting and study design, then translated into French, German and

Spanish (see examples in Table 3). All study designs were included, following Slavin's (1986) best evidence synthesis process for study selection, since discriminating based on research design could reduce the pool of available evidence and bias the review findings in an area where research is already scarce. A data extraction form was created and piloted following the Cochrane Good Practice Guide (Cochrane Effective Practice and Organization of Care, 2017) and Boland, Cherry and Dickson (2017). The Mixed Methods Appraisal Tool (Hong et al., 2018; Pluye, Gagnon, Griffiths & Johnson-Lafleur, 2009) was used to evaluate risk of bias (RoB) of included studies.

Included studies are grouped by outcome measure and presented in a narrative synthesis, since a meta-analysis of such a diverse group of studies with incomparable participants, methodologies and outcomes would be inappropriate (Petticrew & Roberts, 2006). Screening, data extraction, and appraisal processes were corroborated at each stage by the second author independently screening 10% of the titles and abstracts, completed data extraction forms, and completed appraisals, assuring that data was screened, documented, and appraised accurately.

Findings

As Figure 1 illustrates, 2868 records were identified, including 1007 duplicates. After screening, 61 empirical studies published in 23 countries between 1978 and 2021 were included in this review (see summary Table 4). 47 studies (77%) were published since 2009 (Figure 2) and 51 (85%) in English. 42 studies (69%) are peer-reviewed journal articles, and 19 are theses. 36 studies (59%) were conducted in primary schools, with the remaining studies split almost equally between 12 preschool and 13 secondary school contexts (Figure 3). Three studies reported exclusively qualitative data, 11 used mixed methods designs, and 47 reported exclusively quantitative findings. Median sample size was 53 (range 5 to 573) shown in Figure 4. Random allocation of participants to conditions was rare: only 10 studies reported random allocation, with four reporting the strategy used. 26 studies (43%) allocated intact classes, while 24 (39%) did not report any allocation strategy. 47 studies had control groups, and 14 had none. Studies ranged in duration from one hour to four years, with 50% lasting between two and nine weeks (Figure 5). Figure 6 illustrates the outcome types reported by included studies: vocabulary, grammar, four skills (listening, reading, writing, speaking), attitudes, or other outcomes. Figure 7 shows the total studies reporting each outcome type. The following presents key findings from studies measuring vocabulary, grammar and skills outcomes.

Vocabulary

The largest group of studies (n=35) report vocabulary outcomes, as summarised in Table 5 (13 receptive vocabulary studies) and Table 6 (15 productive vocabulary studies). 10 studies do not report how vocabulary was measured. Four comparable studies using standardised measures of receptive vocabulary report mixed findings. Two claim that songs perform worse than stories or combined songs and stories (Albaladejo, Coyle and Larios, 2018; Leśniewska and Pichette, 2014) and one (Schunk, 1999) found a positive effect of singing paired with signing, but that this was not significantly different to speaking paired with signing, and neither signing condition was significantly better than singing by itself. Chen (2011) found a positive effect for vocabulary learning with 9–10-year-olds in the song condition compared to a control group taught using traditional methods. Due to methodological flaws such as not having clearly stated research questions and not accounting for confounding variables such as cognitive ability, the first three studies provide somewhat unreliable evidence about songs' influence on learning receptive vocabulary. 15 studies reporting productive vocabulary measures found predominantly positive results, but only one study received a low RoB appraisal (Busse, Hennies, Kreutz and Roden, 2021) and claimed no effect of singing on productive vocabulary other than improving spelling. Given these varied designs, mixed findings and low cumulative reliability, no overall picture of songs' influence on vocabulary acquisition emerges.

Grammar

Of the 15 studies reporting grammar outcomes for primary and secondary pupils (Table 7), 7 used measures of verb learning that varied from completing cloze sentences by choosing the correct multiple-choice verb form (receptive knowledge, e.g. Busse et al., 2021) or providing the correct verb form themselves (productive knowledge, e.g. Ludke, 2010). Siebring (2004) found no significant effect of singing on improving primary children's fossilised verb errors. Klohs (1994) did not isolate songs as a variable, instead investigating several mnemonic strategies simultaneously. The overall picture is thus unclear about songs' influence on verb learning.

Two studies investigated how songs influence learning L2 word order. The most promising, Campfield and Murphy (2013), investigated 83 Polish primary EFL learners' ability to judge English word order correctly when exposed to nursery rhymes, prose, or control conditions. They found a significant effect of rhythmically salient input (nursery rhymes) on pupils' scores in grammaticality judgement tasks, particularly for verb-last structures. Since this study was well conducted and rigorous, accounting for multiple confounding factors, it provides more trustworthy evidence than the second study, Lowe

(1995), which claimed a positive effect of songs on French immersion learners' understanding of French word order. Lowe (1995) hindered the generalisability of the findings by allocating intact classes to conditions, hence $n=2$ classes not $n=53$ individuals as reported.

Since these studies focus on different aspects of grammar in different ways, some with high RoB, few conclusions can be drawn beyond the studies themselves about songs' influence on grammar learning.

Four skills

A similarly mixed and unreliable picture emerges from studies measuring speaking (17), listening (11), reading (10), and writing (3) outcomes (see Tables 8–11 for studies which report measures clearly). No overall conclusions can be drawn from any subsection of included studies, despite predominantly positive claims made by their authors.

Cumulative weight of evidence

Most studies found a positive effect of singing on their included measures, but the cumulative weight of evidence is weak, as illustrated in Figure 8. Positive effects are noted on vocabulary (receptive and productive), grammar, speaking, and listening measures from studies with low RoB ratings, but also some neutral effects for grammar and productive vocabulary. No overall conclusions can be drawn about the effectiveness of using songs to teach second or foreign languages to YLL in compulsory formal education. It is clear from these results, however, that this research field generally lacks methodological rigour.

Significance

Despite this somewhat bleak outlook, the appetite for research into songs' putative educational benefits for YLL is clear from the number of included studies and positive bias in the literature: practitioners' intuition that songs 'work' is undeniably strong. This review reveals that the field is ripe for a renewed approach that takes strong theoretically driven research questions forward, rather than circling around the same 'starter' questions. We suggest that the prosodic bootstrapping hypothesis is a more promising SLA theoretical foundation than the unsubstantiated din, 'song stuck in my head', learning styles, and multiple intelligences literatures. Studies need to build upon each other, and this can only happen when methodologies, analyses and findings are reported transparently and rigorously. The current research landscape is marred by opaque and unreliable reporting.

This lack of rigour does a disservice both to teachers, whose valued practices should be underpinned by high quality research, and to learners, who should be provided with the

best teaching methods guided by the experiential wisdom of their teachers *and* empirical evidence gathered through well-conducted, truthfully reported research. As Paran (2017) argues, intuition and research are not competing foundations for teaching practice, but basing teaching practice on unanalysed intuition rather than evolving experiential and empirical knowledge may result in teaching becoming “merely the transmission of self-perpetuating, unsupported beliefs and prejudices, based on experience that is never examined” (Paran, 2017:506). Currently, there is no consensus on how or even whether songs influence YLL linguistic outcomes. This could have real consequences for YLL and is something researchers and teachers can address together through carefully designed and conducted classroom research.

REFERENCES

(* indicates paper included in systematic review)

- *Albaladejo, S. A., Coyle, Y., & Larios, J. R. de. (2018). Songs, stories, and vocabulary acquisition in preschool learners of English as a foreign language. *System*, 76, 116–128.
- *Alinte, C. (2013). Teaching Grammar through Music. *The Journal of Linguistic and Intercultural Education*, 6, 7–28. <https://doi.org/10.29302/jolie.2013.6.1>
- *Allen-Tamai, M. (2000). *Phonological Awareness and Reading Development of Young Japanese Learners of English*. Doctoral thesis, Temple University.
- *Alley, D.C. (1988). *The role of music in the teaching of listening comprehension in Spanish*. Doctoral thesis, University of Georgia.
- *Al-Mosawi, F.R.A.H, (2018). Finger Family Collection YouTube Videos Nursery Rhymes Impact on Iraqi EFL Pupils' Performance in Speaking Skills. *Opción*, Año 34, Especial No.17(2018), 452–474.
- *Amiri, M. & Sobouti, F. (2016). The effect of using short stories and songs on the second language achievement of Iranian young learners. *Modern Journal of Language Teaching Methods (MJLTM)*, 6(5), 401–412.
- *An, G-H. 안근행 (2009). The effects of teaching English song through Korean song in vocabulary acquisition and affective attitude, 우리 동요를 활용한 영어 노래 지도가 어휘력과 정의적 태도에 미치는 영향. *Journal of the Korea English Education Society*, *영어교과교육*, 8(1), 37–57.
- *Au, T. K. (2013). Songs as Ambient Language Input in Phonology Acquisition. *Language Learning and Development*, 9(3), 266–277. <https://doi.org/10.1080/15475441.2013.753819>
- *Augustine, C. (2015). How the use of music and movement impacts the learning of reading skills by preschoolers. *Malaysian Music Journal*, 4(2), 74–90.
- Ávila, J. (2016). Musical training and foreign language learning. In M.C. Fonseca-Mora & M. Gant (Eds.). *Melodies, Rhythm and Cognition in Foreign Language Learning*. pp.57–68. Newcastle: Cambridge Scholars Publishing.
- *Becerra Vera, B. & Luna, R.M. (2013). Teaching English through music: a proposal of multimodal learning activities for primary school children. *Encuentro*, 22, 16–28.
- *Boey, L.K. (1978). The Unified Language Project. *RELC Journal*, 9(1), 19–27.
- Boland, A., Cherry, M. G., & Dickson, R. (2017). *Doing a systematic review: A student's guide* (2nd edition.). London: SAGE.
- *Busse, V., Hennies, C., Kreutz, G., & Roden, I. (2021). Learning grammar through singing? An intervention with EFL primary school learners. *Learning and Instruction*, 71, 101372.
- *Caleya, M.F, Nieto, M. & Espejo, A. (2013). Music, poetry and fun activities in English teaching: an early childhood education experience. *EDULEARN13: 5th International Conference on Education and New Learning Technologies*, 0(0), 1473–1481.

