"Song-advantage" or "Cost of Singing"? A Research Synthesis of Classroom-based Intervention Studies Applying Lyrics-based Language Teaching (1972–2019)

Valentin Werner, *University of Bamberg*

Abstract

While lyrics-based activities are increasingly popular in foreign language education, and a number of theoretical arguments have been suggested by applied linguists and SLA researchers why they are beneficial for learners, only a few empirical assessments of the actual effectiveness of lyrics-based teaching have been conducted. This contribution reviews relevant classroom-based intervention studies (N = 28) that employ a pre-/post-test multiple group design. A main aim is to assess previous claims of a "song-advantage" (Busse et al., 2018) when input is presented with the help of songs and their lyrics vs. a "cost of singing" (Racette & Peretz, 2007) in terms of an additional processing burden lowering rates of verbal recall, for instance. Results suggest that, overall, lyrics-based teaching is effective in comparison to control conditions. Effectiveness may vary, however, when different subareas (e.g. grammar vs. vocabulary) are compared. In summary, it is argued that lyrics-based activities can be viewed as a valuable means in the foreign language classroom, even though they cannot serve as universal remedy and are most effective when combined with other materials and activities. Eventually, it is suggested that both additional primary studies at a high level of methodological transparency and rigor and quantitative meta-analyses (e.g. to assess the influence of moderator variables and to systematically compare control groups with different teaching conditions) are desirable.

Keywords*: lyrics; songs; music; intervention studies; research synthesis*

Introduction

Even though they are underrepresented in textbooks and school curricula (Summer, 2011; Tegge, 2015, 2018; V. Werner, 2018), lyrics-based classroom activities,[1] for instance those using lyrics to introduce an area of grammar, vocabulary, etc., are a popular feature in the modern language classroom, and seem to be overcoming the traditional "low culture" association and their status as "time fillers" (Abbot, 2002, p. 11) or "relief" (Smith, 2003, p. 115), becoming a "legitimate alternative to traditional classroom tasks" (Engh, 2013). This is shown both by the availability of relevant practical materials for teachers (e.g. Arnold & Herrick, 2017; Lorenzutti, 2014; Paterson & Willis, 2008) as well as by the increasing research interest devoted to the topic by both applied linguists and language educators, especially over the last two decades (e.g. Davis, 2017; Kao & Oxford, 2014; Mobbs & Cuyul, 2018; Scott Langeland, 2013; Sposet, 2007, 2008; Summer, 2018; Tegge, 2017, 2018; Ziegler, 2016; see also the list of studies mentioned subsequently).

Research has identified a number of arguments for using song and lyrics in the foreign language classroom, drawing from the areas of psychology of learning and motivation, language pedagogy, applied linguistics, and Second Language Acquisition (SLA) theory, among others. A major line of argumentation suggests that teachers by using lyrics-based activities may increase their chances of connecting to their students' lifeworld and are enabled to incorporate learner interests, simultaneously fostering a "real life" connection by providing contextualized and meaningful content and focusing on authentic language (Coats, 2016). Additionally, such studies have argued that lyrics contain structures and linguistic phenomena regularly introduced in the instructed setting anyway, and thus students may appreciate lyrics more fully if they have the linguistic means to do so (cf. the concept of "cultural interest" as motivational factor; Dörnyei, 2010, p. 76). In addition, engaging with lyrics may lead students to expand their personal horizon to "hold out for

new cultural and linguistic relations and for new possible modes of identity" (Pennycook, 2010, p. 65).

A closely related argument derived from SLA theory is that students may benefit from "affective engagement" (Tomlinson, 2017, p. 8) – that is, the emotional quality of the lyrics material may lead to a lowered stress level, which in turn facilitates language learning. Note that all of the aforementioned factors have been claimed to raise the (intrinsic) motivation of students, a key psychological variable in successful language learning (e.g. Loewen, 2015).

Cognitive arguments pertain to the rate of verbal presentation as well as to the multimodal nature of song as a pop culture artifact. As to the former, it has been found that sung lyrics are presented at 75% of the rate of speech, which may facilitate processing and memorization (Busse et al., 2018; Kilgour et al., 2000). As to the latter, beyond linguistic information, encoding happens in a second mode – music – which may lead to multiple encoding and parallel information processing and may facilitate the retention of structures and content (e.g. Ginsborg & Sloboda, 2007; Jentschke, 2016).

Finally, from an applied linguistic perspective, an important issue is that using lyrics in the classroom may offer students the opportunity to engage with both standard (as commonly represented in textbooks and teaching material) and non-standard language (Werner, 2012). This opens opportunities for introducing them to actual language use in a naturalistic way and simultaneously raising their critical awareness for varieties and registers (Duff & Zappa-Hollman, 2013; Werner, 2019, 2021). In addition, it has been recognized that lyrics-based teaching may develop multiple skills/competencies, including listening comprehension, vocabulary, literacy as well as cultural understanding at the same time (Ludke, 2020). Further, lyrics have been found to be suitable as input material in mobile-assisted language learning (Werner, Lehl, & Walton, 2017).

In view of the aforementioned theoretical and practical rationales, which collectively have been termed the "song-advantage for language learning" (Busse et al., 2018), a number of classroom-based intervention studies have been conducted, especially in the last two decades. They address a central question for all language educators – is our teaching effective?[2] Relating this to the area of lyrics-based instruction, practitioners already employing lyrics in their classroom may have wondered about its actual effectiveness, while others may have been reluctant to rely on lyrics-based instruction for a perceived lack thereof in the first place.

A generic factor that may play a part here is that multiple encoding (music + lyrics), rather than having a facilitating effect (cf. the "song-advantage" mentioned before), may initially come with a "cost of singing" (Racette & Peretz, 2007, p. 250). This concept has been established through an experimental approach (with $N = 36$ subjects), testing whether music could be used as a mnemonic aid for verbal learning. In this experiment, where verbal recall was tested with both musicians and non-musicians both in a sung and spoken condition, it emerged that the sung condition results in an additional processing burden for all types of subjects, actually lowering rates of verbal recall, for instance, a result contrary to the expectations as suggested by previous research (see also Ferreri & Verga, 2016; Hahn, 1972; Ludke et al., 2014). The findings of this study are further suggestive of melody and lyrics being processed separately when new songs are learned (vs. multiple encoding as described above).

Given these contrasting viewpoints as outlined in the foregoing passages, the question arises whether lyrics-based instruction is effective (also in comparison to more traditional approaches), or whether it actually is subject to the "cost of singing". As lyrics-based instruction has been applied to different linguistic competencies/skills areas as described above, a related topic will be whether there are differences in effectiveness across these areas. These are the central issues to be addressed subsequently.

While findings in the domain of lyrics-based instruction have previously been assessed by way of shorter literature reviews (see, e.g., Romero, 2017; Sposet, 2007), the present synthesis aims at providing an updated and extended picture to familiarize readers with the body of empirical work focusing on lyrics-based teaching in foreign language education. In addition, it seems that researchers are not aware of each other's studies, so the present article may serve as a starting point informing prospective work. Further, in the spirit of Norris and Ortega (2006), weaknesses in study design are highlighted, also intended to inform future research efforts.

The rest of the article is structured as follows: First, the research domain covered is defined in more detail, followed by an outline of the principles that guided the selection of classroom-based studies. The next part will present the findings of the individual studies included in the review, structured according to the individual domains tested in the primary works (vocabulary, grammar, overall proficiency, etc.). In the discussion, commonalities and differences between studies will be highlighted to identify the overall potential of lyrics-based foreign language education. In addition, limitations of the present study are considered. The concluding part will provide an outline of the overall implications of the findings for applied linguists and language educators, finishing with some pointers to avenues for further research.

Defining the research domain

To address the research question of whether lyrics-based instruction is helpful, or whether using a lyrics-based approach is subject to the "cost of singing" as stipulated before, the current work focuses on specific primary studies. The ones included have tested whether there is an effect of lyrics-based instruction on the development of (a) overall language proficiency or (b) individual areas, such as grammar, vocabulary or reading skills. All studies in the analysis are classroom-based intervention studies with a (quasi-) experimental design, and employ scenarios involving pre- and post-tests (and

occasionally delayed post-tests). The focus is largely on within-group changes (pre- and post-test) and comparisons with control groups.

