Songs and Singing in Foreign Language Learning

Karen M. Ludke

A thesis submitted in fulfilment of requirements for the degree of Doctor of Philosophy

The University of Edinburgh

2010

Declaration

I hereby declare that this thesis, submitted in candidature for the degree of Doctor of Philosophy at the University of Edinburgh, and the research contained herein is of my own composition, except where explicitly stated in the text, and was not previously submitted for the award of any other degree or professional qualification at this or any other university.

Karen M. Ludke, 16 July 2010

Abstract

Educators have claimed that listening to music in a second or foreign language (L2) can provide fun and motivating educational material and that singing can enhance the L2 learning process by improving listening and speaking skills, pronunciation, intonation, and vocabulary. Experiments have shown that under certain conditions, a sung presentation of linguistic material can facilitate verbal learning in the native language. To date, however, there is very little research evidence that singing can increase L2 skills. This thesis begins to methodically evaluate whether listening to songs and singing in a new language can facilitate L2 learning, compared to practising L2 material through more traditional, speech-based instructional methods. The research studies also explore the extent to which individual differences (IDs) between learners may mediate any observed benefits of using songs in L2 instruction.

The first two studies examine under controlled experimental conditions whether singing can support adults' beginning-level modern language learning compared to speech over a short time period. Results indicate that when no significant group differences exist for the ID measures, an instructional method that incorporates L2 singing can facilitate short-term learning and memory. Results also showed that IDs between learners, in particular previous language learning experience, musical abilities, mood, and motivation, can mediate the benefits of L2 learning through a singing method. The third study describes a four-week, classroom-based arts intervention exploring the effects of incorporating songs and dramatic dialogues into the L2 curriculum, both in terms of learning outcomes and the adolescents' opinions. In addition to increases in French skills, many children reported that the dramatic and musical activities had increased their confidence to speak in French. There was also an overall preference for listening to songs and more children reported that the songs repeated in their heads after class. The thesis concludes by discussing practical implications for L2 instruction and proposes a framework to guide future research exploring how and why singing can support modern foreign language learning.

Acknowledgments

I would like to thank my doctoral supervisors, Dr. Katie Overy and Professor Fernanda Ferreira, whose insights, encouragement, and support were invaluable. Their extensive knowledge of experimental psychology, the neuroscience of music, and psycholinguistics were a perfect combination for helping me to successfully navigate this complex research project.

I would like to warmly thank Dr. Krisztina Nagy, Zsuzsanna Kontra, and Lucinda Smallman for their time and help creating the stimuli for the two Hungarian language studies.

Thank you to Florence Bonacina, Damien Guillaud, and Kevin Hay for their help recording the French dramatic dialogue. A big thank you to Mary Dugan, Rachel Roberts, and especially to Sylvie Berthelot and the S1 and S2 French pupils at St. Thomas of Aquin's High School in Edinburgh. Thank you to Dr. Lara Crossland for helping me locate the school through the Researchers in Residence programme.

Thanks also to Justin Takichi Vitello, Suzy Moat, Madeleine Campbell, Ana Oliveira and her parents, Raquel Bennett and her brother and nephews Stanlley, Davi and Philippe, and Professor Peter Dayan for providing helpful resources for this project.

I wish to thank all of the teachers who have shared their practical ideas and resources for teaching modern languages through songs, especially Johanna van Gendt, Lisa R. Ludke, Sylvie Berthelot, Adelheid Kramer, and from the University of Edinburgh, Gareth Roberts, Cyprian Laskowski, Dorothea Gschwandtner, Derek Murphy, Florence Bonacina, Francesca Filiaci, Mara Mari Kirkwood, Ana Oliveira, Shelly Coyne, Mark Hamilton, June Chan, Janet de Vigne, JC Ng Shi Ing, Jillian A.E. Mathews, Hyejin Shin, Li-han Huang, and Liu Yi.

Thank you to Bethan Owen and to Dr. Bróna Murphy for their help in advertising the knowledge exchange workshop for modern foreign language teachers and Masters degree students in Edinburgh. I would like to warmly thank all of the teachers and teachers in training who participated in the two workshop sessions and made them so interesting and successful.

I would like to thank Tom Howey for his superb work, under tight deadlines, creating the beautiful workbook design for the knowledge exchange workshop with modern foreign language teachers. I also wish to thank Kirstin Anderson for her excellent suggestions about the workbook materials.

I never would have finished this thesis without the immense quantities of love and support I have received from my family and friends, especially from Michael, Lisa, Kurt, Juanita, and Roger Ludke. Thank you, Zack Moir, for being such a great officemate and for always knowing which microphone or equipment I needed, Hannah Cornish for our PhD reality check meetings to keep ourselves on track, Francesca Filiaci and Florence Bonacina for our lunchtime chats, and Gareth Roberts, Hae-Sung Jeon, Suzy Moat, Kirstin Anderson, Theresa Steward, Luis Sanchez, Melissa Avdeef, Mary Fogarty, Pieter Blue, Cyprian Laskowski, Olle Blomberg, Sebastian Andersson, Marian van der Meulen, Pit Péporté, Andrew Fenton, Nelly Lakestani, Kendra DeBusk, and Ian Smith for helping to keep me sane. A very special thank you is due to Patrick Power for all the years of support in the emotional, logical, and technical domains.

Music is a world within itself

With a language we all understand

With an equal opportunity

For all to sing, dance and clap their hands

– Stevie Wonder, ‘Sir Duke’

My doctoral research was supported by a College Award from the College of Humanities and Social Sciences at the University of Edinburgh, by funding from the Institute for Music in Human and Social Development (IMHSD) and Music in the School of Arts, Culture and Environment, by the Sir Richard Stapley Educational Trust, and by a Knowledge Transfer grant awarded to Dr. Katie Overy in October 2008 by the University of Edinburgh.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Learning and teaching are complex, interconnected and interdependent processes (Ritter, Nerb, Lehtinen, & O'Shea, 2007). There are many factors that can influence and support successful teaching and learning, particularly for challenging, time-consuming skills such as learning a foreign language. To increase the likelihood of meeting pedagogical aims, instructional techniques should aim to efficiently introduce and reinforce material while simultaneously engaging students' attention and increasing their motivation to learn. Further research into effective methods of teaching and learning is important because current instruction methods are often based on educational theories that may have limited evidence in support of their claims and recommendations. Using the findings of studies in psychology and neuroscience to inform educational practice is an exciting development, but it is important not to leap too quickly from the empirical findings to pedagogical recommendations without taking the intermediate step of substantiating whether new, 'evidence-based' instructional techniques are effective in the classroom; this is especially true of neuroscience, where the state of current knowledge is still in rapid development and most studies do not investigate the learning process (Stern, 2005). From a practical perspective, it is also important for any new pedagogical technique to work without requiring a great deal of extra training or preparation on the part of teachers, unless extra resources to support planning are available. Another important consideration for practice is the fact that each learner is unique and brings different experiences, abilities, and motivations to the classroom. In educational contexts, teachers must take into account the effects of individual learner differences and may need to make modifications to their teaching to enhance the learning process for their students.

Many acknowledge that knowing at least one modern foreign language is an important skill in an increasingly globalised world, but there is continuing debate about both when and how best to teach L2 skills to learners with different backgrounds, needs, and abilities (McColl, 2005). Despite pressure from other European countries (Garner, 2002), low levels of foreign language proficiency are often attained in areas where English is the majority language because many learners do not enjoy it and do not see any point in learning another language (Reilly, 2002), and institutions, administrators, and policy-makers do not consistently make L2 learning a priority (MacLeod & Ross, 2009; Schofield, 2006; Turner, 1974). The Scottish Executive decided that all pupils are entitled to modern language education from primary 6 and 7, which is being achieved through the *Curriculum for Excellence* for modern languages (HMIE, 2008). In England, many secondary school pupils chose to give up foreign language study after the previous government requirement was dropped, so the Department for Education and Skills (DfES) introduced a new foreign language requirement at a younger age (Key Stages 1 and 2; or 7 to 9 years of age) with the aim of improving continuation of L2 learning to Highers level and beyond (Garner, 2007; The Independent, 2007).

Is there reason to believe that music and songs could be a particularly effective way to support foreign or second (or third, fourth, etc.) language (L2) learning? For many years, educators have reported that songs can help students learn and practise a variety of L2 skills (Spicher & Sweeney, 2007; Fomina, 2006; Fomina & Merkulova, 2000; Medina, 1993). Teachers have also reported that songs can quickly set a positive classroom tone, improve L2 intake in the learning process, and attract learners' attention to the language contained in the song (Murphey, 1992). As a pedagogical tool, songs might be ideal for achieving modern language teachers' pedagogical aims, potentially providing positive affective, motivational, and cognitive benefits for students at different ages and stages of learning (Spicher & Sweeney, 2007; Ting, 2002; Medina, 1993; Rees, 1977). Songs also present challenging, authentic L2 material (Spicher & Sweeney, 2007). Listening to songs and singing in the new language may also result in the enjoyable repetition of language sounds and structures after class (Krashen, 1983; Smith Salcedo, 2002; Murphey, 1990). Teachers have written numerous articles and even books with suggestions and anecdotes illustrating the benefits of using songs to support classroom learning (Jensen, 2005; Murphey, 1992) and pedagogical techniques have been advanced which encourage the incorporation of music and song into the foreign language curriculum (Bancroft, 1995; Anton, 1990; Felix, 1989). But to date, no consensus about the possible benefits of including songs and singing in the L2 classroom can be drawn from the empirical evidence available, because many of the studies which have been conducted did not use very robust methods, and some results have been unclear or contradictory (Sposet, 2008).

The research studies presented in this thesis explore whether songs can support L2 learning in comparison to using more traditional auditory learning materials, such as spoken phrases or dramatic dialogues. In other words, can we find empirical support for the claim that songs can provide an effective means of teaching material in a new language? The secondary question under consideration is whether any benefits observed when using songs or singing to teach material in a new language depend upon or are influenced by particular individual differences between learners, or if songs can provide a useful method of presentation for a variety of L2 learners.

To answer these research questions, several methods of data collection were used to enable a deeper exploration of the factors involved in the use of songs and singing for modern foreign language learning. The effects of teaching L2 material through songs and singing are explored in two different language learning contexts: (1) experimental psychology studies with adult participants using randomised group assignment, controlled learning conditions, tests, and questionnaires; and (2) a quasi-experimental arts intervention study in the L2 classroom with secondary school pupils, using questionnaires and tests to measure L2 learning outcomes over a one-month period. Research conducted in a laboratory environment can offer strong evidence that certain materials or learning procedures are effective because it can reduce the number of unknown or uncontrolled variables. Questionnaires can provide another rich source of information, and classroom-based research can offer valuable, practical information for teachers. Previous research about the use of songs and singing to support language learning, along with the results of the three studies conducted for this research project, were later shared with modern foreign language teachers through an afternoon workshop developed to facilitate the exchange of ideas between this key group of stakeholders and the research community. The previous and current research findings were also used to develop a framework that can help direct future research explorations of whether and how songs might support L2 learning.