- *Campfield, D. E., & Murphy, V.A. (2013). The influence of prosodic input in the second language classroom: does it stimulate child acquisition of word order and function words? *The Language Learning Journal*, 45(1), 81–99.
- *Chae, Y., & Yoon, E. 채영신; 윤은자 (2013). The Effects of the Songs of Children's Literature on the Primary School Students' Long-term Memory, Grammar Learning, and Affective Domains, 영어동화노래수업이 장·단기 기억과 문법습득 및 정의적 영역에 미치는 효과. *Primary English Education, 초등영어교육*, 19(2), 241–270.
- *Cheippe, E. (2012). *La voie musicale pour remédier aux difficultés de prononciation des voyelles de l'allemand dans des textes lus: expérimentation dans une classe bilingue: analyse acoustique*. Doctoral thesis, Université de Strasbourg.
- *Chen, J-J. (2011). *The effects of music activities on English pronunciation and vocabulary retention of fourth-grade ESOL (English for Speakers of Other Languages) students in Taiwan*. Doctoral thesis, University of Florida.
- *Chiang, M. (2003). *The effect of chanting activities on the comprehension of English of first graders and college freshmen in Taipei*. Doctoral thesis, Texas A&M University-Kingsville.
- *Chou, M. (2014). Assessing English vocabulary and enhancing young English as a Foreign Language (EFL) learners' motivation through games, songs, and stories. *Education 3–13*, 42(3), 284–297. <https://doi.org/10.1080/03004279.2012.680899>
- Cochrane Effective Practice and Organisation of Care (EPOC). (2017). Data collection form. Retrieved December 2021 from https://epoc.cochrane.org/sites/epoc.cochrane.org/files/public/uploads/Resources-for-authors2017/good_practice_data_extraction_form.doc
- Coffield, F., Moseley, D., Hall, E., & Ecclestone, K. (2004). *Learning styles and pedagogy in post-16 learning. A systematic review*. London: Learning Skills and Research Centre, Department for Education.
- *Coyle, Y., & Gracia, R. G. (2014). Using songs to enhance L2 vocabulary acquisition in preschool children. *ELT Journal*, 68(3), 276–285. <https://doi.org/10.1093/elt/ccu015>
- *Cruz-Cruz, M.L. (2005). *The effects of selected music and songs on teaching grammar and vocabulary to second grade English language learners*. Doctoral thesis, Texas A&M University, Kingsville.
- Davanellos, A. (1999). Songs. *English Teaching Professional*, 13, 13–15.
- Davis, G.M. (2017). Songs in the Young Learner Classroom: A Critical Review of Evidence. *ELT Journal*, 71(4), 445–455.
- *Davis, G.M. & Fan, W. (2016). English Vocabulary Acquisition Through Songs in Chinese Kindergarten Students. *Chinese Journal of Applied Linguistics*, 39(1): 59–71.
- Degrave, P. (2019). Music in the Foreign Language Classroom: How and Why? *Journal of Language Teaching and Research*, 10(3), 412–420.

- *Diakou, M. (2014). *Using Songs to Enhance Language Learning and Skills in the Cypriot Primary MFL Classroom*. EdD thesis. The Open University.
- *Dominguez, D. (1991). *Developing language through a musical program and its effect on the reading achievement of Spanish-speaking migrant children*. Doctoral thesis, Western Michigan University.
- Engh, D. (2013). Why use music in English language learning? A survey of the literature. *English Language Teaching*, 6(2), 113–27.
- Fonseca Mora, C. (2000). Foreign language acquisition and melody singing. *ELT Journal*, 54(2), 146–152. <http://dx.doi.org/10.1093/elt/54.2.146>
- *Fonseca-Mora, M. C., Jara-Jiménez, P., & Gómez-Domínguez, M. (2015). Musical plus phonological input for young foreign language readers. *Frontiers in Psychology*, vol. 6, article 286. <https://doi.org/10.3389/fpsyg.2015.00286>
- Forster, E. (2006). The value of songs and chants for young learners. *Encuentro*, 16, 63–68.
- Garton, S., Copland, F., & Burns, A. (2011). Investigating Global Practices in Teaching English to Young Learners. *ELT Research Papers*, 11(01). London: British Council.
- *Geisler, P. (2008). *Musikorientiertes Lernen im Englisch-Unterricht der Grundschule*. Doctoral thesis, Pädagogische Hochschule Freiburg.
- Gervain, J., Christophe, A., & Mazuka, R. (2020). Prosodic Bootstrapping. In C. Gussenhoven & A. Chen (Eds.) *The Oxford Handbook of Language Prosody*. pp.563–573. Oxford, UK: Oxford University Press.
- Gleitman, L.R. (1990). The structural sources of verb meanings. *Language Acquisition*, 1, 3–55.
- *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>
- Gordon, J. & Darcy, I. (2016). The development of comprehensible speech in L2 learners. *Journal of Second Language Pronunciation*, 2(1), 56–92.
- *Gorjian, B., Hayati, A., & Barazandeh, E. (2012). An evaluation of the effects of art on vocabulary learning through multi-sensory modalities. *Procedia Technology*, 1, 345–350. <https://doi.org/10.1016/j.protcy.2012.02.072>
- Graddol, D. (2006) *English Next*. UK: British Council.
- *Haghverdi, H. R. (2015). The Effect of Song and Movie on High School Students Language Achievement in Dehdasht. *Procedia – Social and Behavioral Sciences*, 192, 313–320. <https://doi.org/10.1016/j.sbspro.2015.06.045>
- *Hakozaki, Y. & Nakagawa, Y. (2020). Teaching stress-timed rhythm of English at the Japanese elementary school level: focusing on the effects of using chants. *Asian EFL Journal Research Articles*, 27(2), 173–201.
- Hamilton, C., & Murphy, V. A. (2023). Folk pedagogy? Investigating how and why UK early years and primary teachers use songs with young learners. *Education 3-13, ahead-of-print*(ahead-of-print), 1–22. doi: 10.1080/03004279.2023.2168132

- *Herrera, L., Lorenzo, O., Defior, S., Fernandez-Smith, G., & Costa-Giomi, E. (2011). Effects of phonological and musical training on the reading readiness of native- and foreign-Spanish-speaking children. *Psychology of Music*, 39(1), 68–81.
<https://doi.org/10.1177/0305735610361995>
- Holmes, B., & Myles, F. (2019). *White Paper: Primary Languages Policy in England – The Way Forward*. RiPL: www.ripl.uk/policy/. Retrieved 23 June 2022 from <https://www.ripl.uk/wp-content/uploads/2019/02/RIPL-White-Paper-Primary-Languages-Policy-in-England.pdf>.
- Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M.-P., Griffiths, F., Nicolau, B., O’Cathain, A., Rousseau, M.-C., & Vedel, I. (2018) *Mixed Methods Appraisal Tool (MMAT), version 2018*. Registration of Copyright (#1148552), Canadian Intellectual Property Office, Industry Canada. Retrieved May 2022 from http://mixedmethodsappraisaltoolpublic.pbworks.com/w/file/attach/127916259/MMAT_2018_criteria-manual_2018-08-01_ENG.pdf.
- *Hsu, H. (2009). *The effect of rhythmic teaching methods for kindergarten EFL students in Taiwan*. Doctoral thesis, University of Mississippi.
- *Jarvis, S. (2013). How effective is it to teach a foreign language in the Foundation Stage through songs and rhymes? *Education 3-13*, 41(1), 47–54.
<https://doi.org/10.1080/03004279.2012.710099>
- *Jeong, Y.-J., & Kim, J.-O. 김정옥 (2014). A Study of English-Teaching Model through Stories and Songs, 영어동화와 노래를 결합한 정의적 영어수업모형의 적용 효과. *Wonkwang Journal of Humanities*, *열린정신 인문학 연구*], 15(2), 57–75.
- *Kaminski, A. (2016). *The Use of Singing, Storytelling and Chanting in the Primary EFL Classroom: Aesthetic Experience and Participation in FL Learning*. Doctoral thesis, Swansea.
<https://doi.org/10.23889/suthesis.54359>
- *Kim, J.-S., & Kang, M.-K. (2015). The Effects of Improving English Listening Skills of High School Students with a Lower Level through Pop Song Hunminglish Pronunciation (PSHP) Practice. *Advanced Science and Technology Letters*, 92(Education 2015), 41–45.
- *Kim, Y. & Park, J.-E. 김양희. (2012). Analysis on English vocabulary acquisition by accomplishment levels with an integrated teaching model for English and music through songs, 노래를 활용한 영어 · 음악 통합 수업에서 성취 수준별 영어 어휘 습득 분석. *Primary English Education*, "초등영어교육, 18(3), 31–63.
- *Klohs, L.M. (1994). *Use of mnemonic strategies to facilitate written production of a second language by high school French students*. Doctoral thesis, University of Minnesota.
- Krashen, S. (1983). The Din in the Head, Input, and the Language Acquisition Device. *Foreign Language Annals*, 16(1), 41–44.
- *LeBrun, C. (2019). *The Effects of Music-Infused Instruction on Student Achievement in Secondary school Spanish*. Doctoral thesis, University of South Dakota.

- *Legg, R. (2009). Using Music to Accelerate Language Learning: An Experimental Study. *Research in Education*, 82(1), 1–12. <https://doi.org/10.7227/rie.82.1>
- *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.
- Linse, C. (2006). Using favorite songs and poems with young learners. *English Teaching Forum*, 44(2), 38–42.
- *Lowe, A.S. (1995). *The effect of the incorporation of music learning into the second-language classroom on the mutual reinforcement of music and language*. Doctoral thesis, University of Illinois at Urbana-Champaign.
- *Ludke, K. (2010). *Songs and singing in foreign language learning*. Doctoral thesis, University of Edinburgh.
- *Luo, S. (2019). Influence of Singing English Songs on Vocabulary Learning by Senior School Students in Guangzhou. *International Journal of Information and Education Technology*, 9(11), 843–848.
- *Ma, S. (2004). English Education Activities and English Story Recall Using Story Songs., 이야기노래 (story songs) 를 활용한 영어교육활동과 유아의 영어이야기회상. *Early Childhood Education Research & Review*, 유아교육학논집, 8(2), 57–75.
- *Madani, D., & Nasrabadi, M. M. (2016). The effect of songs on vocabulary retention of preschool young English language learners. *International Journal of Research Studies in Language Learning*, 6(3), 63–72. <https://doi.org/10.5861/ijrsl.2016.1562>
- *Mamdouh, M. (2017). La canción francófona, una herramienta eficaz en el proceso de enseñanza-aprendizaje de la lengua francesa. Thélème. *Revista Complutense de Estudios Franceses*, 32(2), 221–238. <https://doi.org/10.5209/thel.54572>
- *McCormack, B. & Klopper, C. (2016). The potential of music in promoting oracy in students with English as an additional language. *International Journal of Music Education*, 34(4), 416–432.
- *McCormack, B. A., Klopper, C., Kitson, L., & Westerveld, M. (2018). The potential for music to develop pronunciation in students with English as an Additional Language or Dialect (EAL/D). *Australian Journal of Music Education*, 52(1), 43–50.
- *Medina, S.L. (1991). *The effect of a musical medium on the vocabulary acquisition of limited English speakers*. Doctoral thesis, University of Southern California.
- *Moradi, F., & Shahroki, M. (2014). The effect of listening to music on children's segmental and suprasegmental pronunciation. *English Language Teaching*, 7(6), 128–142.
- Morgan, J.L. & Demuth, K. (1996). *Signal to Syntax: Bootstrapping from Speech to Grammar in Early Acquisition*. New Jersey, USA: Lawrence Erlbaum Associates.
- Murphy, V.A. (2014). *Second Language Learning in the Early School Years: Trends and Contexts*. Oxford: Oxford University Press.