The initial literature search took a broad scope and aimed at all relevant studies using lyrics, be it commercially produced pop lyrics, pedagogical lyrics written by the educators themselves (see also R. Werner, 2018) as well as lyrics-based instruction using media channels such as *YouTube.* As a comprehensive literature research constitutes a fundamental step to ensure a maximally possible coverage of primary studies, the net was cast wide and manual keyword-based searches both on *Google Scholar* (scholar.google.com) as well as searches in the archives of SLA journals were employed.[3] In addition, potentially relevant investigations mentioned in the list of references of already retrieved studies were accessed and inspected.

A number of potentially relevant studies had to be ignored due to the fact that full versions were not available in published format and requesting them from the authors was not possible due to missing contact details (e.g. Ayotte, 2004; Hazel-Obarow, 2004). In addition, studies that reported lab-based findings (e.g. parts of Ludke, 2010, 2018; Ludke et al., 2014; Tamminen et al., 2017), that employed a design without a control group (e.g. Coleman, 2014; Coyle & Gómez Garcia, 2014; Fadli, 2017; Karlina et al., 2017; Limbong, 2012; Schwarz, 2013), that only used pedagogical songs as contained in textbooks or specifically written for the purposes of a study (e.g. Busse et al., 2018; Haghverdi, 2015) or that only reported on student attitudes (e.g. Chen, 2016; Hadian, 2015; Lieb, 2008) were excluded.[4]

Note further that a few primary studies (see, e.g., Ludke, 2010) included data from more than one area (e.g. separate scores for grammar and vocabulary) and different proficiency levels (e.g. intermediate and beginner learners). Table 1 provides an overview of the twenty-eight primary studies eventually considered, with information on tests applied to measure the effects of the intervention.

Table 1. *Overview of the Primary Studies Included (in Chronological Order; Abbreviations Used: MC = Multiple Choice; BF = Blank Fill; + = Bigger (Statistically Significant) Effect of Lyrics-based Instruction Compared to Control Setting; - = Smaller (Statistically Significant) Effect of Lyrics-based Instruction Compared to Control Setting; o = Similar Effect of Lyrics-based Instruction Compared to Control Setting)*

Study	Area(s) tested	*N* learners	Learner age(s)	Learner L1(s)	Target language(s)	Delayed post-test	Test(s)	Other relevant information	Effect
Hahn (1972)	vocabulary	38	12	English	German	×	written (MC, BF)	control group taught same content through dialogs; stronger effect for boys	+
Medina (1990)	vocabulary	48	9	Spanish	English	✓	written, picture-based (MC)	additional effect of pictures tested	o
Fisher (2001)	overall proficiency/literacy	80	NA (elementary school pupils)	Spanish	English	×	spoken, written	use of standardized tests (SOLOM, Yopp-Singer, DRA)	+
Shaffer (2004)	vocabulary	84	NA (university students)	Korean	English	✓	written (BF)		+
Legg (2009)	vocabulary	56	12–13	English	French	×	written	control group taught same content through poem	+
Li & Brand (2009)	overall proficiency	105	NA (postgraduates)	Chinese	English	✓	written (MC, sentence completion, short answers)	one control group without music-based instruction, the other using music half of the time (lowest scores)	o
Shen (2009)	overall proficiency	57	NA (undergraduates)	Chinese	English	×	spoken, written		+
Ludke (2010)	grammar, vocabulary	43	11–14	English	French	✓	various listen-and-repeat experiments	beginner and intermediate learner group	o
Smith Salcedo (2010)	verbal recall	94	17–41	English	Spanish	✓	written (BF)		+

Study	Focus	N	Age	L1	L2	Test cond.	Test type	Notes	Effect
Abidin et al. (2011)	vocabulary	68	NA (secondary school students)	Malay	English	×	written	use of *YouTube*	+
Alipour, Gorjian, & Zafari (2012)	vocabulary	60	20–32	Persian	English	✓	written (MC)	stronger effect for boys	+
Höfler (2013)	verbal recall	82	12–14	German	English	✓	written (BF)		o
Kara & Aksel (2013)	grammar	38	NA (undergraduates)	Turkish	English	×	written (MC, BF)		o
Köksal, Yağışan, & Çekiç (2013)	vocabulary, verbal recall	56	NA (elementary school pupils)	Turkish	English	×	written		+
Rahbar & Khodabaksh (2013)	listening comprehension	40	NA (university students)	Persian	English	×	written	female participants only	+
Tavakoli (2013)	grammar	25	16–18	Persian	English	×	written (MC)	male participants only, two control groups (task-based language teaching/present-practice-produce condition)	+
Ashtiani & Zafarghandi (2015)	reading aloud, speaking	40	18–25	Persian	English	×	expert rating of recordings	male participants only	+
Good et al. (2015)	verbal recall, pronunciation, translation	38	9–13	Spanish	English	✓	spoken interaction, rating of recordings through external rater		+
Heidari & Araghi (2015)	vocabulary	68	7–14	Turkish	English	×	written	control group taught same content through pictures, male participants only	-
Tegge (2015)	vocabulary	105	16–21	German, Dutch, Serbian	English	✓	written (MC, BF, free recall)	advanced and intermediate learner groups; two control groups with poem and prose condition	+/o
Alisaari & Heikkola (2016)	writing fluency	34	18–33	Various	Finnish	×	written (cartoon description)	three conditions: singing, listening, reciting (non-music condition)	+
Kuśnierek (2016)	vocabulary	28	11–12	Polish	English	✓	written		+

Rezaei & Ahour (2016)	listening comprehension	40	18–28	Persian, Turkish	English	×	written		+
Javadi-Safa (2018)	vocabulary	60	17–38	Persian	English	×	written (BF, true-false questions, short answers)	male participants only	+
Rukholm et al. (2018)	vocabulary	66	17–30	English	Italian	✓	written (standardized test/questionnaire)	poem condition; moderator variable of "elaboration" and separate testing of productive and receptive skills	+/o
González Arteaga (2019)	listening comprehension	38	19–26	Spanish	English	×	written (BF, MC, true-false questions, short answers)		+/o
Pavia, Webb, & Faez (2019)	vocabulary	300	10–14	Thai	English	✓	written (MC)	spoken cues for the test	+
Tomczak & Lew (2019)	verbal recall	26	18	Polish	English	✓	written (translation)	two test and control groups	+

<center>**Results**</center>

The following subsections contain descriptive summaries of relevant research, categorized according to the linguistic domain tested. The intention is to illustrate the approaches, scope and outcome of work available, with a specific view on the effectiveness of the respective types of lyrics-based instruction taken.

Vocabulary/verbal recall

Hahn (1972) is noteworthy as it represents an early study that systematically tested the effects of lyrics-based instruction on vocabulary storage of German phrases in a population of twelve-year old L1 English students over a six-week period. Hahn elicited various retention scores (recognition and recall) form both a test group ($N = 18$) that was presented songs and a control group ($N = 18$) that was presented dialogs. Overall, Hahn identifies using lyrics and music as an "effective mediating factor in the learning and retention of lexical items" (1972, p. 1). Additionally, in her study social factors emerged as important variables. Boys achieved higher retention scores in the song condition, while girls achieved higher scores in the dialog condition and when an English-cued recall test was used in the song condition. Students' self-reports collected in the study indicate that using lyrics in the classroom is experienced as an enjoyable activity.

Medina (1990) investigated the incidental acquisition of English vocabulary by L1 Spanish elementary pupils (age 9) across four conditions ($N = 12$ each; each group with comparable pre-test vocabulary scores) with the same lexical content: (1) lyrics-based instruction (2) story-based instruction (3) lyrics-based instruction with illustrations and (4) story-based instruction with illustrations. Medina conducted one test after a four-day treatment period and a delayed post-test after one and a half weeks. While higher vocabulary gain scores were observed both in the immediate and the delayed post-test for those groups treated with lyrics-based instruction, and especially the group with combined lyrics and illustrations, they did not differ significantly between the different conditions. Medina's data further suggest that especially low-proficiency students (as determined by the vocabulary pre-test scores) may benefit from lyrics-based instruction in combination with illustrations and (as determined through a questionnaire) that students consider lyrics-based activities as highly motivating.