The next chapter provides a summary of the literature and previous work related to the topic of music, song, and singing in modern foreign language learning. The chapter discusses research studies originating in several different fields, ranging from education and linguistics to neuroscience and psychology.

Chapter 2

Music and Foreign Language Learning

This chapter provides a synthesis of previous research that has explored the use of music to support language learning. Since few empirical studies have examined the use of songs (music with lyrics) to support foreign and second language (L2) learning (Sposet, 2008), the research discussed in this chapter comes from a range of disciplines that were published in journals, dissertations, or at conferences during the past 50 years, but primarily drawing on more recent work. The major sections include: (1) connections between music and language, discussing findings from neuroscience; (2) short- and long-term memory for music and songs compared to memory for verbal material presented through speech, discussing research from experimental psychology; (3) education research, including teaching approaches that use music to support first and second language development, and teachers' observations and research findings when songs are included in instruction; and (4) the role of individual differences (IDs) which may influence learning when presenting material in a new language through singing. The IDs under consideration include: the learner's previous language learning experience and abilities (including phonological working memory); pre-existing musical training and abilities; mood; age; gender; and motivation and attitudes toward learning the new language.

2.1 Connections between music and language

Research findings from neuroscience can provide a picture of how music and language might be connected in the brain. If linguistic and musical material are processed, retrieved, or produced in similar areas or in a similar manner, this would support the idea of a mutually supportive relationship between music and language learning. This section explores whether the neural processing areas for music and language show significant overlaps, or instead take the form of distinct modules in the brain.

Brain imaging research has produced many gains in our understanding of memory and cognition over the past 30 years. There are a variety of types of neuro-imaging techniques available for use with human participants, and each type has its own strengths and weaknesses. An in-depth discussion of neuro-imaging techniques is outside the scope of this thesis, but the AAMCBI (2010) website[1] is a good source of additional information. Neuroscience research that examines music and language processing remains in the early stages, so there are limitations to the conclusions that can be drawn from this work. One issue is that it is not always clear what is happening in an area where a particular pattern of activation appears; for example, Lutz and Thompson (2003) found that activation patterns can be very different depending on task-related factors and depending on an individual's feeling of readiness for the next stimulus presentation – variation that would not have been observed using a method that averages together data across all trials, which is the more common manner of analysing data.

Aphasia is a condition in which sufferers are either unable to speak or to understand speech due to specific brain damage. In a similarly rare condition called amusia, 'lesions of certain auditory cortical regions result in...a highly selective problem with perceiving and interpreting music'; the dysfunctions observed for

[1] Web address: http://www.nmr.mgh.harvard.edu/martinos/research/technologies.php.

aphasia and amusia can result from stroke, tumour, or brain injury as well as congenital defects (Sloboda, 2005). Identifying the specific brain areas where damage has occurred in patients can provide one means of determining whether the neural structures for music and language may interact or overlap in the brain.

One study with an aphasic patient showed that the lyrics of a song can become dissociated from the song's melody, which did not happen with healthy control participants (Hébert & Peretz, 2001). Some aphasic patients are able to sing words, using the intact right hemisphere, after having a stroke which damaged the predominantly left-hemisphere language areas (Racette, Bard, & Peretz, 2006). Preliminary research on the effects of melodic intonation therapy has shown that the technique, where an aphasic patient sings commonly used words and phrases modeled by the therapist as he or she taps the syllables on the patient's hand, can improve patients' productive speaking skills (Overy, Norton, Ozdemir, Helm-Estabrooks, & Schlaug, 2004). A case study of a 76-year-old bilingual woman showed that her aphasia had affected both languages (English and Chinese), although her English skills were slightly better (Filley et al., 2006). Another case study with a trilingual man who had mild aphasia showed that the age at which each language was learned, the amount that each language was used (proficiency) before aphasia began, and the number of shared lexical items (or cognates) between the languages were the most important factors determining his ability to perform word translation tasks (Goral, Levy, Obler, & Cohen, 2006). When his aphasia was treated in his second language (English), preliminary results showed that his grammatical skills improved in both the language treated (English) and in his native language (Hebrew), but not in his third language (French); no significant changes in his generation of lexical items were observed for any of the three languages (Goral, Levy, & Kastl, 2007).

Another study investigated the effect of brain lesions on musical syntax using EEG and event-related potentials (ERP). The study involved individuals who each had damage to the 'pars opercularis of the inferior frontal gyrus (IFG)' or the 'anterior part of the superior temporal gyrus (aSTG)' on one side of the brain, and a control group (Sammler, Koelsch, & Friederici, 2006). While watching a silent film, participants listened to a series of chord sequences that either ended on a syntactically expected chord (the tonic) or an unexpected one ('dominant-to-the-dominant'). Preliminary results were in line with expectations, showing that the patients differentiated between the chords at levels above chance, but the EEG results showed that their brains had a reduced reaction to the chords with irregular syntax, a reduced ability to discriminate the chords compared to the control group, and lower musical ability test scores. Interestingly, the individuals with damage to the left hemisphere had more difficulty with timing-related tasks whereas those with damage to the right side struggled more with pitch processing. The latter group also showed brain responses consistent with the idea that a lesion in the IFG area on the right side would have a greater impact on performance with the unexpected chord sequences (Sammler et al., 2006).

Evidence from neuroscience with healthy individuals has provided support for the idea that musical and language abilities are linked, even at the neural processing level (Besson, Schön, Moreno, Santos, & Magne, 2007; Milovanov, Huotilainen, Välimäki, Esquef, & Tervaniemi, 2008). Rhythm processing and activation can be extracted to show lateralisation to the left side of the brain for most people, while there is a right hemispheric lateralisation in a similar location for melody and pitch processing (Sloboda, 2005), but when tone language speakers listen to lexical pitch contours that are meaningful in their language, processing shifts to the left side (Wong, Skoe, Russo, Dees, & Kraus, 2007). For most right-handed individuals (and the majority of left- handers, although to a lesser extent), Broca's area (BA44) in the right hemisphere is also where linguistic prosody is processed (Zatorre, Evans, Meyer, & Gjedde, 1992). Some evidence has also emerged showing that BA44 is also activated for coordinating hand actions (Fadiga & Craighero, 2006), suggesting that Broca's area is handling more general – not language-, music-, or action-specific –

tasks related to attempting to predict and understand others' actions and intentions, and thus is perhaps related to the mirror neuron system (Koelsch, 2009). Brodman's area, thought to be a processing centre of linguistic structure when a left lateralisation is observed, also processes musical structure more bilaterally (Levitin & Menon, 2003), and polyrhythmic music also activates this area (Vuust, Roepstorff, Wallentin, Mouridsen, & Østergaard, 2006). These results suggest that there is some brain lateralisation for pitch and for temporal processing in both music and language (Patel, 2007), but whether this can be further subdivided into 'linguistic' and 'musical' components is unclear.

One study showed that tone-deaf individuals have difficulty discriminating emotions conveyed by different pitch contours in speech when they were unable to rely on the linguistic information (Patel, Foxton, & Griffiths, 2005), whereas musicians perform better and more quickly than nonmusicians at perceiving small prosodic pitch incongruities (increased by 35%) in an unfamiliar foreign language (Marques, Moreno, Castro, & Besson, 2007). Other work has shown overlaps between the processing of musical and linguistic syntax (Sammler et al., 2009; Slevc, Rosenberg, & Patel, 2009; Bigand, Tillmann, Poulin, D'Adamo, & Madurell, 2001; Patel, Gibson, Ratner, Besson, & Holcomb, 1998; Patel, 1998), prompting Patel (2003) to propose a 'shared syntactic integration resource hypothesis' (SSIRH).

Music has long been acknowledged as very effective for expressing emotion (Feld & Fox, 1994), but incapable of expressing semantic meaning. While recent ERP studies have shown that when a semantically unrelated word was presented after participants heard a spoken sentence (or word) or a short musical excerpt (either 10 seconds or 1 second) as a prime, there was a very similar N400 observed after both types of primes, compared to no effect for words that were semantically related to the prime (Koelsch et al., 2004; Daltrozzo & Koelsch, 2009). Although it is clear that the musical primes were not expressing semantic meaning in the same way that language can (Steinbeis & Koelsch, 2008b, 2008a), this may indicate that expectancies, and violations of them, may play an important role in listeners' understanding of both music and language.

When musician participants were asked to focus on only the words or only the melody of a song, an ERP study indicated that activation patterns for the lyrics and the tunes were not processed in the same place in the brain (Besson, Faïta, Peretz, Bonnel, & Requin, 1998). However, recent ERP work has suggested that a sung presentation of a word list can improve 'coherence (phase-locked synchronization)' (Thaut, Peterson, & Mcintosh, 2005) and result in enhanced verbal memory for adults with multiple sclerosis when they sang back the words in sequence during the free recall tests (Thaut, Peterson, Sena, & Mcintosh, 2008). The authors hypothesised that even without rhyme, semantic, or syntactic structure present in the sung list of 15 words, music can provide a structure by improving the phase-locking of neuronal firing patterns in the brain. The authors claim that increasing the brain's efficiency during learning through 'musical chunking' could have a positive impact on initial memory encoding. Another study showed that verbal learning through a sung version increases 'frontal EEG coherence' compared to listening to speech, although participants in both conditions performed at similar levels on a verbal memory test after hearing the verbal material once during the learning period (Peterson & Thaut, 2007). Yasui, Kaga, and Sakai (2009) conducted an MEG study comparing hemispheric dominance in the auditory cortex for errors in music (melody) and language (lyrics) when participants listened to familiar and newly memorised songs. Results showed a right-dominant frequency mismatch negativity (M130) when participants heard a musical note that was different from the expected one, and a left-dominant mismatch (M140) for a change in lyrics.

Another line of neuroscience research suggests that having a positive emotional response when listening to music may help with memory formation. The amygdala is an ancient area of the brain, located near the brain stem, which has been implicated in emotion, learning and memory (Kleber, Birbaumer, Veit, Trevorrow, & Lotze, 2007). There is speculation that the role of the amygdala is most powerful for learning new material when humans experience a strong (preferably positive) emotion, which can be evoked (along with many other things) by music and pictures (Baumgartner, Esslen, & Jäncke, 2006).

2.2 Learning and long-term memory for music and songs

Understanding psychological theory that attempts to explain how memory and learning function may provide useful insights into whether songs could improve the efficiency of incorporating verbal input in a new language into long-term memory. The ability of music to serve as a mnemonic device for verbal material and a related phenomenon called involuntary mental rehearsal are other important topics that are discussed in this section.