- Murphey, T. (1990). The Song Stuck in My Head Phenomenon: A Melodic Din in the LAD? *System*, 18(1), 53–64. [http://dx.doi.org/10.1016/0346-251X\(90\)90028-4](http://dx.doi.org/10.1016/0346-251X(90)90028-4)
- *Muzammil, L., & Andy, A. (2019). Can Young Learners Utilize Cartoon Picture and Song To Learn?: A teaching model. *Proceedings of the 3rd Asian Education Symposium (AES 2018)*, 512–517. <https://doi.org/10.2991/aes-18.2019.115>
- *Navarro, K.S., Quiroga, C., & Diaz, C. (2018). English pronunciation for first year primary school students: a didactic sequence implementation for its improvement. *Revista Comunicación, Año 39, 27(1)*, 108–121.
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372(71).
- Paquette, K. R., & Rieg, S. A. (2008). Using music to support the literacy development of young English language learners. *Early Childhood Education Journal*, 36, 227–232. <https://doi.org/10.1007/s10643-008-0277-9>
- Paran, A. (2017). 'Only connect': researchers and teachers in dialogue. *ELT Journal*, 71(4), 499–508. <https://doi.org/10.1093/elt/ccx033>
- Pluye, P., Gagnon, M.P., Griffiths, F., & Johnson-Lafleur, J. (2009). A scoring system for appraising mixed methods research, and concomitantly appraising qualitative, quantitative and mixed methods primary studies in Mixed Studies Reviews. *International Journal of Nursing Studies*, 46(4), 529–46.
- Petticrew, M. & Roberts, H. (2006). *Systematic Reviews in the Social Sciences*. Oxford, UK: Blackwells.
- *Priester, M. (2011). *Using Song Lyrics in the Preschool ESL Classroom to Assist Students' English Vocabulary Retention and Use*. Master's Thesis, Caldwell College.
- Saksida, A., Flo, A., Guedes, B., Nespore, M., & Garay, M. P. (2021). Prosody facilitates learning the word order in a new language. *Cognition*, 213, 104686.
- *Santos Jimenez, O.C., Gallegos Ruiz, A., & Gomez Hermosa, C. (2017). Application of children's songs for the learning of the French language vocabulary in children of the initial level of the San Antonio de Padua Educational Institution, Chosica, Lima, Perú, 2016. *Revista Inclusiones*, 4(4), 189–204.
- Saricoban, A. & Metin, E. (2000). Songs, verse and games for teaching grammar. *The Internet TESOL Journal*, 6(10).
- Schoepp, K. (2001). Reasons for using songs in the ESL/EFL classroom. *The Internet TESOL Journal*, 7(2).
- Schulkind, M. D. (2009). Is memory for music special? *Annals of the New York Academy of Sciences*, 1169(1), 216–224.

- *Schunk, H. A. (1999). The Effect of Singing Paired with Signing on Receptive Vocabulary Skills of Elementary ESL Students. *Journal of Music Therapy*, 36(2), 110–124.
<https://doi.org/10.1093/jmt/36.2.110>
- Şevik, M. (2011). Teacher views about using songs in teaching English to young learners. *Educational Research and Review*, 6(21), 1027–1035.
- *Siebring, M.F. (2004). *The effectiveness of a systematic approach based on songs to prevent and correct errors in elementary core French*. Master's thesis, Nipissing University.
- Slavin, R. (1986). Best evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher*, 15, 5–11.
- Sposet, B. (2008). *The role of music in second language acquisition: a bibliographical review of seventy years of research, 1937–2007*. NY, USA: The Edwin Mellen Press.
- *Tomczak, E. & Lew, R. (2019). "The Song of Words": Teaching multi-word units with songs. *The Southeast Asian Journal of English Language Studies*, 25(4), 16–33.
- Walker, R. (2006). Going for a Song. *English Teaching Professional*, 43, 19–21.
- *Wang, Y. (2005). *A study of the effects of teaching English grammar with English songs in junior high schools*. Master's thesis, Beijing Normal University.
- Waterhouse, L. (2006). Multiple Intelligences, the Mozart Effect, and Emotional Intelligence: A Critical Review. *Educational Psychologist*, 41(4), 207–225.
- Werner, V. (2020). "Song-Advantage" or "Cost of Singing"? A Research Synthesis of Classroom-based Intervention Studies Applying Lyrics-based Language Teaching (1972–2019). *Journal of Second Language Teaching and Research*, 8(1), 138–170.
- *Yousefi, A. (2014). The Effect of Modern Lyrical Music on Second Language Vocabulary Acquisition. *Mediterranean Journal of Social Sciences*, 5(23), 2583–2586.
<https://doi.org/10.5901/mjss.2014.v5n23p2583>
- *Zhaku-Kondri, B. (2014). Using Song Lyrics in the Classroom: Assessing the Utility of Song Lyrics for the Acquisition of a Foreign Language. Proceedings of INTCESS14 – International Conference on Education and Social Sciences Proceedings. *California Folklore Quarterly*, 1(4), 1201–1210.

TABLES AND FIGURES

Table 1 Eligibility criteria

Item	Inclusion criterion	Rationale
Bibliographic information	<p>Include 1: Studies with a full reference or sufficient information.</p> <p>Exclude 1: Studies with insufficient bibliographic information.</p>	Without sufficient bibliographic information, retrieval of works is unfeasible.
Date of publication	Include 2: Published on any date.	Attempting to collect the totality of available evidence.
Participants	<p>Include 3: Studies on typically developing foreign language learners. Include studies even if no explicit reference is made to learning ability if reasonable assumption can be made that participants are comprised mainly of typically developing individuals.</p> <p>Exclude 3: Studies that exclusively target non-typically developing learners or learners with Developmental Language Disorder.</p>	This review seeks to assess effectiveness of songs as a pedagogical tool in typically developing school populations. The findings for non-typically developing populations may not generalise to a larger population, thus such results will not be extrapolated or included in this review.
	<p>Include 4: Studies conducted in early years, primary or secondary schools (students aged 2-18) or other formal settings (e.g. playgroups, after-school clubs) worldwide.</p> <p>Exclude 4: Studies conducted in university, or adult educational contexts; informal settings (e.g. at home).</p>	This study focuses on the outcomes of using songs for learners in formal contexts between age 2 and 18, since adult learners (over 18) have different learning capacities and educational goals. Findings from studies conducted in informal settings may not generalise to formal educational settings, thus such results will not be extrapolated or included in this review.
Intervention	<p>Include 5: Studies where singing songs, choral chanting, or nursery rhymes are included as a whole-class or group activity.</p> <p>Exclude 5: Studies where musical instruments are the intervention focus, not singing, chanting or nursery rhymes.</p>	This review focuses on the linguistic outcomes of using songs as pedagogical tools, thus the intervention must include songs with words, not purely an instrumental intervention (e.g. whole-class ukulele lessons).
Outcomes	<p>Include 6: Primary research studies reporting any measure of language acquisition including but not limited to vocabulary, grammar or phonology outcome measures. Include studies that report either quantitative or qualitative measures of outcomes.</p> <p>Exclude 6: Systematic reviews or studies that provide only narrative</p>	A synthesis of empirical findings in this field of literature is impossible without the reporting and evaluation of concrete data.

	<p>evaluation of an intervention but do not include outcome measures of language acquisition including vocabulary, grammar or phonology; studies that measure only non-language outcomes, e.g. satisfaction, happiness, engagement.</p> <p>Include 7: All types of study design.</p> <p>Exclude 7: Do not exclude studies based on research design.</p>	<p>Given the scarcity of research in this area, excluding one type of study design risks systematically neglecting a body of research and narrowing the scope of this review.</p>
Publication status	<p>Include 8: Grey literature.</p> <p>Exclude 8: Do not exclude studies based on publication status.</p>	<p>This paper seeks to offset potential publication bias by including a wider range of research, including grey literature.</p>
Language of publication	<p>Include 9: Studies published in any language.</p> <p>Exclude 9: Do not exclude studies based on the language of publication.</p>	<p>Limiting this review to studies published in English may result in a systematic neglect of a particular body of research.</p>

Table 2 List of databases searched

Discipline	Database			
	English	German	French	Spanish
Education	ProQuest Education Collection (including ERIC), British Education Index EBSCO; Education Abstracts (H.W. Wilson)	Fachportal Pädagogik	n/a	n/a
Linguistics	ProQuest Linguistics Collection (including LLBA); MLA International Bibliography	n/a	n/a	n/a
Psychology	PsychInfo	PsynDEX	n/a	n/a
Multidisciplinary	Web of Science, Scopus	Humboldt University Berlin; Center for Research Libraries Global Resources Network (CRL)	Cairn.info ; SUDOC ; Pascal-Francis ; CRL	CRL
Grey literature	ProQuest Dissertations & Theses Global; OpenGREY; EthOS	n/a	theses.fr	TESEO educacion.gob.es