Shaffer's (2004) study relies on a sample of L1 Korean learners of English (university students) of different ages and proficiency levels (low-intermediate to low-advanced). The intervention

took place over a five-week period. It can be considered innovative as it employed three conditions to determine a potential added value of using songs: Group (1) ($N = 32$) received lyrics-based instruction, group (2) ($N = 22$) received lyrics-based instruction but in a narrative form only (i.e. a non-music condition), while group (3) ($N = 30$) served as control group without any lyrics-based instruction at all. A self-designed pre-test, which determined similar scores across the three groups, and two post-tests were administered. Overall, it emerged that both group (1) and (2) significantly outperformed the control group in the first post-test, and group (1) showed significantly higher vocabulary retention scores in the delayed post-test administered after four weeks, which Shaffer interprets as suggestive of the song condition facilitating longer-term memorization.

Legg (2009) studied whether lyrics-based teaching supports the acquisition of French vocabulary by L1 English speakers (students of a comprehensive school, aged 12–13). He started from the hypothesis that the association of music with vocabulary leads to better short-term memorization. He used a test group ($N = 27$) that received a lyrics-based treatment and a control group ($N = 29$) that received the same content through a poem, with members of the two groups being randomly assigned. Students completed a self-developed vocabulary translation test before and after a one-hour treatment. The students in the test group increased their scores by a larger margin than the controls (53% vs. 39%), with the difference of the average scores between the two groups being significant.

Smith Salcedo (2010) examined verbal recall in L1 English beginner college students of Spanish ($N = 94$; aged 17–41). Comparable to Shaffer (2004), students were divided into three groups that received a treatment over six class periods by the same teacher: Group (1) received lyrics-based instruction (using three songs), group (2) received lyrics-based instruction but with the lyrics read rather than the songs actually played (i.e. a non-music condition), while group (3) did not receive any lyrics-based instruction at all. Smith Salcedo administered both an immediate post-test and a delayed post-test after two weeks and found that group (1) outperformed the other groups for two of the three songs used in the immediate post-test. This finding was not reproduced in the delayed post-test, however, which leads Smith Salcedo to suggest that longer lyrics-based treatment periods are necessary to achieve an effect.

Abidin et al.'s (2011) study is based on data from a cohort of L1 Malay secondary school students learning English. Both a test group and a control group of equal size ($N = 34$ each)

were pre-tested for vocabulary (self-developed test) and underwent a six-week treatment period, the test group being subject to lyrics-based instruction (using *YouTube* clips), the control group being subject to instruction based on traditional materials and approaches. After the intervention, the test group showed a higher gain in vocabulary scores, with differences to the control group being significant. An additional qualitative evaluation of student journal entries revealed that lyrics-based activities were experienced as highly enjoyable.

Alipour et al. (2012) investigated immediate vocabulary recall and longer-term retention in L1 Persian learners of English with high proficiency (as determined through TOEFL scores). Participants were randomly divided into a test and control group ($N = 30$ each), both balanced for gender. Over a two-month period, the test group received lyrics-based instruction (using music and lyrics; focusing on the subgenres pop, country, and rap), while the control group, comparable to the approach used in Shaffer (2004) and Smith Salcedo (2010), received lyrics-based instruction with the lyrics read rather than the songs actually played (non-music condition). The researchers administered multiple-choice post-tests every second week immediately after a series of sessions. A similar procedure was taken after the end of the intervention to determine long-term retention. Results showed that the test group outperformed the control group in vocabulary gains at a significant level. Further, Alipour et al. (2012) tested for gender differences and suggested that males may benefit more from lyrics-based instruction.

Höfler (2013) was modeled on Smith Salcedo (2010) and tested immediate and delayed verbal recall with four classes of Austrian L1 German secondary school students (aged 12–14) from two different school types. In each school type, a test group ($N = 44$) that received lyrics-based instruction with songs (music condition) and a control group ($N = 38$) that received lyrics-based instruction with the lyrics read by a native speaker of American English rather than the songs actually played (non-music condition) was established. A self-developed written blank-fill test was used to determine vocabulary scores as a pre-test, as well as vocabulary recall immediately after the intervention (one song played in one class session) and after two weeks. Höfler observed no significant differences between gain scores of the two groups tested in the immediate and delayed post-test, even though the test group achieved overall higher mean gain scores in the delayed post-test. Additional variables, such as gender, or those that were elicited through a questionnaire, such as song appraisal, did not emerge as significant in influencing participants' recall performances.

Köksal et al. (2013) investigated the increase in vocabulary scores as determined through a self-developed written test in a sample of L1 Turkish elementary school pupils learning English. To this end, two classes ($N = 29$ each) were assigned as test and control group, respectively. Over a period of twelve weeks, the test group was instructed with lyrics-based vocabulary activities, while the controls received traditional teaching. A vocabulary test was administered immediately after the treatment and again after a delay of one month. In this sample, the gain scores of the test group emerges as significantly higher than those of the control group both in the immediate and the delayed post-test.

Good et al. (2015) tested verbal recall of a lyrics passage of one song in a group of L1 Spanish elementary school pupils (aged 9–13) learning English. Over a period of two weeks, one class ($N = 16$) received a lyrics-based treatment, while another class ($N = 22$) was presented the same material as a poem by the same teacher. Both groups were exposed to the material 20 times in the treatment period. In an immediate post-test after the end of the intervention, the pupils having received the lyrics-based treatment yielded significantly higher recall scores than those in the poem condition. This finding was reproduced in a delayed post-test taking place six months after the intervention. In addition, Good et al. (2015) also elicited data on the pronunciation and on the translation of words, where they found a significantly higher accuracy of the test group as regards the reproduction of vowel sounds (but not for consonants) and as regards translation in the immediate, but not in the delayed post-test. In addition, anecdotal evidence (teacher reports) suggested that the students experienced lyrics-based instruction as highly motivational.

Heidari and Araghi (2015) assessed vocabulary gain scores in a sample of L1 Turkish learners of English (aged 7–14) in a language institute. Homogeneity across the learners in terms of comparable proficiency was ensured through the standardized A2 Flyers test, and students were randomly assigned to two groups ($N = 34$ each), one receiving vocabulary instruction through lyrics-based activities, the other one through picture-based activities over a seventeen-week period. A self-developed vocabulary pre- and a post-test were administered to the students and results showed that the vocabulary gain scores for the group receiving picture-based instruction were significantly higher than for the lyrics group.

An intricate study design is presented in Tegge (2015). She tackled the question whether a lyrics-based, poem-based, or prose-based presentation of a text leads to the highest rate of

verbal recall. Learners (undergraduates and secondary school students) with various L1 backgrounds (Serbian, German, Dutch; aged 16–21) were divided according to the three conditions ($N = 30/33/43$) and listened (as a song or as a recorded reading) to the material three times during the one-lesson intervention that was similar for all groups involved. Three self-developed immediate and delayed (after one week) post-tests followed (cued and uncued recall, recognition). Overall, the findings of the quantitative analysis revealed that the lyrics-based intervention resulted in significantly higher rates of verbal recall compared to the other two conditions. An exception were the non-significant differences between the lyrics and the poem condition in some of the individual post-tests in the Serbian and Belgian learner sample, which is suggestive of a comparable effectiveness of using poems for verbal recall. Tegge (2015) further identified an effect of increasing verbal recall with increasing overall learner proficiency.

Kuśnierek (2016) studied vocabulary recall in a sample of 28 L1 Polish beginning learners of English (aged 11–12) in an elementary school classroom, which were equally distributed among a test group that received lyrics-based instruction and a control group that did not. A vocabulary translation test was administered before and after a two-week period intervention period, followed by a delayed post-test after another week. It emerged that the test group achieved higher scores in both post-tests, while the results obtained were not subject to statistical testing. In an additional attitude questionnaire, a large majority of students from the test group (72%) reported that they experienced the lyrics-based lessons as highly enjoyable.