2.2.1 Memory

The most widely accepted model of memory in psychology is one which divides memory storage into the working memory and long-term memory systems. Working memory consists of a temporary storage location which lasts for a few seconds, and a second component where input can be kept available for retrieval or further processing through active rehearsal. By contrast, long-term memory seems to be of immense size and to last indefinitely, although the information must be occasionally accessed to keep the pathways active for successful retrieval in future. Within long-term memory, many distinguish between ‘implicit’ (automatic or unconscious knowledge) and ‘explicit’ memory (learned or conscious knowledge). Procedural, declarative, and episodic memory are often considered to be special cases of long-term memory (J. R. Anderson, 1999).

Baddeley’s model of working memory proposes that this system is comprised of two underlying sub-systems, the visuo-spatial sketchpad and the auditory short-term store (also called the ‘phonological loop’), which are both controlled by the central executive system that selectively directs attention to different aspects of the information coming from the environment (Baddeley, 1990). There may also be other low-level working memory sub-systems. Empirical support has been found for some of the predictions of the working memory model’s two sub-systems, although it remains unclear whether there is a single ‘central executive’ capacity that filters input before sending it on for further processing in other parts of the brain.

Both working memory sub-systems consist of a short-term memory store, where incoming visual or auditory information lasts a short period of time and which is capable of containing 7 (plus or minus 2) elements (Miller, 1956). The number of components which can be stored at one time depends on the modality (auditory information can last for approximately 2 seconds without active subvocal rehearsal) and experience or training. Storage capacity in a particular modality can be increased through practice, such as by ‘chunking’ the information into larger meaningful units. For example, three nonwords could be created from a list of ten letters, since three elements are easier to remember than ten separate items.

It has been shown that the phonological loop is very important in learning to say new words and to learn words in a new language (Baddeley, Gathercole, & Papagno, 1998), most likely because it allows the learner to keep the sounds in the same sequential order in which the phones were originally presented. Research with aphasic patients has shown that it is impossible to learn words in a new language when the rehearsal component of the phonological loop has been damaged, as described in Schacter (2002).

At present, it is unclear where music fits into this model, although some have proposed that music might be processed as a ‘special case’ of the type of auditory input processed by the phonological loop because music can disrupt the phonological loop when participants are not allowed to use subvocal rehearsal (Salamé & Baddeley, 1989; Pring & Walker, 1994). However, memory for music and songs appears to be somewhat different than memory for verbal material, at least for nonmusicians (Twomey & Esgate, 2002; Snyder, 2001). Specifically, memory for unrehearsed musical stimuli does not appear to decay over the very short timescale observed for verbal material (Dowling, Tillmann, & Ayers, 2002). It is hoped that future research will find a place for music in the working memory model.

2.2.2 Learning, mnemonics, and involuntary mental rehearsal

Learning is defined as a process that causes a long-lasting or permanent change in knowledge, skill, attitude, or behaviour. In a monograph published in 1885, Hermann Ebbinghaus showed that learning a sequence of ten nonwords shows an exponential learning curve, with a sharp initial rise followed by slower increases until all of the material can be recalled. Without reviewing the material occasionally, over time it will be forgotten and this process follows an exponential curve that begins with a sharp drop and then the amount of information loss gradually slows over time. He also discovered that material can be overlearned when material is practised beyond the point that it can be accurately recalled; overlearning reinforces the information and makes it less likely to be forgotten over time (Ebbinghaus, 1885). Other work has shown the benefits of testing – specifically, the attempt to retrieve information and receiving feedback about whether the response was correct – for long-term memory of verbal material (Karpicke & Roediger III, 2008; Roediger III & Karpicke, 2006). It has also been shown that active learning can be more effective than passive learning (Iwata, 2005; Hannaford, 2005), although complex skills, such as dance moves, can also be learned through concentrated observation (Grafton & Cross, 2008), and sometimes behavioural rehearsal or practice does not improve learning (Colvin Clark & Mayer, 2008; Michael, 2006).

A learning process in which one item is learned by its association with another element that precedes it, such as pairing a word with its translation in a new language, is called associative learning. Related to associative learning is rote learning, which encourages frequent verbatim repetition during the learning process to better embed the material in memory for later retrieval. Rote learning has been criticised for being inefficient and for leading to a shallow level of learning because the learner does not have to understand the material being taught; the mere ability to accurately recall the information is proof of successful learning. While it need not be rote learning, in most formal learning situations the repetition of input (e.g., from listening to the teacher, reading a textbook, and taking written notes) and active rehearsal or practice on the part of the learner is generally required for the integration of information into long-term memory. Most adults will spontaneously try to use a mnemonic strategy to remember information that is to be learned by rote (Bellezza, 1981).

A mnemonic is a learning strategy that aims to aid memory formation and later recall. Mnemonic devices work by creating multiple associations between the information to be learned, increasing the number of pathways which can be searched and thereby increasing the likelihood of retrieval and recall. There are many strategies for efficiently memorising material, such as mnemonic techniques that incorporate visualisations and real or imagined locations, movements, gestures, rhythm, and rhymes. Combining more than one retrieval cue for memory, such as using illustrations, rhyme, and music, can be more effective than using one type of input in isolation (Ziegler, 2007; Medina, 1993). However, there is also concern from some psychologists and educators about the use of mnemonic strategies that do not necessarily have much relation to the content being learned (Bellezza, 1981). The idea that mnemonic devices only enhance surface-level processing and memory at the expense of deeper, semantic processing is potentially an important issue.

Poetry was used to pass down cultural histories and stories from one generation to the next, and the material was usually sung rather than spoken (Racette & Peretz, 2007; Calvert & Tart, 1993; Rubin, 1995). Stansell (2005) reminds us that the Muses in ancient Greek mythology were the daughters of Mnemosyne, a Titan whose 'main concern was the human memory', and seven of her nine daughters used their powers to inspire 'epic poetry, lyric poetry, sacred poetry, love poetry, comedy, tragedy, and history.' Musical mnemonics have been shown to support verbal learning in the native language (Thaut et al., 2005; Gfeller, 1983). Other studies have shown that mnemonic techniques can dramatically improve vocabulary learning in a new language (see Bellezza, 1981 for more details).

In addition, there is a phenomenon called involuntary mental rehearsal, where verbal or musical material seems to become 'stuck' in the auditory short-term store. This phenomenon seems to be particularly associated with auditory or linguistic stimuli (e.g., after listening or reading), since memory traces for smells, images, or other visual information that repeats off and on for hours (or even days or years) after the original experience, have not been described in the literature.

2.2.3 Psychological research evidence

Under certain conditions, research has shown that short-term verbatim recall for words in participants' native language can be facilitated when words are paired with music during learning, compared to a spoken version of the same material. We turn now to a discussion of research in experimental psychology for the insights it can offer about memory for music and songs.

Morrongiello and Roes (1990) found that preschool-aged children show some integration of words and melody for learned songs, rather than revealing independent storage of the two elements. This integration was stronger for adults than for the 4-year-old children, which suggests that the storage of song lyrics along with the accompanying tune might be a learned strategy for efficiently enhancing memory for both elements of the song. Another study found that adults were much better at recognising the song, melody, or text of unfamiliar folk song extracts when they heard the original words and tune; melody recognition was very poor when heard without the original words (Serafine, Crowder, & Repp, 1984). In addition, Hébert and Peretz (1997) found that adults were better at identifying a song using only the melody (with notes presented isochronously) compared to hearing only the rhythm, but their performance was best when both melody and rhythm were presented together. Thus, combining pitch with the rhythm of phrases in a new language might make verbal material that is presented through a song more memorable over time.

Dowling et al. (2002) conducted a series of experiments which showed that memory for surface level details of short instrumental music passages does not follow the same pattern of decay as other types of learning. However, this effect depended on the participants' expectation that they would be tested on the material; when there was no expectation of needing to recall the passage of music, the effect disappeared. The results of experiments comparing verbatim recognition for short auditory and read passages of text also showed that memory for surface-level detail is better for poetry than for prose in the native language (Tillmann & Dowling, 2007). Interestingly, no difference in performance was observed between the poetry and prose presentations when participants were distinguishing between the text's meaning (original vs. change in meaning) compared to verbatim recognition of the verbal material.

One study found that learning a word list of proper names which was sung to a familiar melody resulted in better long-term memory and quicker relearning of the word list after a 1-week interval, compared to a spoken presentation. However, results of this study showed no benefit for learning speed or memory in the music condition during the initial learning session (Rainey & Larsen, 2002). Another study found an advantage for adults' verbal memory of the Preamble of the Constitution of the United States when the words were learned through a song accompanied by an animated video compared to a spoken version with the same video, both short-term and long-term (after five weeks); again, no benefit for the sung version was observed after the first presentation of the words (Calvert & Tart, 1993).

A statistical learning paradigm developed by Saffran et al. provided evidence that after a 21-minute passive listening session, both adults and infants were able to segment a continuous stream of isochronous sounds into 3-syllable words (1996) and 3-note musical phrases (1999) by implicitly learning where the boundaries were using the statistical probabilities present within the sound stream. This design was later adapted to compare learning for spoken versus sung auditory stimuli with French-speaking adults (Schön et al., 2008). During a 7-minute training period, participants listened to a continuous stream of synthesised French sounds which had been combined to form 3-syllable nonsense words. The presentation methods were speech (monotone pitch for all syllables), consistent song (each syllable always had the same musical note), and variable song (the first pitch of each 3-syllable word was not consistently paired with the first syllable). At test, participants had to decide whether a sequence of syllables constituted a word. Results showed that participants in both melodic conditions were able to learn the word boundaries, although less successfully in the variable song condition, while the speech (monotone pitch) condition only performed at chance levels. The authors suggest that since pairing syllables with a consistent melody can lead to quicker word segmentation of the sound stream, songs might be particularly helpful during the beginning stages of L1 and L2 learning.

A series of four experiments conducted by Wallace (1994) were among the first to provide support for the facilitative effects of music for verbal memory in the native language. During five listening sessions, adults heard three verses of previously unfamiliar folk ballads. Their memory for the words was tested using a written free recall task after the first, second, and fifth listening sessions. Results showed that pairing the lyrics with one melody during the learning process provided the highest performance, using several analyses of verbatim text recall. The music condition was significantly more effective for verbatim written recall than the spoken voice condition (poetry), and more effective than pairing the rhyming words with a rhythm but without the tune. However, pairing a different tune with each verse of the ballad was less effective than hearing the spoken version. Wallace hypothesised that when the three verses had three different tunes, the frequently changing music served as a distraction rather than helping participants memorise the verbal material. She concluded that using songs with a symmetrical, simple, repeated pattern can facilitate music's benefits for verbatim text recall in the native language (Wallace, 1994, p. 1483-1484).