Table 3 Example search strings

LANGUAGE	DATABASE	SEARCH STRING
ENGLISH	ProQuest Education	ab(MFL OR EAL OR ESL OR EFL OR "foreign language*" OR FL OR "second language*" OR L2 OR French OR German OR Spanish OR English OR TEFL OR TESOL) AND ab(KS1 OR KS2 OR KS3 OR KS4 OR "key stage" OR EYFS OR "early years" OR preschool OR kindergarten OR infant* OR junior* OR primary OR secondary OR elementary OR child* OR adolescent* OR "high school") AND ab("nursery rhyme*" OR choral OR chant* OR song* OR music* OR sing*) AND ab(vocabulary OR grammar* OR phonolog* OR acquisition OR speaking OR spoken OR proficiency OR competence or skill*) NOT ab(singapore OR single* OR singular)
FRENCH	Pascal-Francis	(FLE OR anglais OR "langue étrangère" OR français OR FLS OR "langue seconde" OR allemand OR espagnol OR "langue* moderne*") AND (jeune* OR maternelle OR primaire OR collège OR élémentaire OR enfan* OR adolescent OR lycée) AND (vocabulaire OR grammaire OR phonologie OR acquisition OR compétence) AND (comptine* OR choral* OR chant OR chanson* OR chanter OR musique OR musical*)
GERMAN	Fachportal Pädagogik	(Titel: DAZ oder DAZ oder DAF oder DAF oder L2 oder SLA oder TEFL oder TESOL oder TESL oder ENGLISCH oder FRANZOESISCH oder SPANISCH oder FREMDSPRACH* oder ZWEITSPRACH* oder ZWEISPRACHIG) und (Schlagwörter: LERNER oder GRUNDSCHULE oder KIND* oder JUGENDLICH* oder GYMNASI* oder REALSCHULE oder GANZTAGSSCHULE oder GESAMTSCHULE oder HAUPTSCHULE oder FOERDERSCHULE oder SCHUELER*) und (Freitext: LIED* oder REIM oder GESANG oder SING* oder SPRECHCHOR oder SONG oder MUSIK oder RHYTHMUS oder RHYTHMISCH oder MELODIE oder MUSIKALISCH oder MELODISCH)) und (Freitext: VOKABEL* oder GRAMMATIK oder PHONOLOGIE oder ERWERB oder LERN*) und nicht (Freitext: SINGAPUR oder SINGLE)
SPANISH	TESEO educacion.gob.es	("idioma adicional" O "lengua inglesa" O "idioma extranjero" O "lengua* extranjera*" O "segunda lengua" O "segundo idioma" O francés O "lengua castellana" O español O inglés O "lenguas modernas" O "lenguas vivas" O "idiomas modernos") Y (guardería O "jardín de infancia" O "escuela infantil" O "escuela preescolar" O "escuela secundaria" O instituto* O "escuela de primaria" O "enseñanza primaria" O "escuela elemental" O "ciclo primario" O niño* O estudiante*) Y (rimas infantiles O coral O canto* O canción* O música* O cantar) Y (vocabulario O gramática O fonologi* O adquisición O "habilidades lingüísticas" O "conocimientos lingüísticos")

Table 4 Study characteristics

JA = journal article, PhD = doctoral thesis, MSc = master's thesis. Study duration (in weeks unless stated otherwise).
MM = Mixed methods, Quant = quantitative data, Qual = qualitative data

Study	Publication status	Study design	Data type	Country	Sample size	Setting	Study duration	Non-song control group?	General outcomes	Specific outcome measures
1. Albaladejo, Coyle & Larios (2018)	JA	Within-subjects experimental	MM	Spain	17	Preschool	6	No	Vocabulary	PPVT; observation of behaviour
2. Alinte (2013)	JA	Quasi-experimental	Quant	Romania	34	Secondary	15	Yes	Grammar; attitudes	Grammatical knowledge test
3. Allen-Tamai (2000)	PhD	Quasi-experimental, counterbalanced design	Quant	Japan	62	Preschool	11	No	Phonological awareness	Rhyme awareness
4. Alley (1988)	PhD	Quasi-experimental	Quant	USA	47	Secondary	5	Yes	Listening; attitudes	Listening tests; attitudes to presentation mode
5. Al-Mosawi (2018)	JA	Quasi-experimental	Quant	Iraq	40	Primary	12	Yes	Four skills	Four skills: reading, writing, listening and speaking
6. Amiri & Soubouti (2016)	JA	Quasi-experimental	Quant	Iran	60	Preschool	8	Yes	Speaking	Pronunciation, fluency, grammar and vocabulary
7. An (2009)	JA	Quasi-experimental	Quant	Korea	79	Primary	4	Yes	Vocabulary; attitudes	Vocabulary listening, comprehension of vocabulary meaning, speaking skills; attitude towards learning English
8. Au (2013)	JA	Quasi-experimental	Quant	Hong Kong	126	Primary	18	No	Speaking	L2 or 2 nd dialect accent
9. Augustine (2015)	JA	Quasi-experimental	Quant	Malaysia	40	Preschool	6	Yes	Reading	Print knowledge, definitional vocabulary, phonological awareness

Study	Publication status	Study design	Data type	Country	Sample size	Setting	Study duration	Non-song control group?	General outcomes	Specific outcome measures
10. Becerra Vera & Luna (2013)	JA	Observational	Quant	Spain	49	Primary	1 school year	No	Listening	Listening tests
11. Boey (1978)	JA	Quasi-experimental	Quant	Malaysia	573	Primary	2 school years	Yes	Four skills	Speaking, listening, reading, dictation
12. Busse, Hennies, Kreutz & Roden (2021)	JA	Quasi-experimental	Quant	Germany	57	Primary	9	Yes	Vocabulary; grammar; attitudes	Vocabulary recall (name items); grammar translation; multiple choice grammaticality judgement task; affective outcomes of lessons
13. Caleyá, Nieto & Espejo (2013)	JA	Observational	Quant	Spain	193	Primary	1 school year	No	Speaking	Pronunciation, accuracy, fluency, eagerness to repeat, accent, memorising
14. Campfield & Murphey (2013)	JA	Experimental	Quant	Poland	87	Primary	3	Yes	Grammar	L2 word order; knowledge of function words
15. Chae & Yoon (2013)	JA	Quasi-experimental	MM	Korea	60	Primary	12	Yes	Memory; grammar; affective domains	Short/long-term memory (cloze tests); grammar; affective responses to input (story or song) and interest in learning English
16. Cheippe (2012)	PhD	Experimental	Quant	France	20	Primary	7	Yes	Speaking	Pronunciation (L2 vowels)
17. Chen (2011)	PhD	Quasi-experimental	Quant	Taiwan	128	Primary	12	Yes	Vocabulary; speaking; attitudes	Picture vocabulary test; phonemic analysis test; attitudes to music intervention

Study	Publication status	Study design	Data type	Country	Sample size	Setting	Study duration	Non-song control group?	General outcomes	Specific outcome measures
18. Chiang (2003)	PhD	Quasi-experimental	Quant	Taiwan	120	Primary	18	Yes	Listening	Multiple choice listening comprehension & dictation
19. Chou (2014)	JA	Quasi-experimental	MM	Taiwan	72	Primary	5x 100-minute lessons	No	Vocabulary; attitudes	Written receptive vocabulary recognition (true/false, matching) and spelling/productive vocabulary writing (anagrams/gap-filling with pictures)
20. Coyle & Gracia (2014)	JA	Pre-test, post-test intervention (no control)	Quant	Spain	25	Primary	7	No	Vocabulary; attitudes	Receptive (picture recognition) and productive (naming task) vocabulary tests
21. Cruz-Cruz (2005)	PhD	Quasi-experimental	Quant	USA	28	Primary	6	Yes	Vocabulary; grammar	Grammar (productive/judgement): pronouns, pronoun-verb agreement, adjectives, adverbs, articles; vocabulary: circle correct word to complete sentence; definition-word matching
22. Davis & Fan (2016)	JA	Within-subjects quasi-experimental	Quant	China	64	Preschool	7	No	Vocabulary; grammar; attitudes	MLU of productive description of picture card prompts

Study	Publication status	Study design	Data type	Country	Sample size	Setting	Study duration	Non-song control group?	General outcomes	Specific outcome measures
23. Diakou (2014)	PhD	Action research	MM	Cyprus	171	Primary	2	No	Vocabulary; grammar	Pre-post questionnaires assessing participants' vocabulary/grammar attitudes ; focus groups discussing acquisition ; video observations tracing acquisition.
24. Dominguez (1991)	PhD	Post-test only, control group design	Quant	USA	51	Primary	7	Yes	Reading	Basic reading skills (e.g., word recognition, digraphs, end sounds, letter sounds, referents, drawing conclusions, predicting outcomes, etc.)
25. Fonseca-Mora, Jara-Jiménez & Gómez-Domínguez (2015)	JA	Quasi-experimental	Quant	Spain	63	Primary	11	Yes	Reading	Early grade reading assessment: letter name knowledge, oral reading fluency, initial sound identification
26. Geisler (2008)	PhD	Action research/case study	Qual	Germany	Y1: 35 Y2: 37 Y3: 40 Y4: 40	Primary	4 school years	No	Vocabulary, phonetics, intonation, linking	Observation and interpretation, interviews, parent and teacher questionnaires
27. Good, Russo & Sullivan (2015)	JA	Quasi-experimental	Quant	Ecuador	38	Primary	2 weeks (with follow-up test after 6 months)	Yes	Speaking; vocabulary	Pronunciation (vowel & consonant production); recall words/phrases from lyrics; translate English vocabulary into Spanish

Study	Publication status	Study design	Data type	Country	Sample size	Setting	Study duration	Non-song control group?	General outcomes	Specific outcome measures
28. Gorjian, Hayati & Barazandeh (2012)	JA	Experimental	Quant	Iran	56	Primary	3 months	Yes	Vocabulary	Researcher designed vocabulary test with 14 items
29. Haghverdi (2015)	JA	Unclear	Quant	Iran	60	Secondary	8	Yes	Listening; vocabulary/ grammar; reading; attitudes	Listening; vocabulary/ grammar; reading (not defined further)
30. Hakozaiki & Nakagawa (2020)	JA	Pre-test, post-test intervention (no control)	Quant	Japan	91	Primary	6	No	Speaking	Pronunciation, overall intelligibility
31. Herrera, Lorenzo, Defior, Fernandez-Smith & Costa-Giomi (2011)	JA	Experimental	Quant	Spain	97	Preschool	2	Yes	Phonological awareness	Phonetic awareness, verbal memory, naming speed, name and sound letters knowledge
32. Hsu (2009)	PhD	Quasi-experimental	Quant	Taiwan	47	Preschool	6–8	Yes	Vocabulary; speaking	Pronunciation and oral spelling of colours
33. Jarvis (2013)	JA	Action research	MM	UK	12	Primary	Not reported	Yes	Speaking; listening; attitudes	Speaking assessment of weekly target vocabulary; observation of behaviour; attitudes of staff to introducing MFL in EY setting
34. Jeong & Kim (2014)	JA	Unclear	Quant	Korea	40	Primary	2 months	Yes	Listening; vocabulary; attitudes	Listening; vocabulary; attitudes to learning English
35. Kaminski (2016)	PhD	Case study	MM	Germany	21	Primary	3 school years	No	Speaking; attitudes	Traceback of lexical items produced in interviews to intervention input