Javadi-Safa (2018) investigated vocabulary acquisition in a cohort of adult L1 Persian learners of English at an intermediate proficiency level with a wide age range (17–38). Similar to Tegge (2015), he worked with a test group ($N = 30$) that received lyrics-based instruction, while a control group ($N = 30$) was presented with a recorded and read prose version of the lyrics (i.e. in a non-music condition). Participants were randomly assigned to the groups and the intervention took place over a 20-week term. In the results of a self-developed post-test, it emerged that the test group received significantly higher vocabulary scores than the controls.

Rukholm et al.'s (2018) study probed into incidental acquisition of vocabulary across different conditions, relating both to receptive and productive knowledge. It is based on findings for adult L1 English learners of Italian (aged 17–30) at the beginner level at a post-secondary institution ($N = 66$). Vocabulary knowledge was assessed through a pre-test and a standardized questionnaire (the Vocabulary Knowledge Scale) after a four-week intervention phase and as a

delayed post-test after additional two weeks. It is worth noting that this study employed four test groups with different conditions (lyrics-based and poem-based instruction with high and low levels of "elaboration", i.e. with different additional classroom exercises targeting vocabulary development), as well as one control group that did not receive any relevant instruction. As regards receptive lexical knowledge, the results showed that the group that received lyrics-based instruction in the high elaboration condition improved scores by the widest margin, while no differences were found between the lyrics- and poem-based instruction groups in the low elaboration condition. As regards productive lexical knowledge, the group with lyrics-based instruction in the high elaboration condition is the only one that shows a significant improvement over time. In the longitudinal perspective (delayed post-test), however, also the group treated with lyrics and high elaboration showed a significant regression of scores (cf. Good et al., 2015).

Pavia et al. (2019)[5] also investigated incidental learning of vocabulary (individual items and collocations) across different learning conditions in a large sample ($N = 300$) of L1 Thai beginning learners of English (aged 10–14) from a public school setting over a period of up to five weeks. Participants were assigned to different conditions (six experimental groups using two different songs, presented once, three times, five times with one-week intervals; two control groups with no lyrics-based instruction) according to the classes they attended. No proficiency differences between groups emerged as determined through the standardized Vocabulary Levels Test. A self-developed test was administered as pre-test and immediate and delayed post-test to examine the recognition of spoken forms, the recognition of the meaning (translation task) and the recognition of collocations. Overall, the data yielded significant differences between pre- and both post-test scores for the experimental groups for the recognition of spoken forms for one of the songs but some mixed results for the other song and the further areas tested. This leads the authors to suggest that "learners can incidentally learn L2 vocabulary through listening to songs [but with] relatively small learning gains" (Pavia et al., 2019, p. 17), potentially due to the restricted scope of the intervention and the relatively low overall vocabulary size of the learner sample tested. Secondary results emerging from the study are that multiple repetition of the lyrics input material may lead to higher gains in vocabulary and that there are differences in learning gains determined by the individual songs used.

Tomczak and Lew (2019) represents a recent study that specifically assessed verbal recall of multi-word units, which are conceived as items that increase learners' fluency and idiomaticity. This investigation is based on a sample of intermediate L1 Polish secondary-school students of

English (mean age = 18), which were separated into two test groups (*N* = 12/14), taught in a lyrics condition, and two control groups (*N* = 12/14) that received traditional listening comprehension practice. As determined by a pre-test (translation of multi-word units from Polish into English) there were no significant differences between the groups as regards the knowledge of multi-word units. The intervention took the shape of a 45-minute lesson in which a set of thirteen multi-word units was introduced with the help of two pop songs or several sentences containing target items read out by a native speaker, respectively. Both an immediate post-test (at the end of the intervention) and a delayed post-test (after one week), which again consisted of Polish-English translations of multi-word units, were administered. The overall quantitative results suggest that the test group significantly outperformed the control group in terms of recall rates, especially in the delayed post-test. This finding also held across the two test/control groups and "tall[ies] with the naive [sic] perception of song lyrics being 'memorable'" (Tomczak & Lew, 2019, p. 28), so that multi-word units presented as lyrics potentially are highly salient for learners, which may make lyrics-based teaching an adequate and effective approach for introducing them.

Grammar

Ludke (2010) focused on the learning of French grammar and vocabulary through L1 English secondary school pupils (aged 11–14) over a four-week period. She specifically aimed at assessing the influence of additional factors such as learner age, previous language learning and artistic experience, gender, as well as individual art preferences.[6] The learner sample was divided into a class (*N* = 24) that first received two weeks of lyrics-based instruction, followed by two weeks of instruction through dramatic dialog, and a class (*N* = 19) where the order was reversed. All participants completed a pre-test, a test after two weeks and one after the end of the treatment, consisting of a phrase translation and a blank fill exercise with separate scores for grammatical and lexical accuracy. Results showed that scores increased significantly over both treatment periods for the class starting with lyrics-based instruction, but only significantly increased for the dramatic dialog phase for the second class, so no overall superior effect of lyrics-based instruction was found. However, an additional survey among the participants revealed that the majority of students (62.5%) preferred the lyrics-based approach to the one based on dramatic dialog (16.7%).

Kara and Aksel (2013) represents a study that investigated the effect of lyrics-based teaching on the acquisition of English tense-aspect forms through a sample of L1 Turkish

undergraduates. Students were randomly assigned to a test or a control group ($N = 19$ each). Over ten days, the test group received additional grammar instruction through eight short 10-minute lyrics-based activities, while the controls received comparable instruction through traditional approaches in this time. A gain in accuracy scores was determined through a self-created combined blank-fill and multiple-choice test, which was administered before and immediately after the end of the intervention. In this study, no significant differences emerged between the two groups, while the increase in average gain scores was higher for the test group. An additional questionnaire survey established, however, that students found lyrics-based activities highly enjoyable and motivating.

The investigation by Tavakoli (2013) is conspicuous as it employs a three-group design to study the effectiveness of task-based language teaching (TBLT) with or without a lyrics-based component in comparison with a traditional present-practice-produce approach toward grammar instruction. His sample consists of L1 Persian learners of English ($N = 25$; aged 16–18), with control groups of equal size and proficiency receiving TBLT without lyrics and with the traditional approach. A self-created multiple-choice test with twenty items (from the domains of tense/aspect and conditional clauses) served as both pre- and post-test, while student attitudes were also elicited with the help of a questionnaire. The test group received TBLT in which parts of the instruction were based on pop lyrics, with both the performed songs and the printed lyrics being available as materials.[7] While all three groups received similar scores in the pre-test, it emerged that the group that received TBLT with the lyrics component received significantly higher scores in the post-test than the other two groups. Therefore, Tavakoli suggests that TBLT in combination with lyrics-based instruction is conducive to effective contextualized grammar learning. In line with several other studies discussed in the foregoing, the questionnaire results showed that students expressed highly positive attitudes toward learning with lyrics (while this result also held for TBLT without the lyrics component).

Listening comprehension
Rahbar and Khodabaksh (2013) based their investigation of listening comprehension skills on a sample of L1 Persian adult students of a general English course, which all reached comparable overall proficiency levels as determined by the standardized Oxford Placement Test and were randomly assigned to a test and control group ($N = 20$ each). The test group received listening practice with song lyrics, while the control group received traditional listening practice based on principles of communicative language teaching. A self-developed listening comprehension test

was employed both as a pre-test and as a post-test after the two-month intervention phase, and results suggest that the test group achieved significantly higher scores in the post-test.

Based on a group of bilingual (L1 Turkish and Persian) learners of English (aged 18–28) in a language institute, Rezai and Ahour (2016) studied the development of listening comprehension. Comparable to Rahbar and Khodabaksh (2013), learners were randomly assigned to a test and control group ($N = 20$ each) after their general level of proficiency was determined through a standardized test (the Preliminary English Test). This test also served as a pre-test determining listening comprehension scores, while a different section of the same test was used as post-test. Both groups were taught by the same teacher over a seven-and-a-half week period, the test group receiving lyrics-based teaching, while the control group received traditional instruction. It emerged that the test group reached significantly higher scores in the post-test. Self-reports of students that were informally collected suggested that they experienced the lyrics-based instruction as enjoyable and relaxing.