However, other studies using similar designs have produced contradictory evidence. Kilgour, Jakobson, and Cuddy (2000) found no memory advantage for the sung presentation of verbal material compared to a spoken version, after controlling for the rate of presentation and total duration of the auditory stimuli. While the results showed no significant effect based on the listening condition, Kilgour et al. did find a significant difference between musicians and nonmusicians, starting at the second test (after participants heard the stimuli twice), but not after the first listening session. An important difference between the studies mentioned earlier and this experiment is that Kilgour et al. measured learning using a written, multiple-choice test with questions about the meaning of the text, rather than assessing verbatim recall. A benefit for the music condition would not be expected using this measure, because recognition tests do not reliably distinguish between learning conditions – music shows larger benefits for verbatim recall tasks (Yalch, 1991; Tillmann & Dowling, 2007). Nevertheless, it is important to control for the rate of presentation and duration of stimuli in the different listening conditions when testing whether music may have facilitative effects for memory, since a slower rate of presentation in the music condition might explain any observed learning or memory facilitation.

Another study investigated whether these findings, which used measures of written text recall, would apply for oral recall in the native language. The authors argue that oral recall should be easier and more natural than producing written text recall after hearing auditory input, since speaking or singing in response to an auditory stimulus does not require as large a shift from one modality to another (Racette & Peretz, 2007). Native French speakers listened to three unfamiliar French folk songs which were either spoken or sung *a capella* and each participant learned the verbal material through the three presentation and testing methods: sung-sung (hearing a song and singing at test), sung-spoken (hearing a song and reciting the lyrics at test), or a 'divided' (control) condition where the lyrics were also recited at test. For the stimuli in the 'divided' condition, the singer quietly sang the melody for each line on 'la' in the background, to ensure that the total duration of the spoken and sung auditory stimuli were equal; however, in this control condition, the overall duration of the spoken words was significantly shorter than the duration of the sung lyrics ($p < .01$). Participants initially heard all eight lines of a folk ballad (spoken or sung) and then listened once to the first line and were tested by orally repeating the line (either by reciting or singing). Performance was assessed on-line by the lead researcher. If performance was high enough, the participant would hear the first two lines and repeat them back, followed by hearing the next two lines and repeating all four lines aloud, and so on, up to a possible eight lines. If the participant reproduced less than 80% of the words accurately during a test, the researcher would stop the learning session and proceed to the next song.

Results did not show a facilitative effect for the sung-sung presentation when compared to the spoken presentation or the sung presentation with spoken recall of the lyrics. In fact, the opposite result was found, with participants showing better immediate, short-term (after 20 minutes) and long-term (after several months) recall for the words in the 'divided' spoken condition. In addition, no difference in performance was observed between the trained singers and nonmusicians recruited for this study; verbal recall was lower when participants had to sing the lyrics at test. The authors conclude that memorising the words of a song prior to (or separate from) learning the melody will result in better short- and long-term retention.

A possible explanation for this unexpected result is that the folk songs used by Racette & Peretz did not meet the requirement of having a symmetrical, easy-to-learn melodic structure which Wallace (1994) reported would be facilitative for verbal memory. In addition, since participants heard a few lines of the song and immediately tried to say or sing them back after only hearing the song once before, the test was likely measuring participants' working memory, rather than learning. If so, it is plausible that the shorter duration of the spoken lyrics (which were embedded in the middle of each line of the 'la' melody) made them easier to rehearse in working memory before being asked to say the words aloud. This explanation fits with the findings of Kilgour et al. (2000) that a performance advantage for musicians was only observed after the second time hearing the song, not after the initial exposure.

There are fewer studies of longer-term recall, but findings generally support the claim that songs can support memory for material in the native language. One study found an advantage for recognition of a sung (vs. spoken) advertising slogan paired with a product, but only for a difficult recall task; no effect was found for an easier recognition task (matching the slogans with the products) (Yalch, 1991). Another study investigated nonmusicians' long-term memory for lyrics and pitch with overlearned popular songs chosen by the participants and found that most of their sung productions were very accurate, especially for the first verse and the chorus

(Levitin, 1994). A third study measured the consistency of participants' tempo judgements using popular songs such as 'Yesterday' by the Beatles and a Greek dance piece which had appeared in a recent film (Lapidaki, 1996). The hypothesis was that familiarity with a certain style of music might increase participants' accuracy when making tempo judgments. Results showed that accurate tempo judgments were highly influenced by the initial tempo at which the subjects heard the piece, and was not significantly influenced by familiarity with the style. However, familiarity with the style of music did improve the subjects' consistency across trials; and the best performers, in terms of consistency, were both composers (Lapidaki, 1996). Calvert and Tart (1993) conducted a naturalistic study to investigate adults' long-term memory (after ten years) for verbal material in the native language that was originally learned through a song about the Preamble of the US Constitution on 'Schoolhouse Rock.' Results showed that verbal recall was significantly higher than recall for the text among participants who reported they had infrequently or never watched 'Schoolhouse Rock.'

To summarise, while some studies have shown that music can facilitate verbal learning and memory in the native language, counter-examples can also be found in the literature. Studies which have found benefits for musical presentations of verbal material suggest that the benefit will be greatest for verbatim memory tasks (Tillmann & Dowling, 2007; Dowling et al., 2002; Wallace, 1994; Yalch, 1991). The benefits of a sung presentation may disappear when the rate of stimulus presentation is carefully controlled (Kilgour et al., 2000), especially if the test does not require verbatim recall (Yalch, 1991; Kilgour et al., 2000). A sung presentation can even be detrimental for verbal learning and memory when the song's melodic and rhythmic structures are difficult to learn (Racette & Peretz, 2007; Wallace, 1994). The rate of stimulus presentation, the overall duration of stimuli, and the song's complexity are other important factors which can influence verbal learning and memory through song. To date, there is no consensus amongst psychologists regarding whether learning verbal material along with a melody can provide benefits for memory, whether in the native language or in a new language.

2.3 Education research

This section discusses pedagogical approaches and research findings from education and applied linguistics when music and language learning are integrated in the classroom. First, pedagogical approaches and techniques that encourage the use of music in language teaching (and vice versa) are outlined. Next, teachers' observations of the benefits of music to support language skills are presented. Finally, classroom-based research studies exploring the use of music and songs in the modern foreign language classroom are discussed.

2.3.1 Educational approaches for L2 learning and music

In the field of second language education, there is considerable debate about when and how language skills are learned, and different viewpoints have ramifications for the way in which new languages should be taught. Some have claimed that the order of foreign language instruction should follow the same natural sequence as when learning a first language: listening, speaking, reading, and writing. For example, in the Direct Method, the teacher should never use the students' native language in the classroom, instead using the context, gestures, songs, and visual aids (e.g., realia) to facilitate communication, and to teach grammar inductively rather than teaching rules; for older students reading and writing skills are also taught from the beginning level, but this is secondary to aural/oral communication practice (Larsen-Freeman, 2000).

The audiolingual method is an oral approach to L2 learning which was introduced during the 1950s. This teaching approach encourages students to listen, repeat, and memorise increasingly complex phrases, sentences, and dialogues in the target language (Larsen-Freeman, 2000). For practitioners of the audiolingual method, the skills of listening, speaking, reading, and writing should strictly be taught in that order. The goal is for learners to practise perfectly pronounced and grammatical sentences in the new language from the very beginning, so that producing sentences in the new language becomes automatic. However, with misconceptions focusing on the way this method encouraged only rote memorization, resulting in a lack of opportunity for self-expression and creativity, or learning from mistakes (Rivers, 1964), the audiolingual method was disparaged by many researchers and language teachers (H. D. Brown, 2000). However, Wong-Fillmore (1976) and others (R. Mitchell & Martin, 1997; Castagnaro, 2006) have argued for the benefits of this technique because formulaic speech, such as learning certain phrases verbatim, plays a crucial role in foreign language learning. Wong-Fillmore claims that the learned language chunks, rather than individual words, gradually become part of the learners' developing linguistic system, and later these preformulated parts can be combined in ways that do allow creative expression. Thus, learning through songs and related activities in the L2 classroom or language lab might provide a great way to reinforce new L2 material in an enjoyable way, because songs often become more appealing as they become more familiar (Szpunar, Schellenberg, & Pliner, 2004).

Lozanov (1978) developed an approach for teaching the early stages of a new language which results in accelerated vocabulary learning. This approach uses music to help set the mood of the classroom and promote the relaxed conditions necessary for linguistic input to become intake. Suggestopedia (also called Desuggestopedia and Reservopedia) involves the teacher playing a piece of Baroque music while he or she reads 4-second chunks of a text aloud in the new language, varying the rhythm and prosody of the words with the music. This passive listening time is followed by a period of listening to relaxing Classical music to enhance integration of material into long-term memory, thereby improving retention because of decreases in anxiety and other negative emotions that could otherwise block memory formation (Lozanov, 1999). An empirical study showed a significant increase in vocabulary learning using this technique, compared to using traditional methods (Felix, 1989). Language teachers and learners wishing to enrol in courses using Lozanov's accelerated learning technique are screened before their attendance is confirmed (and it is not revealed on what basis applicants are excluded) at his training institute in Vienna. To date, the official technique seems to have only been applied to L2 teaching at the beginning and intermediate levels, and some have criticised his use of 'pseudo-science' to support his often sound pedagogical techniques and recommendations (Scovel, 1979).

Krashen (1982) developed an influential theory of second language acquisition called 'The Natural Method,' which has five key hypotheses:

1. There is a sharp division between material that has been 'acquired' (implicit knowledge) and 'learned' (explicit knowledge), and that learned information can never become fast and automatic the way acquired knowledge can be retrieved.

2. Grammatical morphemes will be learned in a particular order when learning a particular language.

3. The 'learned' information system monitors the language learner's output or production (which was 'acquired') and self-correction after an utterance is performed by this self-monitor – this is the only importance granted to 'learned' information.

4. To reach the next stage of acquisition (i + 1), the *meaning* of new material must be 'comprehended' rather than asking the individual to focus on the form (which would result in learning).

5. There is an 'affective filter' that can interfere with and decrease the amount of information a learner is able to take in when they feel anxious about their abilities in the new language.

Of most interest in this research project is Krashen's later 'Din in the Head hypothesis,' which claims that when an individual reads, writes, or hears words and phrases in the new language which later repeat in the learner's head ('din'), this is proof that acquisition is taking place because these new words are being integrated into memory through the help of the 'language acquisition device' (Krashen, 1983).