Study	Publication status	Study design	Data type	Country	Sample size	Setting	Study duration	Non-song control group?	General outcomes	Specific outcome measures
36. Kim & Kang (2015)	JA	Unclear	Quant	Korea	128	Secondary	10 months	No	Listening; attitudes	National listening comprehension tests
37. Kim & Park (2012)	JA	Unclear	MM	Korea	87	Primary	3 months	Yes	Vocabulary	Vocabulary proficiency test
38. Klohs (1994)	PhD	Experimental	MM	USA	72	Secondary	4.5	Yes	Grammar; writing; attitudes	Verb tenses; written paragraph assessed for communicative skills; attitudes to mnemonic skills taught/perceived vs actual usage of mnemonics in the tests
39. LeBrun (2019)	PhD	Quasi-experimental	Quant	USA	142	Secondary	15 lessons	Yes	Vocabulary; grammar; reading; listening; attitudes	Vocabulary: matching/cloze/multiple choice Grammar: cloze sentence to fill with correct verb conjugation Reading/listening comprehension
40. Legg (2009)	JA	Experimental	Quant	UK	62	Secondary	1 hour	Yes	Vocabulary	Translate English phrases containing passé 22erano22/imperfect verbs into French equivalent; translate weekdays
41. Leśniewska & Pichette (2014)	JA	Within-subjects experimental	Quant	Canada	24	Preschool	4	No	Vocabulary	PPVT

Study	Publication status	Study design	Data type	Country	Sample size	Setting	Study duration	Non-song control group?	General outcomes	Specific outcome measures
42. Lowe (1995)	PhD	Quasi-experimental	Quant	Canada	53	Primary	5 months	Yes	Vocabulary; grammar; reading; speaking; music skills	Vocabulary: cloze/matching; oral grammar (put words in correct order); reading: true/false, gap-filling; pronunciation; music skills – describe, create, perform
43. Ludke (2010)	PhD	Quasi-experimental	MM	UK	59	Secondary	4	Yes	Vocabulary; grammar; attitudes	Cloze test of song lyrics; translation French > English
44. Luo (2019)	JA	Unclear	MM	China	50	Secondary	3	Yes	Vocabulary; attitudes	Use target words in a sentence; Chinese > English word translation
45. Ma (2004)	JA	Quasi-experimental	Quant	Korea	48	Preschool	4	No	Vocabulary; story recall	Picture vocabulary test: point (receptive) and label (productive); child prompted to complete sentences by reading/singing along with story
46. Madani & Nasrabadi (2016)	JA	Experimental	Quant	Iran	112	Preschool	1 month	Yes	Vocabulary	Vocabulary learning/retention
47. Mamdouh (2017)	JA	Quasi-experimental	Quant	Spain	19	Secondary	10	Yes	Listening	Listening comprehension
48. McCormack & Klopper (2016)	JA	Action research	Qual	Australia	5	Primary	6	No	Speaking	Graphic melodic contouring to measure oral fluency

Study	Publication status	Study design	Data type	Country	Sample size	Setting	Study duration	Non-song control group?	General outcomes	Specific outcome measures
49. McCormack, Klopper & Westerveld (2018)	JA	Single-subject experimental design	MM	Australia	6	Primary	8	No	Speaking	Pronunciation
50. Medina (1991)	PhD	Experimental	Quant	USA	48	Primary	6	Yes	Vocabulary	Picture vocabulary test: circle item that matches the word read aloud
*61. Moradi & Shahroki (2014)	JA	Experimental	Quant	Iran	30	Primary	5	Yes	Speaking	Pronunciation, intonation, stress patterns
51. Muzammil & Andy (2019)	JA	Unclear	Quant	Indonesia	31	Preschool	Not reported	Not reported	Vocabulary; speaking; phrases	Receptive/productive vocabulary; phrases: matching
52. Navarro, Quiroga & Diaz (2018)	JA	Action research	Quant	Chile	25	Primary	5	No	Speaking	Pronunciation: words, phrases and sentences
53. Priester (2011)	MSc	Quasi-experimental	MM	USA	15	Preschool	5	No	Vocabulary	Oral productive task and journal pictures
54. Santos Jimenez, Gallegos Ruiz & Gomez Hermosa (2017)	JA	Quasi-experimental	Quant	Peru	48	Primary	Not reported	Yes	Vocabulary	Measures unclear
55. Schunk (1999)	JA	Experimental	Quant	USA	80	Primary	1–2	Yes	Vocabulary	PPVT
56. Siebring (2004)	MSc	Quasi-experimental	Quant	Canada	53	Primary	2	No	Grammar	Fossilised errors tested orally – complete sentence/respond to question with correct form

* reference identified through citation chaining

Study	Publication status	Study design	Data type	Country	Sample size	Setting	Study duration	Non-song control group?	General outcomes	Specific outcome measures
57. Tomczak & Lew (2019)	JA	Unclear	Quant	Poland	31	Secondary	3 per study (x2)	Yes	Vocabulary	Multi-word unit productive knowledge
58. Wang (2005)	MSc	Quasi-experimental	Quant	China	133	Secondary	4.5 months	Yes	Grammar; attitudes	Formative grammar, summative grammar and listening comprehension tests
59. Yousefi (2014)	JA	Experimental	Quant	Iran	60	Secondary	2 months and 11 days	Yes	Vocabulary	Provide L1 equivalent of English vocabulary item
60. Zhaku-Kondri (2014)	JA	Quasi-experimental	Quant	Macedonia	57	Primary	8	Yes	Vocabulary; grammar; attitudes	Grammar (verb tenses) in pre/post-tests; vocabulary in post-test

Table 5 Studies reporting receptive vocabulary measures

	Study ID	Receptive Vocabulary Measure	Claim made by authors about findings	MMAT trustworthiness rating
P R E S C H O O L	9. Augustine (2015)	Definitional vocabulary	Claim positive effect for songs for definitional vocabulary	MODERATE
	45. Ma (2004)	Picture vocabulary test*: point (receptive) and label (productive); child prompted to complete sentences by reading/singing along with story	Claim positive effect of singing on word recognition/labelling	WEAK
	1. Albaladejo, Coyle & Larios (2018)	PPVT	Claim positive effect for songs but that songs alone performed worse than story or song/story combination	WEAK
	41. Leśniewska & Pichette (2014)	PPVT	Claim singing presentation condition worse than story or story/singing combined	MODERATE
P R I M A R Y	17. Chen (2011)	Picture vocabulary test from standardised Test of Language Development-Primary, 3 rd Edition (TOLD:P-3: Newcomer & Hammill, 1997)	Positive effect claimed for vocabulary learning and pronunciation in song condition compared to control group with traditional methods	STRONG
	19. Chou (2014)	Written receptive vocabulary recognition (true/false, matching)	Claims positive combined effect of songs, games and stories on vocabulary learning	WEAK
	20. Coyle & Gracia (2014)	Receptive (picture recognition*)	Claims positive effect of songs on receptive vocabulary	WEAK
	21. Cruz-Cruz (2005)	Circle correct word to complete sentence; definition-word matching	Claims positive effect of songs on vocabulary and grammar	WEAK
	50. Medina (1991)	Picture vocabulary test: circle item that matches the word read aloud by tester	Claims positive effect of music for low proficiency learners but since it is not significant at 0.5 level unclear why they claim this. No other positive effects. Very small samples of ¼ per group	MODERATE
	55. Schunk (1999)	PPVT	Claims positive effect of sung condition and spoken condition with signs compared to spoken text only. Sung/spoken with signs not significantly different to singing-only condition.	WEAK
	42. Lowe (1995)	Vocabulary: cloze/matching	Claims positive effect of music programme on composite French post-test but no Group X Time interaction effects for vocabulary alone	MODERATE
S E C O N D A R Y	39. LeBrun (2019)	Vocabulary: matching/cloze/multiple choice	Claims significant differences in vocab scores for junior high group for experimental group compared to control, but not when all groups added together	MODERATE
	59. Yousefi (2014)	Provide L1 equivalent of English vocabulary item	Claims positive effect of music on short and long-term retention of vocabulary	WEAK

*unclear whether standardised PPVT was used

Green = positive, Orange = mixed, Red = negative

Table 6 Studies reporting productive vocabulary measures

	Study ID	Productive Vocabulary Measure	Claims made by authors about findings	MMAT trustworthiness rating
P R E S C H O O L	22. Davis & Fan (2016)	MLU of productive description of picture card prompts	Claim singing/chanting equally effective compared to no presentation control	MODERATE
	32. Hsu (2009)	Pronunciation and oral spelling of colours	Claims positive effect of singing on oral vocabulary and spelling of target words	WEAK
	45. Ma (2004)	Picture vocabulary test*: point (receptive) and label (productive); child prompted to complete sentences by reading/singing along with story	Claims positive effect of singing on oral target word & phrase recall	WEAK
	53. Priester (2011)	Oral productive task and journal pictures	Claims positive effect of singing on oral vocabulary and use of target words when drawing in journals	WEAK
P R I M A R Y	19. Chou (2014)	Spelling/productive vocabulary: writing (anagrams/gap-filling with pictures)	Claims positive combined effect of songs, games and stories on vocabulary learning	WEAK
	23. Diakou (2014)	Pre/post questionnaires assessing participants' vocabulary ; focus groups discussing acquisition ; video observations tracing acquisition.	Claims positive effect of introducing songs on pupil interest/motivation, which in turn has positive effect on vocabulary uptake	WEAK
	26. Geisler (2008)	Observation and interpretation, interviews, parent and teacher questionnaires	Claims positive effect of songs on L2 development (including productive vocabulary)	MODERATE
	27. Good, Russo & Sullivan (2015)	Pronunciation (vowel & consonant production); recall words/phrases from lyrics; translate English vocabulary into Spanish	Claim positive effect of songs on vocabulary recall	WEAK
	35. Kaminski (2016)	Group interviews (n=21) of 28–36 minutes during which the researcher analysed the nature of the participants' FL output, whether they used lexical items they had learnt in connection with songs, chants and stories, and whether any traces of out-of-school exposure to the FL could be found.	Claims traces of songs and chants in learners' FL output, including chunks that were used in an appropriate context and to fulfil a communicative purpose 12 months after having been exposed to the songs in class. Less successful in reproducing language from texts (e.g. illustrated stories) than songs/chants.	STRONG
	12. Busse, Hennies, Kreutz & Roden (2021)	Vocabulary recall (name items)	No effect of singing claimed, except for spelling	STRONG
	20. Coyle & Gracia (2014)	Productive naming task	Claim non-significant effect of singing on productive vocabulary	WEAK
S E C O N D A R Y	40. Legg (2009)	Translate English phrases containing passé 27erano27/imperfect verbs into French equivalent; translate weekdays	Claims positive effect of music condition on learning song words & Eng>Fre translation	WEAK
	43. Ludke (2010)	Cloze test of song lyrics; translation French > English	Claims positive effect of singing on French language skills compared to visual art/drama	MODERATE
	44. Luo (2019)	Use target words in a sentence; Chinese > English word translation	Claims positive effect of singing on vocabulary learning	WEAK
	57. Tomczak & Lew (2019)	Multi-word unit productive knowledge	Claims positive effect of songs for learning MWU	MODERATE