González Arteaga (2019) explored listening comprehension of English among a group ($N = 39$) of Ecuadorian L1 Spanish university students (aged 19–26) enrolled for non-language subjects. Both the test group and a control group not undergoing the intervention ($N = 38$) but doing traditional listening comprehension exercises completed a written test with multiple answer formats before and after the intervention. The intervention took place over a four-month period, with repeated short lyrics-based phases of twenty minutes each four times a week. The test group showed higher gains in the listening comprehension post-test scores than the control group. Differences were not statistically significant, however. Qualitative written interviews revealed that that the vast majority of students experienced the lyrics-based activities as enjoyable and found it a helpful approach for learning English they would like to increasingly use in their future studies.

Reading/speaking
Ashtiani and Zafarghandi (2015) is a study that is noticeable as it explicitly tested the use of lyrics-based instruction on the development of spoken skills. The learner sample consisted of intermediate L1 Persian learners of English at a language institute and the two domains of reading aloud and free speech were targeted. Participants all reached comparable proficiency levels (as determined through the standardized Nelson English Language Test) and were randomly assigned to a test and control group ($N = 20$ each) taught by the same instructor. The

test group repeatedly was instructed with lyrics-based activities focusing on pronunciation. Both groups completed a self-developed pre- and post-test (after an intervention phase of seven weeks). Free speech was assessed on the basis of student answers to structured interview questions loosely based on those used in the TOEFL. Scores were independently assigned through two expert raters per student that judged the accuracy of features of connected speech. Results showed that the test group received significantly higher scores in both the reading aloud and free speech condition.

Writing fluency

The study by Alisaari and Heikkola (2016) is unique in that it explores the effect of lyrics-based instruction on the development of written fluency. They divided their learner sample of L2 Finnish learners (with various L1 backgrounds; age range 18–33), which all received instruction over a four-week period, into three groups subject to different types of lyrics-based instruction: (1) a group singing songs ($N = 14$), (2) a group listening to songs ($N = 20$) and (3) a group with a non-music condition only reciting songs ($N = 18$). All groups were tested with the help of a test (serving as both pre- and post-test) that required cartoon descriptions, with fluency operationalized as the number of words used in the individual texts produced. It emerged that the largest increase in writing fluency was observable for the "singing" group (1), while also the "listening" group (2) showed a gain in fluency scores. This gain was significantly higher than the one of the group in the non-music condition.

Overall proficiency/literacy

Fisher (2001) studied the development of overall literacy in a sample of L1 Spanish elementary pupils over a nineteen-month period. He randomly assigned pupils to groups and contrasted a test group ($N = 40$) partly subject to lyrics-based instruction with a control group ($N = 40$) only subject to traditional types of instruction (other variables held constant or at least approximated). Three standardized tests assessing oral production (SOLOM), phoneme segmentation (Yopp-Singer test) and reading/re-telling (DRA) were administered to the pupils before and after the treatment period. The test results showed that the test group outperformed the controls for all areas at a statistically significant level, suggestive of a positive effect of lyrics-based instruction on literacy development. Fisher further emphasizes that lyrics-based activities should be part of regular curricular activities, rather than be viewed as "additional" material.

Li and Brand's (2009) investigation considered both vocabulary acquisition and overall proficiency development of English in a sample ($N = 105$) of L1 Chinese postgraduate law students. They randomly assigned participants, which all attended the same classes with the same instructor and achieved similar average values in a pre-test, to three groups: (1) exclusive use of lyrics-based instruction (for areas such as vocabulary, listening and reading comprehension, pronunciation, and grammar), (2) use of lyrics-based instruction half of the time, and (3) no lyrics-based instruction at all. All groups underwent treatment of six 90-minute classes, followed by a self-developed immediate post-test and a delayed post-test after three weeks. Group (1) achieved the highest scores in both post-tests, with significant differences to group (2) but not to group (3). Li and Brand (2009) also elicited attitude scores in the post-tests and found that the group receiving lyrics-based treatment had the most positive attitude toward language learning.

Shen (2009) compared results in a written and oral final test between a test ($N = 31$) and control group ($N = 26$). Over one semester he taught English to L1 Chinese electronics and communication engineering undergraduates that started from comparable scores as determined by a national entry exam. Otherwise using the same approaches and methods as the control group, the test group was presented lyrics-based blank-fill exercises to foster listening comprehension, songs were used to illustrate specific phonological, lexical and grammatical features and lyrics-based writing tasks were given. Eventually, Shen (2009) observed higher scores for the test group compared to the control group, while no test for significance was applied.

Discussion

A general pattern emerging from the review of relevant studies is that the test groups (lyrics conditions) usually outperform control groups (non-lyrics conditions) in the respective areas tested. This pattern was found for the majority (22/28) of the studies included, while among those only marginally advantageous effects could be observed in three studies. In five studies, no advantage of lyrics-based instruction over the alternative approaches used was found, while merely one study found lyrics-based instruction to be less effective than the control condition (picture-based instruction). In response to the overall research questions, it thus first emerges that only very limited evidence for a "cost of singing" is observable in intervention studies in a classroom-based environment, and that the "song-advantage" seems to prevail. In addition, there is tentative evidence that actually singing the song may be even more beneficial than

merely listening, a result that emerged at least for the area of writing fluency (Alisaari & Heikkola, 2016). The bird's eye view of the studies further implies that the stronger effect of lyrics-based instruction compared to the control settings is persistent (a) across learner ages and proficiency levels, (b) across the language pairs involved, and (c) across the individual linguistic areas addressed. An exception to (c) seems to be constituted by the area of grammar, where findings have been more ambiguous. However, only three studies could be considered in this category, and it emerges overall that studies focusing on vocabulary acquisition and verbal recall are overrepresented.

A secondary finding applies with respect to the motivational potential of lyrics-based instruction. An observation that was consistent across the primary studies in which student perceptions and attitudes were elicited was that participants in test groups perceived lyrics-based instruction as an enjoyable and motivating activity. As student motivation has been found to be a crucial variable in any language learning process (see, e.g., Dörnyei, 2010), it is probable that the "hedonic value" (Good et al., 2015, p. 637) of this type of instruction may translate into a long-term positive effect on overall learning gains, or at least to a lowering of foreign-language learning anxiety (see Dolean, 2016).

Despite the positive effects emerging from the overview of intervention studies, it is evident that lyrics-based instruction does not represent a silver bullet, and that there are other teaching approaches that are equally effective. More concretely, even though it emerged in some of the primary studies that groups with lyrics treatment outperformed control groups, this was by no means an unambiguous result. While control groups that only received input through dialogs or poems as a rule obtained lower results in the post-tests, this was not the case if control groups were taught through picture-based activities (e.g. Heidari & Araghi, 2015; Medina, 1990). This suggests that combining lyrics-based instruction with other materials may be even more effective than using lyrics alone, a result also emerging from recent investigations such as Rukholm et al. (2018). In this survey the "high elaboration" samples, that is, those with related practice in addition to working with a song, showed markedly better gain in scores between pre- and post-test, so that the authors submit that "song and high elaboration activities should be implemented in the L2 curriculum" (Rukholm et al., 2018, p. 153). Such statements are supported by the present results.

While the general picture is an encouraging one for language educators that use lyrics-based activities, a number of limitations of the present analysis are worth discussing. Minor points are that the intervention phase of many studies was comparatively short (with studies such as Legg, 2009 or Tomczak & Lew, 2019 administering the post-test after merely one teaching session), and that only a restricted number of studies (13/28) administered delayed post-tests at various points in time. Thus, it would be crucial to investigate the long-term effects of lyrics-based instruction in greater detail. Another caveat is that a substantial number of studies retrieved in the search of potentially relevant work used research designs without control groups and thus could not be considered. Others (e.g. Javadi-Safa, 2018; Smith Salcedo, 2010) did not administer pre-tests to ensure homogeneous samples in the test and control groups, respectively. This means that the results of these studies have to be taken with a grain of salt and are less generalizable.