Anton's Contemporary Music Approach (1990) encourages teachers to incorporate songs into the modern language classroom for grammar practice for beginning-level university students. The method involves a 3-week learning period for each Spanish song, all of which were created by the author (who is also a musician) to accompany a variety of grammar lessons. During the first class period when the song is introduced, the teacher spends approximately 30 minutes helping learners understand the grammar and lyrics of the song. Approximately 5-10 minutes are spent on the song during subsequent class periods for each two-week Song Introduction period. For the second phase, the students have one week (as homework) to write their own lyrics using the same grammar forms that were in the original song. To write new lyrics, students must review a large variety of L2 material, including vocabulary, grammar, adjectives, and idiomatic phrases.

In addition, two music education methods use the native language to teach musical skills to very young children. The Kodály method is a means of teaching music through folk songs in children's native language. Language and music are structured together during instruction to make each more accessible to young children as they proceed through stages of development (Choksy, 1998). The Suzuki method (also called the 'Mother Tongue Approach') is based on ear training and learning to play a musical instrument primarily through listening and imitation, rather than reading a musical score (Liperote, 2006). The Suzuki method has also been adapted for use with singers and some versions also include musical literacy.

Finally, Gardner's theory of multiple intelligences has also had a large influence on educational recommendations and practice (H. Gardner, 1983, 1999). This theory proposes that musical intelligence is one of eight types of intelligence that educators should be aware of as they teach their students any subject, because people learn in many different ways and some individuals are more receptive to or have stronger abilities in certain areas than in others. Because of similarities in input between musical and linguistic structure, he claims that musical intelligence is closely tied to verbal-linguistic intelligence as well as others, such as logical-mathematical and bodily-kinesthetic intelligence (H. Gardner, 1983). While there is evidence for an underlying general intelligence, *g*, upon which these other intelligences or abilities depend (Visser, Ashton, & Vernon, 2006), it is nevertheless useful to teach using a variety of pedagogical materials because every individual has different abilities and interests that they bring to the L2 (or any other) learning context.

2.3.2 Educators' observations of the benefits of music

In the modern language classroom, educators have recommended the use of songs as a way to reinforce important material in the target language, including listening and speaking skills, vocabulary, grammar, pronunciation, rhythm and intonation, and cultural understanding (Lê, 1999; Shtakser, 2001; Adkins, 1997; Murphey, 1992; Felix, 1989; Fomina & Merkulova, 2000; Fomina, 2006). One teacher recommended the use of L2 Russian songs in teaching phonological skills (Tumanov, 1986). Music teachers have also cited the appeal of learning songs in new languages for pupils in a language immersion context, using a variation of the Kodály method (N. Brown & Lamb, 2004).

Researchers have argued that rhythm, prosody, and intonation – in other words, the musical features of language – are essential to the native speaker's comprehension of speech (Parker, 2000) and that mastery of prosodic features (as opposed to perfect pronunciation of particular sounds or syllables) is crucial for communication in a foreign language (White, 2005). And as Smith Salcedo (2002) writes, 'Fluency in the use of the language was one of the most valuable contributions of songs, according to Bartle (1962), who believed that: "some songs lend themselves to the incidental revision of grammatical points or of verb tenses. Songs are a definite advantage in memorization of phrase constructions. They are more easily learned and tend to 'stick' longer than straight-out grammatical examples" (p. 11).' (p. 76).

However, when music and language are combined into song, clashes can occur between the rhythm and melody of the music versus the natural stress pattern and intonation of those words and phrases in the language (Arnold & Jusczyk, 2002; P. Martin, 2004), and the effect of these clashes on learning and memory have not been adequately explored (but see Fomina, 2000a). From a practical perspective, teacher-researchers have claimed that traditional music and folk songs preserve most features of both the music and language of a particular culture, following the natural intonation, stress, and rhythmic patterns of the language and so folk songs might be the best musical material to use in the L2 classroom (Spicher & Sweeney, 2007; Fomina, 2000). This claim has been substantiated through a comparative analysis of instrumental music created by English- and French-speaking composers, which showed that the melody and rhythm of their music closely followed the natural rhythms of their native languages (Patel & Daniele, 2003). In addition, an experimental study showed that participants are capable of categorising instrumental music using their knowledge of speech rhythm to decide the language from which a piece of music originated (Hannon, 2009). However, it is important not to use only folk music, because this could reinforce stereotypes that L2 learners may have about speakers of the target language (Failoni, 1993). The best recommendation is to choose songs from a variety of genres, including famous songs that everyone from that culture would know.

In some ways, 'din' for material in the target language seems very similar to involuntary mental rehearsal, which was discussed previously. Murphey (1990) claims that songs are very effective at triggering 'din' in the second language, which he terms the 'song-stuck-in-your-head' phenomenon. Lake (2002) believes this benefit of songs is substantial for his ESL learners, claiming that '[a]dding rhythm and melody to chunks of language invites rehearsal and transfers words into long-term memory.' Another related phenomenon is called 'earworms' – the involuntary repetition of music, with or without lyrics, in your mind (Kellaris, 2003). Typical attributes of musical earworms are:

- The structure of the song is simple and predictable (Kellaris, 2003, 2001; Wallace, 1994)
- The melody line is symmetrical, often consisting of a rise in pitch, followed by a fall (Wallace, 1994)
- The end of the song or phrase leads naturally back to the beginning (Kellaris, 2003)
- The rhyme scheme of the lyrics is important because it can facilitate recall (Wallace, 1994; Rubin & Wallace, 1989).

All of these characteristics seem to enhance learners' memory-based predictions. In essence, if the learner can more easily predict what word will come next because they remember the melody of the song, that may in turn improve retrieval of linguistic material which is paired with that melody.

Another suggestion about the potential of music to support learning is that listening to music is aesthetically appealing and enjoyable, which can lead to increased emotional arousal and engagement with the material to be learned, which can then lead to temporary enhancements in learning or spatio-temporal IQ

(Schellenberg, 2006). Listening to the words of a song in a new language can make learning the material seem effortless, energising, and fun; and this, in turn, might increase the student's motivation to learn. Indeed, one teacher recommended the use of rap music with a positive message in his urban elementary school classroom both to motivate and teach pupils, including for history and science lessons (E. Anderson, 1993).

2.3.3 Educational studies using music to support language skills

Music and songs can provide powerful support for short- and long-term memory of verbal material in the native language, using both naturalistic and experimental research designs (Calvert & Tart, 1993; Rainey & Larsen, 2002). Music listening, musical activities and games have also been shown to support literacy (Gfeller, 1983),[2] vocabulary (Ziegler, 2007; Schuster, 1985), and spelling and phonological skills (Overy, 2003; M. Martin, 1983), both for typically developing children and for children with special needs. Several classroom-based studies have also found benefits for music when teaching verbal material in the native language (Foster, Kersh, & Masztal, 1999; Albers & Bach, 2003; Ahlkvist, 2001; Ziegler, 2007; Walczak & Reuter, 1994).

A few research studies have reported benefits for modern foreign language vocabulary learning when material is presented through music. One study found that second-grade Spanish-speaking children with low English proficiency learned more English vocabulary from a story when it was presented through a song compared to when the story was read aloud over a two-day period (Medina, 1993). In this study, the presentation method which combined illustrations with a song produced the largest vocabulary gains. A project with kindergarteners in Singapore showed that

[2]For a review of some of the literature on music and L1 reading skills, see Chang (2000).

integrating weekly music sessions (partly based on the Kodály method) into the language arts curriculum for one year were very effective for improving the children's oral skills in English, as well as their musical and social skills (Gan & Chong, 1998).

Iwata (2005) showed that active participation by music students who repeated foreign language phrases aloud during the learning process was more effective than merely listening while watching the teacher and another student sing and sign. The learners who were actively engaged in the learning process also had a more positive emotional state at the end of the experiment session. Another study with English as a Second Language at the primary school level, with four treatment groups with 20 children each, showed that while all four groups improved significantly from the pre-test to post-test ($p < .01$), the singing-with-signs, speaking-with-signs, and singing-without-signs conditions had a significant beneficial effect on children's learning of 20 vocabulary words compared to the speaking-without-signs (control) condition ($p < .05$), although the difference between the speaking-without-signs condition and the singing-without-signs condition was not significant (Schunk, 1999).

Hearing L2 material presented through songs might increase the incidence of 'din,' where words repeat in the learner's mind and reinforce the material to be learned. The hypothesis that using songs for foreign language learning would result in increases in L2 performance and also in higher reported experiences of 'din' compared to speech was tested in a classroom-based study with four classes of adult, beginning-level Spanish learners (Smith Salcedo, 2002) over the course of one academic term. Results showed that students' performance on Cloze tests was significantly better in the two classes which had learned the new Spanish material through songs, compared to a class that heard the same material presented as poetry and a control group which never heard the material before the test. There was no performance benefit observed for the class that heard the song (without the words) during the test. In addition, experimental evidence in the native language suggests that songs are more likely to involuntarily repeat in the learner's mind after listening than do the same words presented as speech, $p < .05$ (Calvert & Tart, 1993).

In sum, as Sposet (2008) concludes in her review of the use of music in second language education over a 70-year period, 'The data suggests that while the quantity of studies purporting a positive outcome (i.e. 15 of 23), almost three-quarters (75%) of those studies can be classified as non-experimental. Conversely the majority of the studies indicating a negligible or negative outcome can be classified as experimental [5 were experimental and 3 were non-experimental studies, with some conducted in the L2 classroom and some in the laboratory]. Until such time as the studies using a non-experimental approach can be replicated with a control and experimental group, it is the opinion of this researcher that there is no definitive conclusion as to the value of the positive or negative impact of the use of contemporary music on second language acquisition' (p. 90). Although it would have been more useful had she chosen to include only experimental and quasi-experimental studies in her review if only empirical results would convince her, Sposet nevertheless offers a valuable recommendation to researchers by calling for more controlled studies, and also acknowledges that there may be emotional and motivational justifications for including music and songs in the L2 curriculum.

2.4 Individual differences relevant to L2 learning through songs

It is important to consider the extent to which individual differences (IDs) might influence learning, especially when exploring the effectiveness of a pedagogical technique for learning material in a new language through songs or singing. Every learner is different, and it is important to to establish whether participants assigned to different learning conditions are well matched for factors that might influence L2 learning through musical presentation methods. This would provide evidence that any differences in language learning observed between the experimental groups were due to the learning condition, rather than other factors that could explain the variance in overall results, because otherwise different variables that also influence modern foreign language learning could influence or even confound the findings of the studies in this research project. The individual differences which are most relevant to second language learning through songs are discussed in this section, and include: language learning ability and prior experience; musical ability and training; mood; handedness; gender; age; and motivation to learn the new language.