Green = positive, Orange = mixed, Red = negative

Table 7 Studies reporting grammar measures

	Study ID	Grammar Measure	Claims made by authors about findings	MMAT trustworthiness rating
P R E S C H O O L	6. Amiri & Soubouti (2016)	Combined pronunciation, fluency, grammar and vocabulary in "YLE" (Young Learner English) test	Claim large effect of singing on grammar test	WEAK
	12. Busse, Hennies, Kreutz & Roden (2021)	6 question-answer pairs presented in English (3 from songs, 3 new): participants choose correct form of verb 'to do' in multiple choice.	Claim students in the singing group identified correct form of verb "to do" better than speaking/control group when sentences were already provided, with progress retained over retention period.	STRONG
P R I M A R Y	14. Campfield & 4 th author (2013)	L2 word order (70 sentences) and knowledge of function words (64 sentence pairs) tested with grammaticality judgement tasks.	Claim significant effect of song input on GJT for word order (particularly verb-last structures).	STRONG
		L2 word order and knowledge of function words tested with grammaticality judgement tasks.	No effect detected for function-words.	STRONG
	21. Cruz-Cruz (2005)	Grammar (10 questions in 6 sections) included productive (choosing the correct pronoun), judgement task (which agreement is correct?), cloze with articles provided to fill in a/an, 'spot the adjective' sentence, knowing if an adverb is of time or manner.	Claim experimental group outperforms control on grammar post test	WEAK
	23. Diakou (2014)	Questionnaire and focus group questions about how songs help them learn grammar	Claims songs helped students memorise grammar structures	WEAK
	42. Lowe (1995)	Oral grammar: students asked to rearrange words to form a sentence (5 items) and read it aloud. Words in the incorrect order lost a mark.	Claims significant difference in favour of treatment group for oral grammar post-test, when achievement in French and maths taken as covariates.	MODERATE
	56. Siebring (2004)	Oral interviews targeting fossilised verb error structures	No significant effect detected of treatment on improving fossilised verb errors	MODERATE
	60. Zhaku-Kondri (2014)	Target verb tenses – I would/Would I?/I wouldn't – but unclear how exactly these are tested	Claims significant effect of using song lyrics on grammar test score, helping pupils practise the grammar and understand spoken and written English	WEAK
S E C O N D A R Y	38. Klohs (1994)	Change French sentences into past tense, then write justification in English of chosen tense	Claims significant effect of mnemonic strategies on learning grammar	MODERATE
	39. LeBrun (2019)	Cloze sentences: fill in blank with correct form of verb. Write a response to the question in Spanish.	No effect detected in singing condition. Significant difference in favour of control group (total participants – all ages added together)	MODERATE

	43. Ludke (2010)	Translate 5 sentences Fre>Eng from song and 5 from dialogue with "acceptable" scores used as basis for statistical analysis when Eng meaning was close to correct Fre meaning (e.g. only one incorrect verb tense or form).	Both age groups improved grammar from pre- to mid-point test, but only older age group improved from mid- to post-test (and younger group's score decreased)	MODERATE
	58. Wang (2005)	Form-changing and picture-writing test of 3 English verb tenses (unclear what this means in practice)	Claim experimental group is more competent in using target grammatical rules, as shown by them scoring significantly higher on form-changing and picture-writing (but not multiple choice) tests	WEAK

Green = positive, Orange = mixed, Red = negative

Table 8 Studies reporting speaking measures

	Study ID	Speaking Measure	Claims made by authors about findings	MMAT trustworthiness rating
P R E S C H O O L	6. Amiri & Soubouti (2016)	Combined pronunciation, fluency, grammar and vocabulary in “YLE” (Young Learner English) test.	Claim that all four subskills of speaking (pronunciation, fluency, grammar and vocabulary) were statistically significantly improved in the song group, compared to the control.	WEAK
	32. Hsu (2009)	Oral test of colours (can the child recall and pronounce the colour that corresponds to the colour card and “what colour is this?” question) and give the oral spelling. 1 point for correct pronunciation, 1 point for correct spelling.	Claims rhythmic teaching methods help EFL kindergarteners acquire target vocabulary pronunciation and spelling.	WEAK
P R I M A R Y	8. Au (2013)	Participants read two illustrated stories aloud (one English, one Putonghua) after hearing NS of each language read story aloud. Accents rated on five-point scale by three NS of each language.	Claim significant positive effect on pronunciation of ambient Putonghua music on Cantonese L1 second-dialect learners of Putonghua. No measurable benefits detected for English songs on L2 pronunciation not closely related to L1.	MODERATE
	16. Chieppe (2012)	Participants read text aloud and recordings are transcribed, with target German vowels/diphthongs rated by NS for NS norm pronunciation.	Claims improvement of singing groups on target German vowel and diphthong sounds.	WEAK
	17. Chen (2011)	Phonemic analysis test from TOLD-P: 3. 14 items measured children’s pronunciation of phonemes and their ability to break down spoken words into shorter phonemic portions.	Claims students’ pronunciation gain scores were statistically and significantly affected by music treatment, even when taking current private music lessons into account as a covariate.	STRONG
	27. Good, Russo & Sullivan (2015)	Pronunciation of vowels tested with support of lyrics handout: children asked to reproduce the lyrics (not specified whether to sing or speak them). 15 target vowels/consonants rated 1 for correct pronunciation (i.e. English not Spanish norms).	Claim sung condition better than spoken condition for teaching vowel sounds, but no significant difference in pronunciation of consonant sounds.	WEAK
	30. Hakozaki & Nakagawa (2020)	Participants read a familiar text aloud. Segmental features, sentence level stress, and overall intelligibility all scored on a scale of 1–5 (1 = poor, 5 = high) by three native English speakers.	Claims that chants had a significant effect on intelligibility of English pronunciation by helping Japanese EFL learners focus on prosodic features of English.	MODERATE
	42. Lowe (1995)	Read five sentences aloud. Pronunciation scored on a five-point scale by five French immersion teachers. Average pre- and post-test scores for each student are used in analyses.	No effect of music condition found for pronunciation measure alone, but overall composite French score was significantly different for treatment group.	WEAK
	48. McCormack & Klopper (2016)	L2 oracy progress measured with graphic contouring (visual representation) of pronunciation of a marker sentence once a week for six weeks.	Claim increased oracy and fluency in all six students.	WEAK
	49. McCormack, Klopper & Westerveld (2018)	Weekly speech samples collected and analysed using the Student Oral Language Observation Matrix [SOLOM] (California Department of Education, 1981), and the EAL/D Rating Scales designed by the research team.	5/6 EAL/D participants’ English pronunciation improved, 1 decreased according to both measures. Although students’ native accent was retained, their speech was more coherent post-intervention in comparison to their pre-intervention results.	WEAK

	61. Moradi & Shahroki (2014)	Post-test of pronunciation, intonation, stress recognition (each marked out of 10). Recordings of post-test compared with original song input pronunciation.	Claim positive effect of treatment on pronunciation (segmental), and intonation and stress recognition (suprasegmental articulation).	WEAK
	52. Navarro, Quiroga & Diaz (2018)	English pronunciation evaluated at the level of words, phrases, and sentences. Repeat words after hearing recording (1); choose three objects and describe them (2); do an oral presentation (3). Marked according to whether Adequate, sufficient, or insufficient for 1 & 2; or Excellent, good, sufficient, and insufficient (3).	Claim positive effect of treatment on students' pronunciation. None remained in 'insufficient' grading after interventions.	WEAK

Green = positive, Orange = mixed, Red = negative

Table 9 Studies reporting listening measures

	Study ID	Listening Measure	Claims made by authors about findings	MMAT trustworthiness rating
S E C O N D A R Y	4. Alley (1988)	Weekly unit test where text was spoken/sung to match treatment conditions. Comprehensive end-of-treatment exam testing all content, through narrative or dialogue only (no sung presentation). No report of the actual test content.	No significant differences between either song or listening skills (active control) treatment groups on weekly unit tests or post-test. Treatment groups scored significantly higher than no treatment groups in post-test. Inconclusive: both treatment groups did better than groups with no focus on listening skills.	WEAK
	39. LeBrun (2019)	Test from the textbook/course. Listening comprehension of 3-minute Spanish audio recording with 10 yes/no questions to check understanding.	No significant differences detected between treatment and control groups (composite group or within age categories).	MODERATE

Green = positive, Orange = mixed, Red = negative

Table 10 Studies reporting reading measures

	Study ID	Reading Measure	Claims made by authors about findings	MMAT trustworthiness rating
P R E S C H O O L	3. Allen-Tamai (2000)	Rhyme awareness tested by children raising a pink flag if a word rhymes, or green if it does not when told a word and asked which of the two words read aloud (supported with visuals) shares the same end sound (hold up pink or green flag for each). 30 questions with rhyming words taken from two taught nursery rhymes with implicit (nursery rhyme) or explicit (rhyming word game) conditions. Tests video recorded and researcher noted children's responses afterwards.	Claims no significant differences in mean scores between groups: children improved their rhyme awareness regardless of type of instruction. Children acquired rhyme knowledge equally well from explicit (rhyming games) or implicit (nursery rhyme) conditions, thus author claims nursery rhymes are useful as semantic material for developing L2 rhyme awareness.	MODERATE
	9. Augustine (2015)	Print knowledge, definitional vocabulary, phonological awareness tested with TOPEL (Test of Preschool Early Literacy).	Claims positive effect of music treatment on overall reading scores: significant differences on print knowledge and definitional vocabulary, but not phonological awareness.	MODERATE
	31. Herrera, Lorenzo, Defior, Fernandez-Smith & Costa-Giomi (2011)	Phonetic awareness, verbal memory, naming speed, name and sound letters knowledge	<p>Rhyme oddity task: both treatment groups outperformed the control group, with musical treatment significantly outperforming non-musical phonological training ($p < .05$) regardless of L1 status.</p> <p>Syllabic tapping and initial phoneme oddity task: both treatment groups outperformed controls at post-test, but it does not report if the treatment groups' mean scores were significantly different from each other.</p> <p>Naming task: treatment groups outperformed controls. Tamazight (L2 Spanish) learners in the music group significantly outperformed Tamazight learners in the control group.</p>	WEAK
P R I M A R Y	24. Dominguez (1991)	Basic reading skills (e.g., word recognition, digraphs, end sounds, letter sounds, referents, drawing conclusions, predicting outcomes, etc.)	Only the word recognition test (1/15 tests) had a significant difference in mean scores between the treatment and control groups.	MODERATE
	25. Fonseca-Mora, Jara-Jiménez & Gómez-Domínguez (2015)	Early grade reading assessment (EGRA): letter name knowledge, oral reading fluency, initial sound identification.	Claim that performance of the phonological training and phonological training with music groups increased significantly compared to control group for correct letter names test, but not for correct words read in a dialogue or initial sound identification tests.	WEAK