A further restriction on the generalizability of the findings is the fact that the vast majority (21/28) of the primary studies included focused on English as target language. At the same time, it came as a surprise that a large part of the primary studies was conducted in non-Western settings, so the cultural bias we would commonly expect (see also Pavia et al., 2019) seemed to be reversed in this particular area of inquiry.[8] This also means that in the future we need more intervention studies from different L1 and target language settings to arrive at a comprehensive picture. This will also serve to rectify another potential weakness of the present work, namely that the number of primary studies that could be included was comparatively small, and that some areas (e.g. vocabulary/verbal recall) have received much broader coverage than others (e.g. grammar or reading). On a related note, it may be worthwhile to use more spoken tests, which are underrepresented in the set of studies (5/28), to determine the effect of instruction that to a large degree happens aurally.

It would further be desirable for future primary studies (a) to describe in more detail how the interventions were implemented and which types of tests and interventions were conducted (as some descriptions stay at a rather vague level; see, e.g., Abidin et al., 2011; Shen, 2009) and (b) to state more explicitly whether and how other forms of input were controlled during the intervention phase to exclude potentially confounding factors or to be in a position to acknowledge them as moderator variables (see also Pant, 2014). At the same time, it is clear that there is an inherent trade-off between being able to control all forms of input (as potentially

possible in a formalized experimental setting) and working within an ecologically valid, but potentially less controllable, classroom environment.

The point of testing for moderator variables is also crucial for future quantitative meta-analytic work. This could include more detailed considerations on gender differences (e.g. Hahn, 1972), the influence of the learners' age, proficiency levels and first-language background (e.g. Ludke, 2010; Tegge, 2015), differences between receptive and productive skills (e.g. Coyle & Gómez Garcia, 2014; Rukholm et al., 2018), as well as the role of the individual educator and teaching approach taken.

Conclusion

In sum, the results of this comparative review suggested that, in addition to the positive effects on student attitudes, employing lyrics-based instruction is an effective approach in teaching foreign languages, even though the extent of the effects differs for individual areas, with grammar potentially representing an area where lyrics-based instruction is less effective than other approaches. From a practical perspective, this may come as an encouraging result for those that already incorporate lyrics-based activities in foreign language education and may be viewed as an incentive for others to do so regularly in the future. It has to be conceded, though, that areas other than vocabulary acquisition/verbal recall and scenarios not featuring English as a target language are less well explored, so that the summary results presented in this study are of a tentative nature.

From a methodological point of view, a generic point of criticism that could be raised against syntheses of the kind presented is that they aim to represent a set of rather heterogeneous studies (in terms of subject areas, learner ages, etc.). Yet, the principle of comprehensiveness arguably outweighs concerns about being too inclusive in the choice of primary studies assessed. In this light, the statement, originally formulated for quantitative meta-analyses, that "comparing apples to oranges is sensible when the goal is to learn about fruit" (Oswald & Plonsky, 2010, p. 91) can be considered equally valid for the present purposes (assessing the overall effect of lyrics-based instruction). On a related note, the present study hopes to have illustrated the benefits of a systematic narrative synthesis in terms of accessibility for language educators as well as in terms of highlighting key patterns in the extant specialist literature, as forcefully supported by Ellis (2015, 2018). This does not mean, however, that this type of review is inherently superior to purely quantitative meta-analyses, which are helpful for identifying

160

specific moderator variables, as previously discussed. For the set of studies included above, a severe restriction in terms of the variety of aims, designs and measures of proficiency and learning gains applies (see Table 1), which would render a quantitative meta-analysis that compares effect sizes difficult. At the same time, this means that such an effort should stay on the agenda for the future once more comparable primary intervention studies become available.

To eventually widen the perspective of the present review to the area of pop-culture informed language education in general (see, e.g., Summer, forthcoming and contributions in Werner & Tegge, 2021), it will be worthwhile to review empirical intervention studies that relied on other types of pop culture artifacts, such as cartoons (e.g. Hua & Li, 2015) or movies and TV series (e.g. Dose, 2013).

Biodata

Valentin Werner is assistant professor of English linguistics at the University of Bamberg, Germany. He researches and teaches in the areas of applied linguistics (esp. learner Englishes and application of linguistic findings in language education), corpus linguistics, language variation and change (esp. World Englishes), and stylistics. He has recently co-edited the volumes Tense and Aspect in Second Language Acquisition and Learner Corpus Research *(John Benjamins, 2020) and* Pop Culture in Language Education. Theory, Research, Practice *(Routledge, 2021).*

Notes

1 The relevant literature also features the alternative terminology "music-based" or "song-based" instruction/teaching/etc. (e.g. Busse et al., 2018; Tegge, 2015). The focus in the language classroom will in all probability be on the lyrics, as they contain the linguistic material with which educators work, so "lyrics-based" will be used throughout this article. This terminology is certainly not intended to disregard the inherent value of musical aspects (e.g. melody and rhythm) for drawing attention to linguistic surface structures and fostering memorization (Lehmann & Seufert, 2017; Purnell-Webb & Speelman, 2008; Wallace, 1994) or for raising students' motivation.

2 A similar question was lately raised in Bennett (2019). His argument for using lyrics-based activities in language education does not exclusively draw on and synthesize the results of

empirical intervention studies, however. In fact, only one intervention study (Tavakoli, 2013) is summarized and presented as evidence.

3 The keywords used were the words *song* or *lyrics* or *music* in combination with the items *foreign language, second language, SLA, language instruction, language development, language acquisition, language education, language pedagogy, singing*. All studies retrieved were written in English. Following precedence such as Reichelt (2001) or Thomson and Derwing (2015), unpublished work (MA theses and doctoral dissertations) was included.

4 Note that the present review also ignores studies that exclusively tested the effect of lyrics-based instruction on pronunciation. Potentially relevant studies (such as Farmand & Pourgharib, 2015, or Izzah & Sukrisno, 2017, for instance) did not contain sufficient information (e.g. on participants and procedures) to make them comparable to each other.

5 This study is a revised and shortened version of Maneshi (2017).

6 Overall, previous language learning experience emerged as most important variable in determining test scores. As attempting to sum up the influence of all individual factors would go beyond the scope of the present synthesis, please refer to the relevant sections of the primary study (Ludke, 2010, pp. 267–292).

7 Unfortunately, the duration of the intervention phase is not specified for this study.

8 A potentially connected limitation of the study is that several primary studies have appeared in low-prestige or newly founded journals. This may either reflect a publication bias on part of established publishers and "gatekeepers" (journal editors, reviewers, etc.) to accept studies on lyrics-based instruction or pop culture issues in foreign language education or point towards the need for increasing the methodological rigor of such studies.

References

Abbott, M. (2002). Using music to promote L2 learning among adult learners. *TESOL Journal, 11*(1), 10–17.

Abidin, M. J. Z., Pour-Mohammadi, M., Balbir Singh, K. K., Azman, R., & Souriyavongsa, T. (2011). The effectiveness of using songs in *YouTube* to improve vocabulary competence among upper secondary school studies. *Theory and Practice in Language Studies, 1*(11), 1488–1496.

Alipour, M., Gorjian, B., & Zafari, I. (2012). The effects of songs on EFL learners' vocabulary recall and retention: The case of gender. *Advances in Digital Multimedia, 1*(3), 140–143.

Alisaari, J., & Heikkola, L. M. (2016). Increasing fluency in L2 writing with singing. *Studies in Second Language Learning and Teaching, 6*(2), 271–292.

Arnold, J. L., & Herrick, E. (Eds.). (2017). *New ways in teaching with music.* Alexandria, VA: TESOL Press.

Ashtiani, F. T., & Zafarghandi, A. M. (2015). The effect of English verbal songs on connected speech: Aspects of adult English learners' speech production. *Advances in Language and Literary Studies, 6*(1), 212–226.