2.4.1 Language learning experience and abilities

Gathering information about learners' previous language experience is important because this factor could influence the results of L2 tests, particularly because there are correlations between L1 and L2 reading abilities (Abu-Rabia, 2004), which may also exist in other language areas. It is logical that learning another language would increase future language learning abilities because learners can develop more effective strategies for L2 learning over time. Some differentiation of brain activation has been shown for bilingual and multilingual speakers when they use one language as opposed to another (Halsband, 2006), although an ERP study investigating word processing with both bilinguals and second language learners suggests that frequency of language use (not only proficiency level) is also an important factor (Midgley, Holcomb, & Grainger, 2009). Another study showed that early bilinguals have an advantage for learning skills in new languages, including reading (Sagasta Errasti, 2003), at least when they attain and keep a high level of competence in both languages (Thomas & Collier, 2002).

Learners' underlying ability or aptitude for language learning is another important variable to consider because it could have a large influence on test results, particularly in a brief language learning study (Gilleece, 2006; Skehan, 1989). Language aptitude is often assessed by measuring certain underlying abilities that combine together to predict success in language learning, using measures such as the Modern Language Aptitude Test (MLAT) which was developed by Carroll & Sapon (1959) or the Pimsleur Language Aptitude Battery (PLAB), designed for use with secondary school students and which also includes motivation as a factor (Pimsleur, Reed, & Stansfield, 2004).

Evaluating participants' phonological working memory was another important measure in the experimental studies because this type of working memory correlates with both first and second language learning outcomes (Kormos & Sáfár, 2008; Gathercole, 2006; Abu-Rabia, 2003; Masoura & Gathercole, 1999; Baddeley et al., 1998) and can provide a rough measure of verbal IQ (Alloway & Alloway, 2010; Gilleece, 2006).

2.4.2 Musical skills and experience

It is also important to consider learners' musical abilities and prior musical training, because musical skills have consistently been shown to correlate with verbal skills, both in the native language and in a second language. Gilleece (2006) found that even after controlling for the effect of nonverbal intelligence, small to moderate positive correlations existed between 11-year-old children's musical and linguistic abilities, both for receptive and productive skills in these domains. Slevc and Miyake (2006) found that pronunciation and speaking ability in a foreign language (English) correlated highly with musical ability. However, an unequal number of men and women were involved in the study (41 of the 50 total participants were female), which might have had an influence on results. In addition, there were few measures of musical ability in comparison to the number and variety of language tests. Another study investigated the role of early music training on adults' reading skills in their native language, and found a strong positive correlation between these two factors (Zimmermann, 2006).

Research in psychology and neuroscience has also identified differences between musicians and nonmusicians, both in terms of verbal memory and recall (Kilgour et al., 2000; Jakobson, Cuddy, & Kilgour, 2003) and brain structure (Schlaug, Norton, Overy, & Winner, 2005). Chan, Ho, and Cheung (1998) found that adult musicians' verbal memory was improved after music training compared to nonmusicians', although in this study, it was difficult to draw a conclusion since the overall level of education was also higher for the group of musicians. However, similar studies with children have shown benefits for verbal memory in the native language as a result of musical training (Ho, Cheung, & Chan, 2003; Fujioka, Ross, Kakigi, Pantev, & Trainor, 2006) and that 15 months of musical training in early childhood is sufficient to change the structure of children's brain development (Hyde et al., 2009).

Marques et al. (2007) found that musicians are better at perceiving intonational pitch violations in an unfamiliar foreign language, which they hypothesise is due to a more general ability to perform fine pitch discrimination tasks. Milovanov et al. (2008) used ERP with children who had superior foreign language pronunciation or musical skills, and concluded that musical and linguistic skills could be partly based on shared neural mechanisms. Even at the low level of brain stem response, musicians performed better than nonmusicians on a passive pitch detection task using lexical tones in Mandarin, which was an unfamiliar language for all participants, who were not tone language speakers (Wong et al., 2007). A review paper on pitch processing found that musical training can improve pitch perception, which the authors suggest might have a positive influence on foreign language learning as well as verbal learning in the native language (Besson et al., 2007). However, individuals without formal musical training also have very sophisticated musical skills that were learned implicitly – in certain domains, at virtually the same level as trained musicians (Bigand & Poulin-Charronnat, 2006).

Taken together, these findings suggest that it is very important to gather information about learners' musical skills and training when investigating language learning abilities, especially for an L2 learning method that includes songs and singing. There might also be an interaction between music and language abilities; for example, it is possible that musical activities could most help individuals who have lower language learning abilities.

2.4.3 Mood

Strong emotions (both positive and negative) can increase the likelihood that input from the environment will be processed and converted into long-term memory, due to increased cooperation between the amygdala and hippocampus (J. R. Anderson, 1999; McGaugh, 2004; Richter-Levin & Akirav, 2000). By contrast, high stress levels and negative emotions can impede working memory and long-term memory formation (Roozendaal, 2008; Hulse, Allan, Memon, & Read, 2007).

Some researchers have argued that since positive mood and emotional arousal has benefits for learning and memory (Hulse et al., 2007), music's facilitation for verbal memory (Thaut et al., 2005; Rainey & Larsen, 2002) or temporary improvements in spatio-temporal IQ (Rauscher, Shaw, & Ky, 1993) may arise simply due to an increase in positive mood or emotional arousal, rather than a cognitive benefit (Schellenberg, 2006; Schellenberg & Hallam, 2005). Therefore, whenever possible it is important to measure learners' emotional states at the beginning and at the end of the experiment sessions to determine whether there is a relationship between mood and L2 test performance, particularly in the learning conditions that use musical presentations of the L2 material.

2.4.4 Handedness

The role of handedness is also an important variable to include, since right-handed individuals tend to slightly outperform left-handers on both L1 and L2 tasks (Bishop, 2001; Andreou, Vlachos, & Andreou, 2005). One large randomised study found that right-handed children aged between 7 and 12 years old had advantages on native language tests compared to left-handers, especially for a sentence completion test (Natsopoulos, Kiosseoglou, Xeromeritou, & Alevriadou, 1998). When possible, handedness is another factor considered in this research project.

2.4.5 Gender

Gender is another important factor to consider because research has shown a slight female advantage for a variety of language tasks using both behavioural and neural measures (Burman, Bitan, & Booth, 2008; Deary, Strand, Smith, & Fernandes, 2007), although some have argued that this has arisen due a publication bias (Sommera, Alemanc, Somersa, Boksa, & Kahna, 2008). By contrast, there is some self-report evidence that men practise musical instruments more than women (Zetterberg, Backlund, Karlsson, Werner, & Olsson, 1998). Another important consideration is that L2 learning anxiety shows a gender difference in adolescents, with girls reporting higher average anxiety than boys (Abu-Rabia, 2004). Since anxiety can interfere with learning, it is important to consider the influence of gender on L2 test performance. To do so, it is important to have approximately the same number of male and female participants in each group of language learners.

2.4.6 Age

Age is another important consideration. Research has shown that that inadequate exposure to the native language from an early age can produce severe and long-lasting consequences for first language skills. This ‘critical period hypothesis’ has also been extended to the field of second language acquisition and claims that some aspects of a new language system are substantially more difficult – or even impossible – to learn to the same degree as the same language skill could have been mastered if they had been learned from an earlier age (Johnson & Newport, 1989). Pronunciation and syntax in a new language are the L2 skills often cited as the most difficult for older children and adults to learn. Evidence suggests that certain aspects of language learning become progressively more difficult as age increases (Hernandez & Li, 2007; Hannon & Trainor, 2007; Skehan, 1989; Diller, 1981). In addition, Scherag, Demuth, Rösler, Neville, and Röder (2004) showed that late L2 learning (English native speakers who had learned L2 German to a high standard) has a greater effect on reaction times (but not overall accuracy) for a lexical decision task of German adjective-noun (vs. pseudo- noun) morphosyntactic agreement, while learning the semantic aspects of the new language appeared to be more robust for the late L2 learners.

But counter-examples can be found for many of the L2 language skills that are considered to be greatly affected by the strict ‘critical period’ view (Abello-Contesse, Chacón-Beltrán, López-Jiménez, & Torreblanca- López, 2006; Abu-Rabia & Kehat, 2004; Bongaerts, Summeren, Planken, & Schils, 1997). There is considerable debate about the extent to which the inability to learn particular skills in a new language to a native-like standard of accuracy is mediated by the learner’s age (which may result in decreases in auditory discrimination abilities or neural plasticity for comprehending new linguistic structures) as opposed to other factors, such as the total amount of time spent learning and practising a new language. A study comparing students’ Russian ACTFL proficiency after participating in immersion and non-immersion L2 learning programmes showed almost a straight line of improvement with hours spent practising the language (Rifkin, 2005), irrespective of the student’s age at the start of learning – some of whom started instruction at age 18.

Trainor (2005) has hypothesised that there might be a musical critical period for certain aspects of music learning, similar to the critical period which is debated in linguistics. To date, there has been much less discussion of the possibility of a critical period for certain aspects of music perception than has been discussed in language learning research. In fact, brain structure changes as a result of musical training is often cited as evidence of neural plasticity (Hyde et al., 2009; Pantev et al., 2003; Rauschecker, 1999).

Since age may be an important factor in L2 learning performance, learners' ages will be kept within a small range and any effects of age on L2 test scores will be examined in the research studies.

2.4.7 Motivation and attitudes in L2 learning

Motivation to learn a new language and attitudes towards speakers of that language have long been identified as important factors in L2 learning. One correlation study investigated the influence of many individual differences factors, including 'attitudes, motivation, self-confidence, anxiety, aptitude, and learning strategies,' on L2 French learning with adult students (R. C. Gardner, Tremblay, & Masgoret, 1997). Five factors were extracted from the IDs data: self-confidence, language learning strategies, motivation to learn French, language aptitude, and attitudes toward learning French and French speakers. Results showed that affective factors, including motivation, attitudes, and language aptitude, were the best predictors of French performance, although one weakness of this study is that data were collected for many more women than men (82 female vs. 20 male participants).

Although both extrinsic factors (instrumental reasons, such as to improve career prospects) and intrinsic factors (when an individual finds the activity of learning to be enjoyable for its own sake) contribute to an individual's motivation to learn a new language, researchers have argued that intrinsic factors are more important for sustaining motivation for foreign language learning long-term (Oxford, 1999). Failoni (1993) observes that adding 'music to the foreign language classroom as a teaching method may be a way to focus students' attention, and produce a more committed learner.'