	39. LeBrun (2019)	¿Así se dice! End-of-unit test: read two paragraphs about the weather unit – 10 points for reading comprehension questions.	No significant between-groups differences in mean reading scores.	MODERATE
S E C O N D A R Y	42. Lowe (1995)	Reading comprehension – the comprehension section of the test consisted of a short text to read, after which students were asked to answer five items as “true” or “false” and five items which required them to fill in a blank.	Claims a significant effect of music treatment on reading comprehension: the experimental group made more progress than control group from pre- to post-test, when maths and French prior achievement are covariates.	MODERATE

Green = positive, Orange = mixed, Red = negative

Table 11 Studies reporting writing measures

	Study ID	Writing Measure	Claims made by authors about findings	MMAT trustworthiness rating
P R I M A R Y	11. Boey (1978)	10 marks for sentence dictation as part of end-of-year assessment.	No significant difference between experimental and control groups in their English dictation	WEAK
S E C O N D A R Y	38. Klohs (1994)	Write one paragraph (scored out of 15) about an event from the weekend or from childhood. Include negatives, questions, and sentences about other people. Marked according to Semke’s Communicative Rating Scale of 1 (unintelligible) to 5 (Mostly intelligible). Three French NS hired to rate the essay task.	Only predictor of success in the essay task was the previous quarter grade, not treatment condition, according to the stepwise regression model used.	STRONG