Ayotte, S. (2004). *The acquisition of verb forms through song.* (Unpublished doctoral dissertation). Michigan State University, USA.

Bennett, P. A. (2019). The effects of music in the foreign language learning classroom. *Relay Journal, 2*(1), 6–16.

Busse, V., Jungclaus, J., Roden, I., Russo, F. A., & Kreutz, G. (2018). Combining song- and speech-based language teaching: An intervention with recently migrated children. *Frontiers in Psychology, 9*, 2386. doi.org/10.3389/fpsyg.2018.02386

Chen, A.-H. (2016). Perspectives upon integrating music into freshman English pronunciation training classroom. *Asian Journal of Education and Training, 2*(1), 16–22.

Coats, G. (2016). Analyzing song lyrics as an authentic language learning opportunity. In A. J. Moeller (Ed.), *Fostering connections, empowering communities, celebrating the world: 2016 report of the Central States Conference on the Teaching of Foreign Languages* (pp. 1–22). Richmond, VA: Terry.

Coleman, R. (2014). Exploratory practice: Researching the impact of songs on EFL learners' verbal memory. *Journal of Second Language Teaching & Research, 3*(1), 53–70.

Coyle, Y., & Gómez Garcia, R. (2014). Using songs to enhance L2 vocabulary acquisition in preschool children. *ELT Journal, 68*(3), 276–285.

Davis, G. S. (2017). Songs in the young learner classroom: A critical review of evidence. *ELT Journal, 71*(4), 445–455.

Dolean, D. D. (2016). The effects of teaching songs during foreign language classes on students' foreign language anxiety. *Language Teaching Research, 20*(5), 638–653.

Dörnyei, Z. (2010). Researching motivation: From integrativeness to the ideal L2 self. In S. Hunston & D. Oakey (Eds.), *Introducing applied linguistics: Concepts and skills* (pp. 74–83). New York/London: Routledge.

Dose, S. (2013). Describing and teaching spoken English: An educational-linguistic study of scripted speech. (Doctoral dissertation). University of Giessen, Germany. Retrieved from d-nb.info/1068589043/34

Duff, P. A., & Zappa-Hollman, S. (2013). Using popular culture in language teaching. In C. Chapelle (Ed.), *The encyclopedia of applied linguistics* (pp. 5997–6002). Malden, MA: Wiley-Blackwell.

Ellis, R. (2015). Complementarity in research syntheses. *Applied Linguistics, 36*(3), 258–259.

Ellis, R. (2018). Meta-analysis in second language acquisition research: A critical appraisal. *Journal of Second Language Studies, 1*(2), 231–253.

Engh, D. (2013). Effective use of music in language-learning: A needs analysis. *Humanising Language Teaching, 15*(5). Retrieved from www.hltmag.co.uk/oct13/mart03.htm

Fadhli, A. (2017). The use of song worksheets to enhance EFL elementary school students' vocabulary mastery. *Journal of ELT Research, 2*(1), 37–43.

Farmand, Z., & Pourgharib, B. (2015). The effect of English songs on English learners' pronunciation. *International Journal of Basic Sciences and Applied Research, 2*(9), 840–846.

Ferreri, L., & Verga, L. (2016). Benefits of music and verbal learning and memory: How and when does it work? *Music Perception, 34*(2), 167–182.

Fisher, D. (2001). Early language learning with and without music. *Reading Horizons, 42*(1), 39–49.

Ginsborg, J., & Sloboda, J. (2007). Singers' recall for the words and melody of new, unaccompanied song. *Psychology of Music, 35*(3), 421–440.

Gonzáles Arteaga, A. J. (2019). *Can the use of English songs improve A1 EFL college students' listening comprehension?* (Unpublished master's thesis). University of Cuenca, Ecuador. Retrieved from http://dspace.ucuenca.edu.ec/handle/123456789/33060

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

Hadian, M. (2015). The use of song lyrics in teaching listening. *Journal of English and Education, 3*(1), 96–105.

Haghverdi, H. D. (2015). The effect of song and movie on high school students language achievement in Dehdasht. *Procedia – Social and Behavioural Sciences, 192*, 313–320.

Hahn, S. M. (1972). *The effect of music in the learning and retention of lexical items in German.* Education Resources Information Center. Retrieved from files.eric.ed.gov/fulltext/ED119455.pdf

Hazel-Obarow, S. E. (2004). *The impact of music on the vocabulary acquisition of kindergarten and first grade students.* (Unpublished doctoral dissertation). Widener University, USA.

Heidari, A., & Araghi, S. M. (2015). A comparative study of the effects of songs and pictures on Iranian EFL learners' L2 vocabulary acquisition. *Journal of Applied Linguistics and Language Research, 2*(7), 24–35.

Höfler, M. A. (2013). *Can't get it out of my head: The pop song in the EFL classroom and its effect on text recall.* (Unpublished master's thesis). University of Vienna, Austria. doi.org/10.25365/thesis.30258

Hua, C., & Li, B. (2015). Bringing fun and meaning into grammar learning: A case study of a secondary-level EFL class in Hong Kong. *Cogent Education, 2*(1), 1111039. doi.org/10.1080/2331186X.2015.1111039

Izzah, U., & Sukrisno, A. (2017). The effectiveness of using songs and dialogues to teach students' pronunciation. *English Education Journal, 7*(2), 179–193.

Javadi-Safa, A. (2018). Effects of using songs on adult EFL learners' vocabulary learning. *Journal of Applied Linguistics and Language Research, 5*(3), 101–112.

Jentschke, S. (2016). The relationship between music and language. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 343–355). Oxford: Oxford University Press.

Kao, T., & Oxford, R. L. (2014). Learning language through music: A strategy for building inspiration and motivation. *System, 43*, 114–120.

Kara, Z. E., & Aksel, A. S. (2013). The effectiveness of music in grammar teaching on the motivation and success of the students at preparatory school at Uludağ University. *Procedia – Social and Behavioral Sciences, 106*, 2739–2745.

Karlina, H., Raja, P., & Suka, R. G. (2017). Teaching tense through modified song lyrics for junior high school. *Neliti.* Retrieved from media.neliti.com/media/publications/213694-teaching-tense-through-modified-song-lyr.pdf

Kilgour A. R., Jakobson L. S., & Cuddy, L. L. (2000). Music training and rate of presentation as mediators of text and song recall. *Memory & Cognition, 28*(5), 700–710.

Köksal, O., Yağışan, N., & Çekiç, A (2013). The effects of music on achievement, attitude and retention in primary school English lessons. *Procedia – Social and Behavioral Sciences, 93*, 1897–1900.

Kuśnierek, A. (2016). The role of music and songs in teaching English vocabulary to students. *World Scientific News, 43*(1), 1–55.

Legg, R. (2009). Using music to accelerate language learning: An experimental study. *Research in Education, 82*(1), 1–12.

Lehmann, J. A. M., & Seufert, T. (2017). Can music foster learning? Effects of different text modalities on learning and information retrieval. *Frontiers in Psychology, 8*, 2305. doi.org/10.3389/fpsyg.2017.02305

Li, X., & Brand, M. (2009). Effectiveness of music on vocabulary acquisition, language usage, and meaning for mainland Chinese ESL learners. *Contributions to Music Education, 36*(1), 73–84.

Lieb, M. M. (2008). Music and listening: Learning gain without pain. In K. Bradford Watts, T. Muller, & M. Swanson (Eds.), *JALT2007 conference proceedings – Challenging assumptions: Looking in, looking out.* Tokyo: JALT. Retrieved from jalt-publications.org/archive/proceedings/2007/E082.pdf

Limbong, R. (2012). Enriching students' vocabulary using English pop songs. *Journal of English Teaching*, 2(3), 223–237.

Loewen, S. (2015). *Introduction to instructed second language acquisition.* New York/London: Routledge.

Lorenzutti, N. (2014). Beyond the gap fill: Dynamic activities for song in the EFL classroom. *English Teaching Forum, 52*(1), 14–21.