2.4.8 Intelligence and IQ

Researchers have found that standardised intelligence (IQ) tests provide an extremely strong predictor of academic achievement in a variety of subjects (Mayes, Calhoun, Bixler, & Zimmerman, 2009), including for modern foreign language learning (Deary et al., 2007). In this research project, it was not possible to include any IQ tests because of cost and time constraints. Many IQ tests, such as the Stanford-Binet Intelligence Scale (Thorndike, Hagen, & Jerome, 1986), are expensive to use and these tests often take an hour or longer to administer. For more information about the interplay between IQ, second language learning skills, and musical abilities, readers can consult Gilleece (2006). Gilleece's doctoral thesis contains a study with adolescent pupils in Ireland, with results showing that signifiant positive correlations exist between musical skills and language abilities, independent of general fluid intelligence, *g*F (which can be measured separately from linguistic skills, as opposed to crystallised intelligence, which depends on verbal skills).

Conclusion

In summary, this chapter has discussed evidence from neuroscience showing that there are strong connections between music and language perception, processing, and production. A review of experimental psychology studies has shown that under certain conditions, learning music and language together in a song can support long-term verbal memory in the native language. Educational approaches and studies exploring the use of songs to enhance modern foreign language learning and teaching were also presented. In addition, a discussion of individual differences factors which might influence L2 learning through presentation methods that include songs and singing were outlined.

Since very little controlled research has been conducted to date in this area (Sposet, 2008), many open questions remain regarding whether, to what extent, and how songs and singing might support L2 learning. By contrast, there is more empirical support for the benefits of using music to learn verbal material in participants' native language. While there is anecdotal evidence from educators about the benefits of using music and songs to support L2 learning, there is a gap in the literature when it comes to experimental and classroom-based work. Thus, this thesis explores whether modern foreign language learning can be supported by the inclusion of songs and singing in the learning process. It examines this question under controlled conditions, investigating whether a sung presentation might provide L2 learning or memory benefits compared to more traditional, spoken presentation methods. This thesis also considers the use of songs in a real-world educational setting, exploring the impact of including French songs in the modern language classroom. Since a range of individual differences can influence learning outcomes, the impact of different learner characteristics will be considered in this thesis as we attempt to determine whether songs and singing can support L2 learning.

2.5 Structure of thesis

This section provides a brief preview of each chapter's content, describing the three research studies that were conducted to explore the effects of singing and using songs to teach material in a new language. After these chapters, the outcomes of a workshop developed to facilitate the exchange of ideas between modern foreign language teachers and researchers are presented. Finally, the research findings are summarised and a proposed framework that can help direct future research into whether songs and singing can support L2 learning is presented.

Chapter 3 describes an experimental research study that investigated whether different 'listen-and-repeat' presentation methods (including spoken, rhythmically spoken, and sung phrases) might have different effects for adults learning the meaning of paired-associate phrases in a new language. It also explores the influence of individual differences on verbal memory in the different learning conditions.

Chapter 4 describes a second experiment which investigated whether different 'listen-and-learn' vs. 'listen-and-repeat' presentation methods (including speaking and singing) might have different effects for adults learning the meaning of paired-associate phrases in a new language. This study attempts to directly compare active learning ('listen-and-repeat') to passive learning ('listen-and-learn') procedures using sung and spoken phrases. The experiment further explores whether IDs between learners had differential effects on verbal learning and memory in these learning conditions.

Chapter 5 describes a four-week quasi-experimental, classroom-based research study which was conducted to explore the effects of incorporating L2 songs into the modern foreign language classroom over a longer time period. This arts intervention study was designed to evaluate whether there would be benefits of learning songs or dramatic dialogues in an educational setting. Employing a mixed research design which incorporated both questionnaires and language tests, this study provides information about the potential for songs to support adolescents' L2 learning.

Chapter 6 describes a knowledge exchange project designed to facilitate communication between researchers and modern foreign language teachers. A workbook and afternoon workshop were created to help teachers without musical training incorporate more musical activities into the L2 curriculum and to encourage the discussion and exchange of ideas about the benefits of songs and singing in learning a new language. Post-workshop questionnaires were completed by the attendees to determine whether the workshop format was useful for the exchange of ideas and whether the experience had had an impact on their L2 teaching practice.

Chapter 7 outlines the results of stepwise regression analyses that were conducted on the data from the three research studies. The regression results help provide the basis for a proposed framework for research that can help guide future investigations into whether singing can support modern foreign language learning.

Finally, Chapter 8 offers a discussion and summary of the outcomes of this research project. It also discusses some important questions arising from the findings in this thesis that are worthy of further exploration and presents a framework that can help direct future educational research in this interdisciplinary area.

Chapter 3

Effects of Singing on Paired-Associate Foreign Language Learning

3.1 Introduction

Some experimental work has shown that music can facilitate verbal learning and memory, although most of this work has been conducted using stimuli in participants' native language. The studies that have found benefits for musical presentations of verbal material suggest that the benefit will be greatest for verbatim memory tasks (Thaut et al., 2008; Tillmann & Dowling, 2007; Dowling et al., 2002; Wallace, 1994; Yalch, 1991). Research has also shown that the benefits of a sung presentation may disappear when the rate of stimulus presentation is carefully controlled (Kilgour et al., 2000), especially if the memory test is not challenging (Yalch, 1991; Kilgour et al., 2000). A sung presentation may even be detrimental for verbal learning and memory when the song's melodic and rhythmic structures are difficult to learn (Racette & Peretz, 2007; Silverman, 2007; Wallace, 1994), so the rate of stimulus presentation, the overall duration of stimuli, and the complexity of the music are important considerations which can influence verbal learning and memory with sung stimuli. To date, there is no consensus regarding whether learning verbal material with a melody can provide memory benefits, whether in the native language or a foreign language.

This chapter describes an experimental study designed to investigate whether singing can support foreign language learning under controlled conditions, which has not previously been attempted; the end goal is for this line of research to inform educational practice.

Research questions

1. Is there a benefit for adults' aural/oral (listen-and-repeat) learning of paired-associate phrases in an unfamiliar language when phrases are paired with melodies and/or rhythms during the learning process, compared to a spoken presentation?
2. Do individual differences between learners have an important influence on participants' learning, based on their L2 test performance?

3.2 Method

This section describes the design of the present study and how the original design was modified based on the results of two pilot studies conducted prior to the main study. The section also provides details about the participants involved in this study, development of the stimuli and the measures used, the experimental procedure, and calculation of scores for each measure.

3.2.1 Design

In this study, participants heard 20 paired-associate phrases in English and an unfamiliar language during a 15-minute 'listen-and-repeat' learning period. Three different methods of presentation were developed to compare the relative effects of presenting the material in more musical ways: Speaking, Rhythmic, and Singing.

During the 15-minute learning sessions, participants practiced 20 English-Hungarian paired-associate phrases, listening and repeating the phrases in the new language aloud as best they could. The three learning sessions were followed by a series of production, recall, recognition, and vocabulary tests for the English-Hungarian pairs. Presenting the phrases followed by immediate tests on each item would provide a robust measure of short-term memory. Since the effects of singing on longer-term verbal memory is of interest, this design gave participants time to learn and practise the complete list of phrases three times before performance on the foreign language material was evaluated.

Based on previous research findings for verbal material in the native language, it was predicted that learning paired-associate foreign language phrases with a melody would provide a significant learning advantage for the verbatim spoken language tests (Hungarian Production Test and Delayed Recall Hungarian Conversation) compared to the Speaking condition, but that the benefit for the Singing condition would decrease for the language tests which did not require speaking in the new language (English Recall, Hungarian Recognition, and Multiple-Choice Hungarian Vocabulary Test). Further, it was predicted that performance in the Rhythmic condition would fall between the other two conditions because an emphasis on the rhythm of the spoken stimuli was the only musical element, compared to having both rhythm and melody to support learning in the Singing condition.

The stimuli developed for this experiment were short phrases from a real foreign language. Hungarian was chosen because it was likely to be an unfamiliar language for native English-speaking participants. In addition, Hungarian has different syntactic structures, no lexical cognates, and differences in the sound system compared to the more frequently studied Germanic or Romance languages and, more importantly, compared to English. Using basic phrases in a foreign language, rather than using English-sounding nonsense words, would provide a strong test for whether singing can support foreign language learning. In addition, the stimuli in the three learning conditions were controlled for overall duration, reproducing an important feature of the studies conducted by Kilgour et al. (2000) in participants' native language.

Two major differences between the Racette and Peretz (2007) experiment (which used songs in the native language) and the current study was the use of a longer learning period (15 minutes) and the use of a variety of performance measures to assess learning of the foreign language phrases in this study. If the advantages shown for music in native language studies were not observed in this study, when the foreign language phrases have been carefully controlled for duration and rate of presentation, this would support the claim of Kilgour et al. (2000) that previous experiments showing a benefit for music may have been flawed due to the use of learning materials which were not adequately controlled for these factors. If a benefit for the Singing condition were observed in this study for the verbatim Hungarian tests but not for the language tests which did not require speaking in Hungarian, this would lend support to the idea that singing can support verbal learning, but that a significant advantage for songs may only be observed when challenging, verbatim performance measures are used (Yalch, 1991).

As discussed in Chapter 2, individual differences (IDs) between participants can influence learning outcomes. In this study, it was important to establish whether any significant differences found between the groups were due to the learning condition to which participants had been randomly assigned rather than due to potentially confounding IDs between learners or between groups. Thus, brief assessments of seven IDs which have been shown to either influence learning or to correlate with language skills were collected:

1. Language learning experience and ability

2. Musical experience and ability

3. Mood (at the beginning and at the end of the session)
4. Phonological working memory
5. Handedness
6. Gender
7. Age

The primary purpose of including these IDs measures was to confirm that any difference in Hungarian language performance between the three groups was not simply due to individual differences between participants in the learning conditions. It was predicted that at an individual level, previous language learning experience and language learning ability would strongly predict Hungarian language learning, but the learning condition would also have a significant effect on performance for the verbatim spoken Hungarian language tests. A second prediction was that the measures of musical ability and experience would correlate with performance on the Hungarian language tasks, but to a lesser extent than language learning experience. It was also expected that phonological working memory would correlate with performance on the Hungarian language tests and, to a lesser extent, with the measures of language learning experience and ability. A final prediction was that mood, age, gender, and handedness might influence participants' Hungarian language learning, but that these factors alone would not explain any observed difference in performance between the three groups.

3.2.2 Pilot studies

The stimuli and experiment design was piloted twice prior to beginning the main study, with 9 and 12 individuals, respectively. The results showed very large variation in participants' performance, which provided justification for including several measures of individual differences to explore the IDs' effects on learners' performance. The measures of individual differences are described in more detail later in this chapter, in section 3.4.

Based on feedback from the first pilot study, the instructions for particular tasks were clarified and the decision was made to display the Hungarian phrases on-screen during the first of three 5-minute learning sessions. It was important to show the written Hungarian words once, since participants in the first pilot study found the written Multiple-Choice Hungarian Vocabulary Post-test difficult to complete without having the opportunity to observe how the sound-grapheme correspondences of the Hungarian language differed from those of English (and from other languages they had learned in the past).