Green = positive, Orange = mixed, Red = negative

Figure 1 PRISMA 2020 flow diagram of study selection process

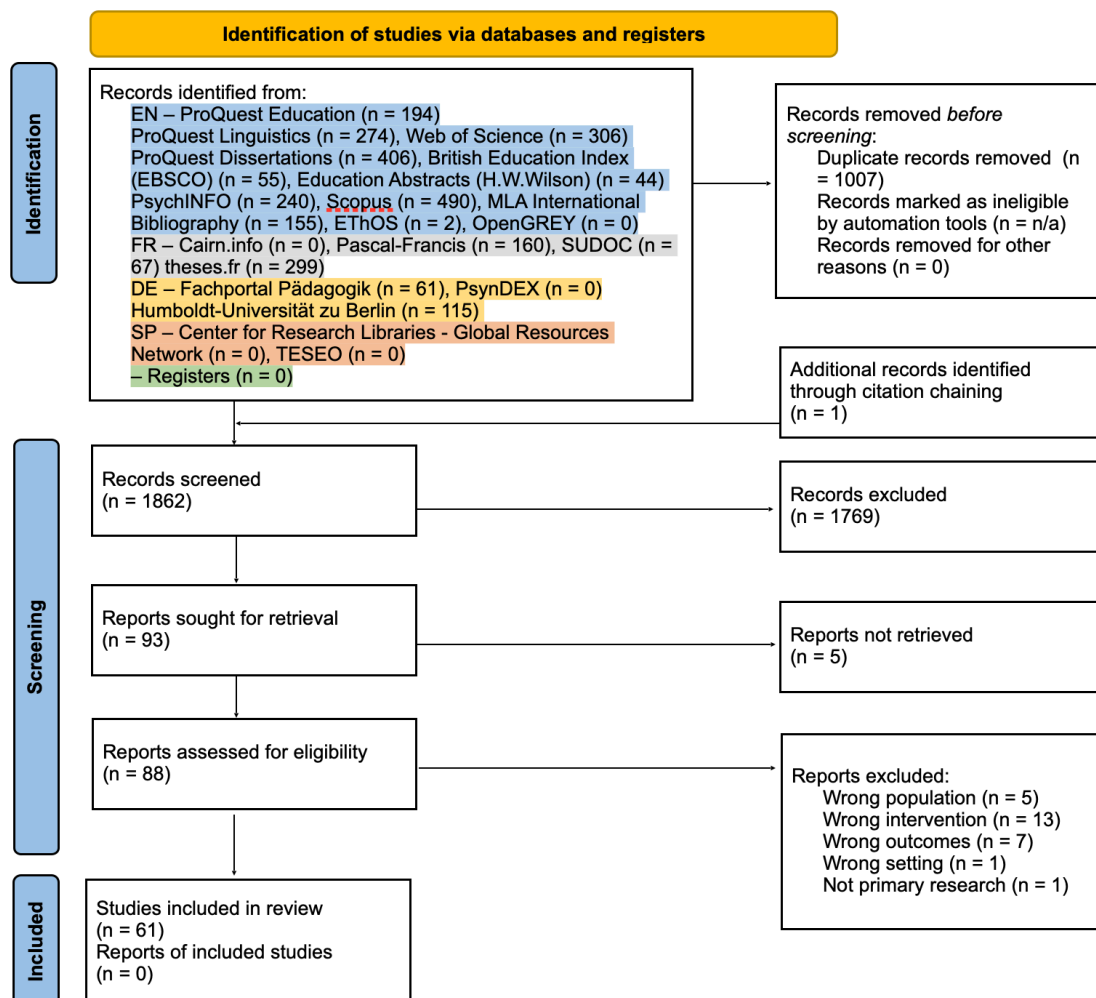


Figure 2 Number of included studies by publication year and educational context

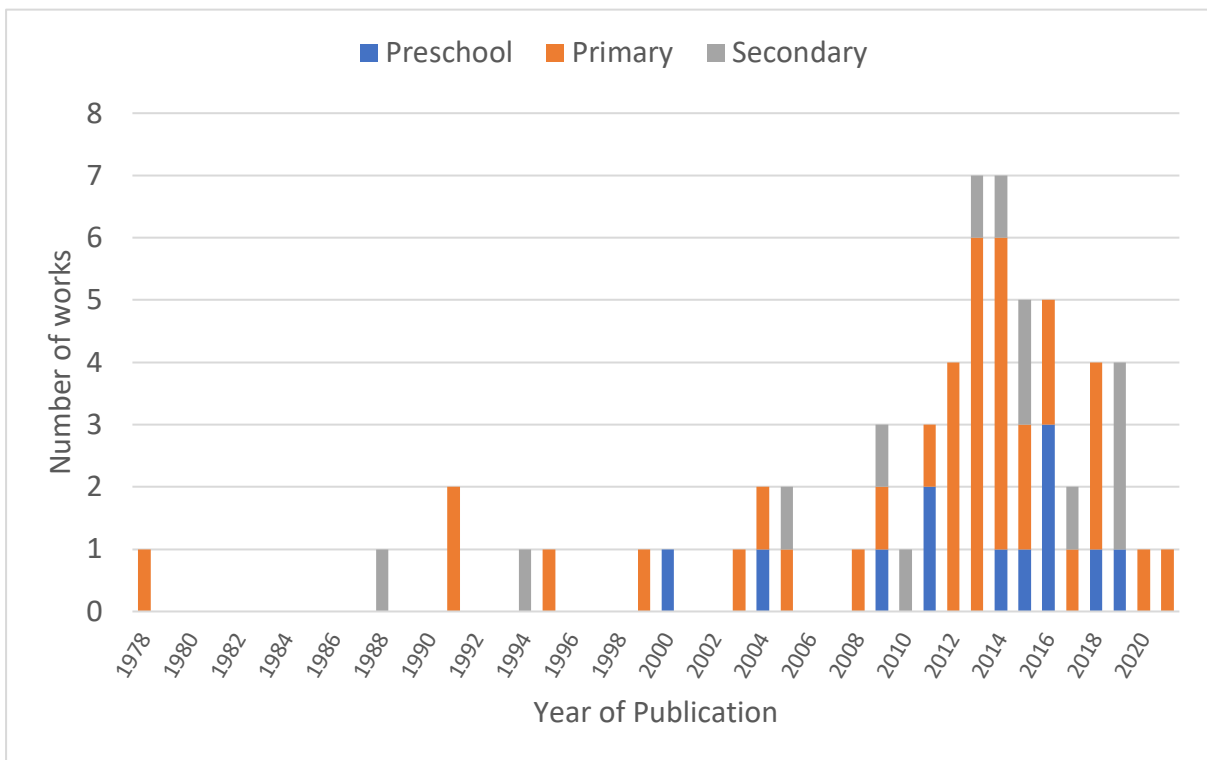


Figure 3 Instructional context of included studies

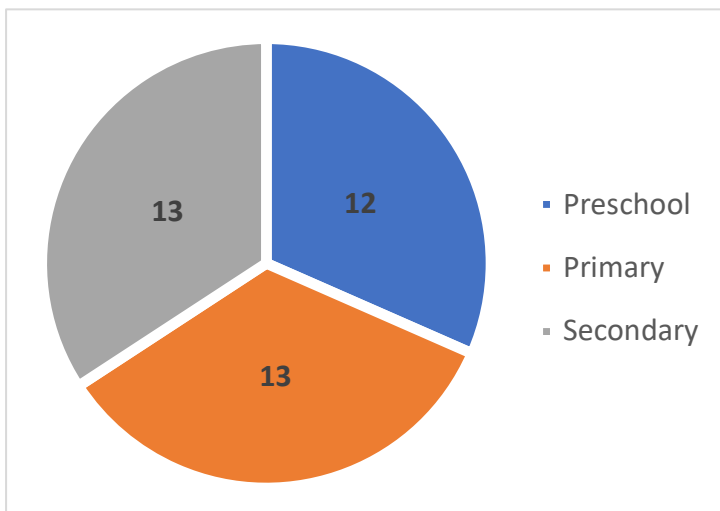


Figure 4 Sample size of included studies

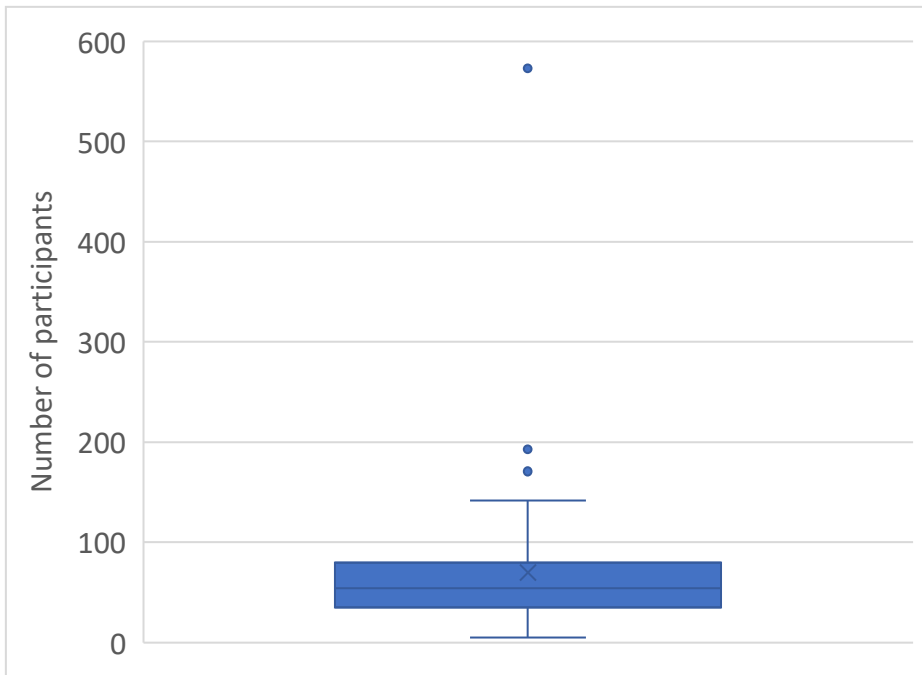


Figure 5 Study duration

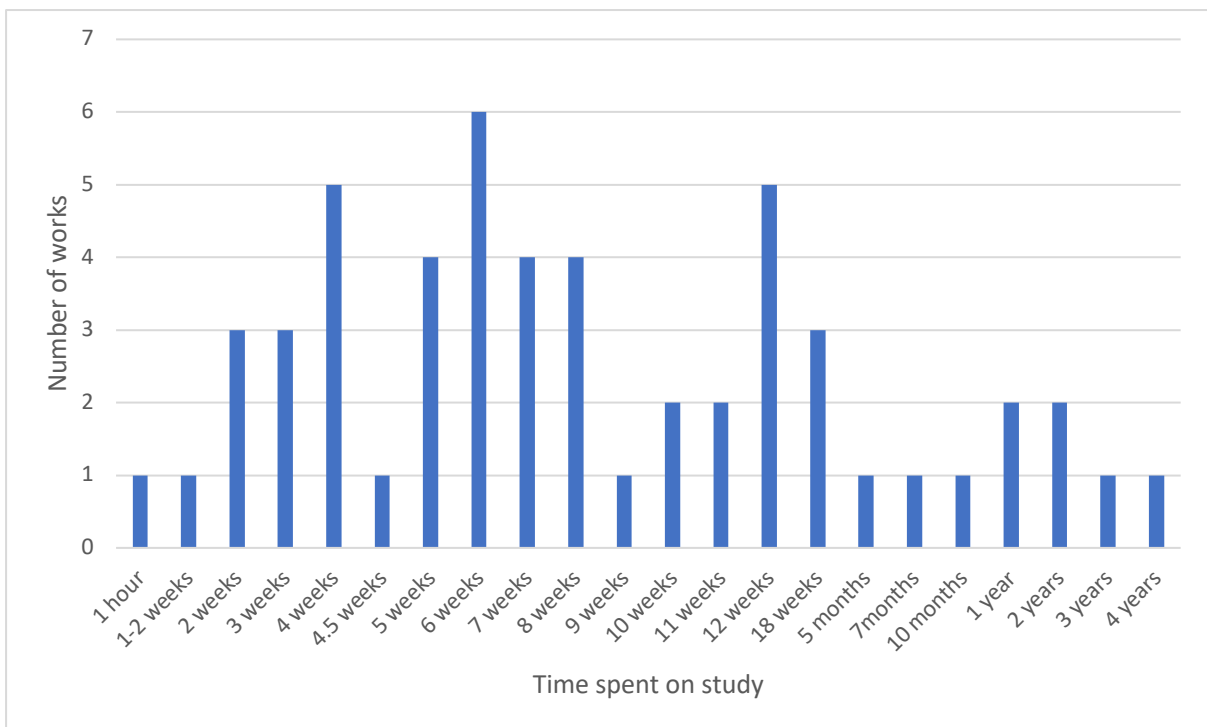


Figure 6 Outcome type by study

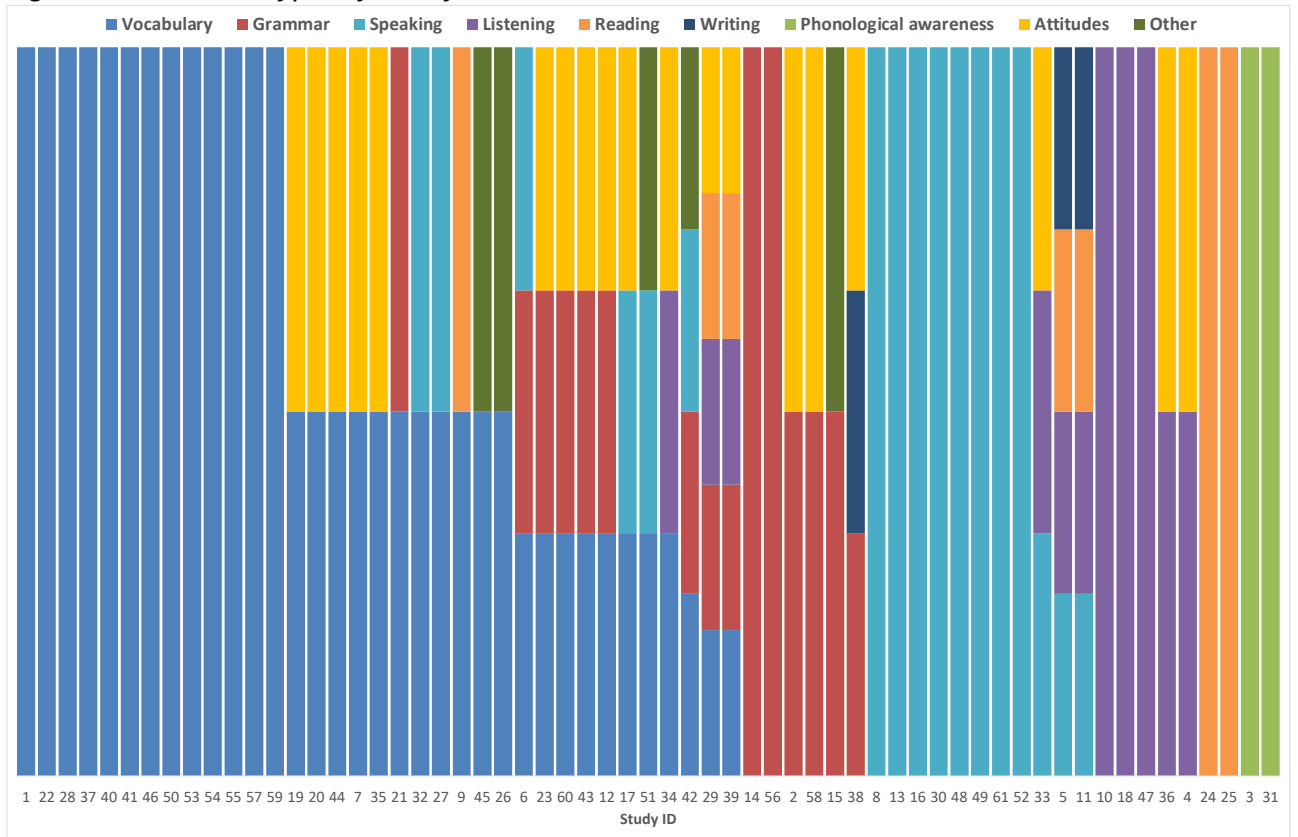


Figure 7 Outcome frequency

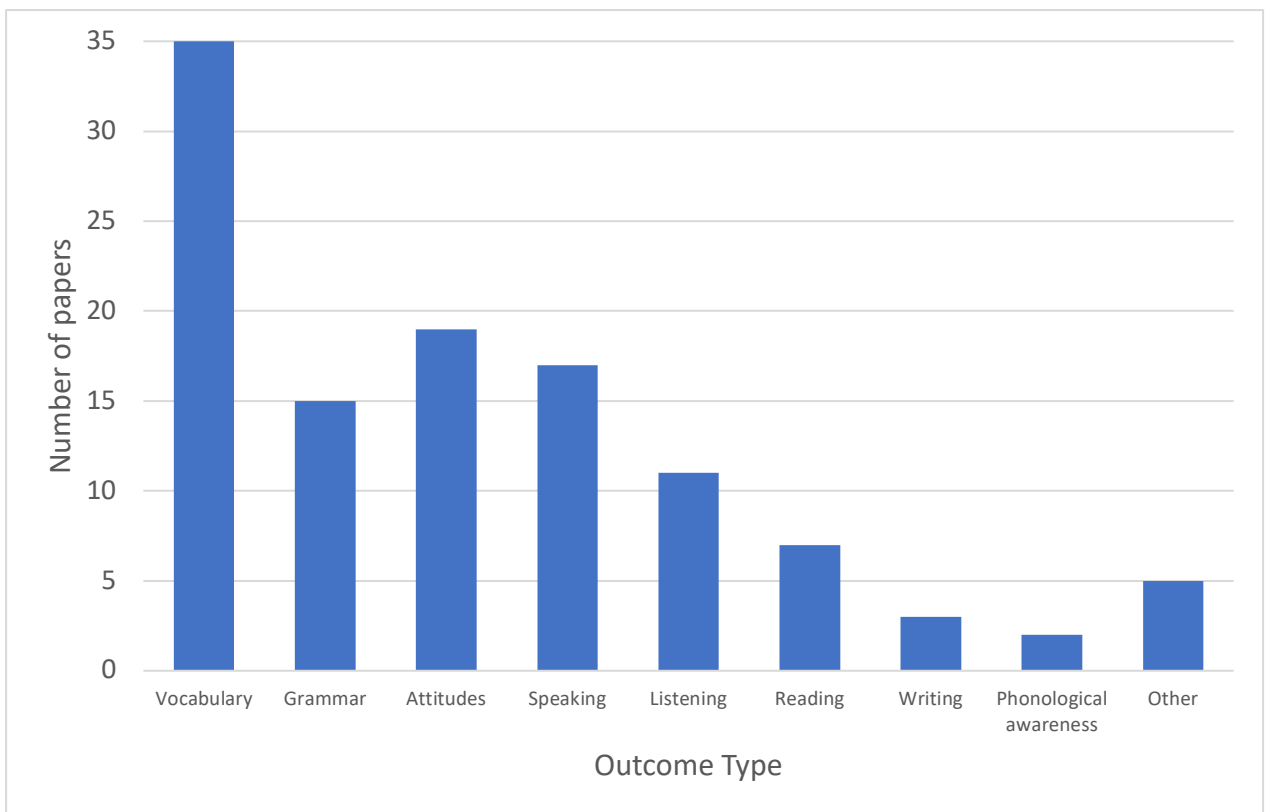


Figure 8 Reported effect of singing and weight of evidence of included studies