Ludke, K. M. (2010). *Songs and singing in foreign language learning.* (Doctoral dissertation). University of Edinburgh, UK. Retrieved from www.era.lib.ed.ac.uk/handle/1842/5500

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.

Ludke, K. M. (2020). Songs and music. In M. Dresman & R. W. Sadler (Eds.), *The handbook of informal language learning* (pp. 203–213). Malden, MA: Wiley-Blackwell.

Ludke, K. M., Ferreira, F., & Overy, K. (2014). Singing can facilitate foreign language learning. *Memory & Cognition, 42*, 41–52.

Maneshi, N. (2017). Incidental vocabulary learning through listening to songs. (Master's thesis). University of Western Ontario, Canada. Retrieved from ir.lib.uwo.ca/etd/4783

Medina, S. L. (1990, March). *The effects of music upon second language vocabulary acquisition.* Paper presented at the Annual Meeting of the Teachers of English to Speakers of Other Languages, San Francisco. Retrieved from files.eric.ed.gov/fulltext/ED352834.pdf

Mobbs, A., & Cuyul, M. (2018). Listen to the music: Using songs in listening and speaking classes. *English Teaching Forum, 56*(1), 22–29.

Norris, J. M., & Ortega, L. 2006. The value and practice of research synthesis for language learning and teaching. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 3–50). Philadelphia/Amsterdam: Benjamins.

Oswald, F. L., & Plonsky, L. (2010). Meta-analysis in second language research: Choices and challenges. *Annual Review of Applied Linguistics, 30*, 85–110.

Pant, H. A. (2014). Aufbereitung von Evidenz für bildungspolitische und pädagogische Entscheidungen: Metaanalysen in der Bildungsforschung [Representing evidence for political and pedagogical decisions: Meta-analyses in educational research]. *Zeitschrift für Erziehungswissenschaft, 17*(4), 79–99.

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

Pennycook, A. (2010). Nationalism, identity and popular culture. In N. Hornberger & S. L. McKay (Eds.), *Sociolinguistics and language education* (pp. 62–86). Blue Ridge Summit, PA: Multilingual Matters.

Paterson, A., & Willis, J. (2008). *English through music.* Oxford: Oxford University Press.

Purnell-Webb, P., & Speelman, C. P. (2008). Effects of music on memory for text. *Perceptual and Motor Skills, 106*(3), 927–957.

Racette, A., & Peretz, I. (2007). Learning lyrics: To sing or not to sing? *Memory & Cognition, 35*(2), 242–253.

Rahbar, S., & Khodabaksh, S. (2013). English songs as an effective asset to improve listening comprehension ability: Evidence from Iranian EFL learners. *International Journal of Applied Linguistics & English Literature, 2*(6), 63–66.

Reichelt, M. (2001). A critical review of foreign language writing research on pedagogical approaches. *The Modern Language Journal, 85*(4), 578–598.

Rezaei, M., & Ahour, T. (2016). The effect of listening to English songs on Iranian EFL pre-intermediate learners' listening comprehension. *The Journal of Applied Linguistics, 8*(16), 159–176.

Romero, P. X. (2017). Teaching and learning English through songs: A literature review. *Michigan State University Working Papers in Second Language Studies, 8*, 40–45.

Rukholm, V. N., Helms-Park, R., Odgaard, E. C., & Smyth, R. (2018). Facilitating lexical acquisition in beginner learners of Italian through spoken or sung lyrics. *The Canadian Modern Language Review, 74*(1), 153–175.

Schwarz, M. (2013). Learning with Lady GaGa & Co: Incidental EFL vocabulary acquisition from pop songs. *Vienna English Working Papers, 22*, 17–58.

Scott Langeland, A. (2013). Action research into the use of popular music: A goldmine worth exploring in the ELT secondary-school classroom? *Nordic Journal of Modern Language Methodology, 2*(1). doi.org/10.46364/njmlm.v2i1.88

Shaffer, D. E. (2004). The effectiveness of song-based EFL instruction at the tertiary level. *STEM Journal, 5*(2), 83–100.

Shen, C. (2009). Using English songs: An enjoyable and effective approach to ELT. *English Language Teaching, 2*(1), 88–94.

Smith, G. P. (2003). Music and mondegreens: Extracting meaning from noise. *ELT Journal, 57*(2), 113–121.

Smith Salcedo, C. (2010). The effects of songs in the foreign language classroom on text recall, delayed text recall and involuntary mental rehearsal. *Journal of College Teaching & Learning, 7*(6), 19–30.

Sposet, B. A. (2007). Sing, sing a song! A meta-analysis of programs and research on the use of music in second language acquisition. In C. P. Pascual (Ed.), *Revisiting language learning resources* (pp. 303–322). Newcastle, UK: Cambridge Scholars.

Sposet, B. A. (2008). *The role of music in second language acquisition: A bibliographical review of seventy years of research, 1937–2007*. Lewiston, ME: Edwin Mellen Press.

Summer, T. (2011). *An evaluation of methodological options for grammar instruction in EFL textbooks: Are methods dead?* Heidelberg: Winter.

Summer, T. (2018). An analysis of pop songs for teaching English as a foreign language: Bridging the gap between corpus analysis and teaching practice. In V. Werner (Ed.), *The language of pop culture* (pp. 187–209). New York/London: Routledge.

Summer, T. (forthcoming). *Pop culture in foreign language education.* Berlin: ESV.

Tamminen, J., Rastle, K., Darby, J., Lucas, R., & Williamson, V. J. (2017). The impact of music on learning and consolidation of novel words. *Memory, 25*(1), 107–121.

Tavakoli, M. (2013). The effect of task-based language teaching accompanied by songs on Iranian EFL learners' learning L2 grammar and their attitudes. *The Iranian EFL Journal, 19*(3), 112–135.

Tegge, F. (2015). *Investigating song-based language teaching and its effect on lexical learning.* (Doctoral dissertation). Victoria University, New Zealand. Retrieved from core.ac.uk/download/pdf/41339839.pdf

Tegge, F. (2017). The lexical coverage of popular songs in English language teaching. *System, 67*, 87–98.

Tegge, F. (2018). Pop songs in the classroom: Time-filler or teaching tool? *ELT Journal, 72*(3), 274–284.

Thompson, R. I., & Derwing, T. M. (2015). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics, 36*(3), 326–344.

Tomczak, E., & Lew, R. (2019). "The song of words": Teaching multi-word units with songs. *3L: The Southeast Asian Journal of English Language Studies, 25*(4), 16–33.

Tomlinson, B. (2017). Achieving a match between SLA theory and materials development. In B. Tomlinson (Ed.), *SLA research and materials development for language learning* (pp. 1–22). New York/London: Routledge.

Wallace, W. T. (1994). Memory for music: Effect of melody on recall of text. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(6), 1471–1485.

Werner, R. (2018). Music, movement and memory: Pedagogical songs as mnemonic aids. *TESOL Journal, 9*(4), 1–11.

Werner, V. (2012). Love is all around: A corpus-based study of pop lyrics. *Corpora, 7*(1), 19–50.

Werner, V. (2018). Linguistics and pop culture: Setting the scene(s). In V. Werner (Ed.), *The language of pop culture* (pp. 3–26). New York/London: Routledge.

Werner, V. (2019). Lyrics and language awareness. *Nordic Journal of Modern Language Methodology, 7*(1), 4–28.

Werner, V. (2021). Teaching grammar through pop culture. In V. Werner & F. Tegge (Eds.), *Pop culture in language education: Theory, research, practice* (pp. 85–103). New York/London: Routledge.

Werner, V., Lehl, M., & Walton, J. (2017). Pop lyrics and mobile language learning: Prospects and challenges. *International Journal of Mobile and Blended Learning, 9*(4), 33–48.

Werner, V. & Tegge, F. (Eds.). (2021). *Pop culture in language education: Theory, research, practice.* New York/London: Routledge.

Ziegler, N. (2016). Enhancing the use of music in language learning through technology. In A. J. Moeller (Ed.), *Fostering connections, empowering communities, celebrating the world: 2016 report of the Central States Conference on the Teaching of Foreign Languages* (pp. 59–78). Richmond, VA: Terry.