Results from the second pilot study (N = 12) showed comparable scores for average overall performance on the Hungarian tests (M = 8.95) compared to participants who completed the first pilot study (N = 9; M = 8.18). No significant differences were observed between participants in the two pilot studies for the measures of individual differences, including age, language learning ability (as measured by Modern Language Aptitude Test-style language structure and memory sub-tests), or musical abilities (for receptive skills as measured by musical ability test scores for rhythm, melody, and pitch discrimination and for productive musical skills as measured by singing and tapping to the syllables of ‘Happy Birthday’), with all p-values > .05.

3.2.3 Participants

Sixty self-selecting adult students (30 male and 30 female) participated in this study. Participants were recruited through a university website advertising an auditory memory study to learn foreign language phrases. The mean age of participants was 21.7 years and was similar across the three conditions. The Speaking condition had the smallest age range and the mean age in the Speaking condition was approximately

1 year lower than the mean for the Rhythmic condition. However, ANOVA showed no difference between the three groups in terms of age ($F(2,57) = 1.247$, $p = .30$). For more details, please see Table 3.1.

Table 3.1: Age of participants – Study 1

Learning Condition	*N*	*M*	*SD*	Range
Speaking	20	21.2 years	1.74	19 - 25
Rhythmic	20	22.4 years	2.66	18 - 28
Singing	20	21.7 years	2.71	18 - 29
Overall (3 groups)	60	21.8 years	2.42	18 - 29

An additional 12 students completed the experiment but were excluded from analysis due to technical problems with the audio recording equipment (four participants), a score higher than 50% on the Multiple-Choice Hungarian Vocabulary Pre-test (four participants), or age (four mature students aged 30 years or older, because a smaller overall age range would likely provide more generalisable results, since learning and memory are influenced by age). This produced a total of 60 participants who were randomly assigned to one of three learning conditions, matched for gender (10 M/10 F in each group).

All participants completed an informed consent form (available in Appendix A) and were treated according to the ethical research standards published by the American Psychological Association (2002). Sessions were completed on an individual basis, with each participant taking about 60 minutes to complete the session. Participants were compensated £6 for their time.

In general, scores on the measures of individual differences confirmed that participants in the three groups were well matched in terms of these potentially confounding factors.[1] Table 3.2 shows the ANOVA calculations for the measures of individual differences. No differences were observed between participants in the three learning conditions for age, mood, phonological working memory, language learning

[1]Details about the measures of individual differences are presented in section 3.4 and the IDs' effects on Hungarian test performance are discussed in section 3.5.

experience, musical experience, or musical ability test scores. However, a trend toward difference between the scores in the three groups on the language memory test (which was modelled on the Modern Language Aptitude Test) was observed. The Singing condition scored highest on this measure, with the Speaking condition performing at the lowest level. A difference between the two groups was found when Student's t-test was conducted ($p < .05$), suggesting that there might have been a pre-existing difference between the Singing and Speaking groups on the language memory measure. Because the results for the Rhythmic condition fell squarely between scores for the Singing and the Speaking conditions, another potential explanation was that in the Singing condition, participants' performance on the test was positively influenced by the musical learning condition to which they had been assigned; these two possibilities are explored in section 3.4.

Table 3.2: ANOVA for ID measures in the Speaking, Rhythmic, and Singing conditions – Study 1

ID Measure	*N*	df	Sum Sq.	Mean Sq.	*F*-statistic	*p*-value
Age	60	2, 57	14.93	7.470	1.287	.28
Working Memory	60	2, 57	2.533	1.267	0.453	.64
Language Experience	60	2, 57	10.40	5.200	0.084	.92
MLAT-style Total	60	2, 57	60.63	35.32	2.003	.14
MLAT-style Structure	60	2, 57	0.233	0.117	0.040	.96
MLAT-style Memory	60	2, 57	63.03	31.52	2.622	**.09**
Musical Experience	60	2, 57	58.60	29.30	0.486	.62
MAT Total	60	2, 57	34.53	17.27	2.092	.13
MAT Receptive	60	2, 57	17.10	8.550	2.198	.12
MAT Productive	60	2, 57	6.633	3.317	0.926	.40
MAT Rhythm	60	2, 57	7.600	3.800	1.836	.17
MAT Pitch	60	2, 57	4.133	2.067	1.476	.24
MAT Melody	60	2, 57	0.433	0.217	0.295	.75

3.2.4 Stimuli

The 20 phrases developed for use in this study were chosen based on both frequency of use and utility in conversation. Useful phrases such as 'Yes, thank you,' 'I don't know,' and 'Where can I buy tickets?' were included as a way to help motivate participants to learn the phrases. In addition, there was some repetition of individual words within the phrases to permit participants to work out which English words corresponded to particular Hungarian words during the learning process, potentially improving performance. The rhythms and melodies created for the stimuli were similar to those found in Hungarian folk songs and exaggerated the natural rhythm and intonation patterns of the language. Three musically trained reviewers, one of whom was also a native Hungarian speaker, checked the phrases' rhythms and melodies before the recording was created.

All stimuli were audio recorded in a sound-proofed room by adult female native speakers of English and Hungarian in a soundproofed recording studio. The recordings were made by an experienced sound engineer using an omni-directional microphone. Digital audio files were recorded onto a Windows computer using the SONAR 4 Studio Edition software.

The Hungarian stimuli were recorded by a native speaker who did not have extensive training in music or singing, but who felt comfortable singing for the recording sessions. First she recorded the spoken stimuli, next the rhythmically spoken stimuli, and finally the sung stimuli. For the spoken stimuli, the Hungarian speaker was asked to speak slowly and clearly, as if she were talking to a non-native speaker of her language. The rhythms created for use in the Rhythmic and Singing conditions were modeled on the natural rhythms of the Hungarian language. The rhythm of the phrases was emphasised in the Rhythmic condition compared to the Speaking condition. The Singing condition added a short melody to the rhythms used for the Rhythmic condition using a small pitch range with Hungarian folk song intervals and melodic lines. A list of the 20 Hungarian phrases with their English meanings and musical notations for the rhythmic and sung stimuli are available in Appendix A.

An important consideration for this study was to control for the duration and rate of presentation of the foreign language phrases in the three learning conditions. Since songs generally present words at approximately half the pace of speech (Murphey, 1990), a slower rate of presentation for the rhythmic and sung stimuli might improve participants' learning, rather than any effects being due to the musical elements found in those phrases. Therefore in this study, the duration of phrases in the new language was carefully controlled. The mean duration for the Hungarian stimuli was approximately equal across the three learning conditions (M = 2.2 seconds), with the shortest, 2-syllable phrases lasting one second and the longest, 8-syllable phrases lasting four seconds. ANOVA comparing the stimuli durations (in milliseconds) across the three learning conditions showed a very close relationship between the phrase durations (p = .97). To control for rate of presentation, the auditory stimuli were presented at the same rate in all three learning conditions: English phrase, 1s pause, Hungarian phrase, 1s pause, Hungarian phrase, 8s pause for participant to repeat the phrase, followed by the next English phrase, and so on.

The order of stimuli presentation in both the learning and testing phases was generated using a pseudo-random number generator based on the Mersenne Twister algorithm (Matsumoto & Nishimura, 1998). The order of presentation was then checked by hand to ensure that a phrase with a repeated word was not placed directly before or after another phrase with the same word. The Hungarian tests were also checked and changed to ensure that the phrases did not follow the same order of presentation as the phrases had appeared during the learning sessions.

3.2.5 Measures

Hungarian language tests

- *Multiple-choice Hungarian Vocabulary Test.* Twenty forced-choice multiple-choice questions were developed as a vocabulary pre-test to assess whether participants had prior knowledge of basic words in Hungarian. The same multiple-choice questions were used as a Hungarian vocabulary post-test after participants had finished all three learning sessions and the first three Hungarian tests. The post-test vocabulary scores provided a measure of whether participants had analysed the foreign language material to learn some of the word meanings which were embedded in the phrases. Generalisations about the meaning of individual words in the sequences of sounds could have been made during the three learning sessions because five of the words were presented as part of two phrases, rather than only appearing in one context. It was predicted that post-test vocabulary scores for the items which only appeared in one phrase would be lower than scores for items which had been presented in two phrases.

- *Hungarian Production Test.* Participants heard the 20 English prompts from the learning sessions – presented in a different, randomised order – and attempted to remember and say the corresponding Hungarian phrase aloud, as best they could. Participants were encouraged to guess if they were not certain of the correct response.

- *English Recall Test.* Participants heard the 20 Hungarian phrases as prompts – presented in a different order – and attempted to remember and say the English translation of the Hungarian phrase. Again, participants were encouraged to guess if they were not certain of the English meaning.

- *Hungarian Recognition Test.* Participants were asked to make same/different judgments for spoken versions of the 20 Hungarian phrases they had heard. Ten phrases were presented with all syllables in the original order. In the remaining 10 items, two syllables within the phrase were swapped, resulting in phrases containing exactly the same syllables, but in a different order from the originals (e.g., 'Megismételné, kérem' was changed to 'Megisteméiné, kérem'). The syllable order of the 10 'different' phrases were modified by hand by a native

English speaker and spoken by the same Hungarian speaker who recorded the other stimuli. Because the 10 'different' phrases still had all of the same syllables, the phrases sounded very similar (but not identical) to the phrases heard during the learning sessions. A list of the recognition test items is available in Appendix A.

- *Delayed-Recall Hungarian Conversation Test.* After a 20-minute delay, when all other tests had been completed (including the seven measures of individual differences), participants attempted to have a conversation in Hungarian. Participants heard an audio recording of 5 simple Hungarian phrases, separated by 8s pauses, which functioned as one side of a brief conversation. Participants were instructed to respond to each statement using a Hungarian phrase that would make sense in the context, and they were encouraged to guess or to say 'I don't know' or 'I don't understand' in Hungarian if they were unsure how to respond.

Equipment

High-quality audio recordings were made during each experiment session using a MicroTrack digital audio recorder. Participants completed the experiment using Firefox on an Acer (Windows) desktop computer and using high-quality noise-cancelling headphones. Responses to the web-based items were sent to a MySQL database running on an Apache web server on an Apple eMac computer which was connected to the same local network.

3.2.6 Procedure

Experiment sessions were held in a quiet room at a comfortable temperature and with appropriate lighting. Participants first completed a phonological working memory test (CNRep), followed by a brief mood pre-session questionnaire (PANAS) and

Hungarian vocabulary pre-test presented on the Windows desktop computer. Because the Firefox web browser was displayed full-screen without displaying the URL, participants could neither return to a previous screen nor proceed to the next page until all required responses for each web page were completed.

Before beginning the Hungarian learning session, each participant was given spoken and written instructions to listen to the recording and to repeat the phrases they heard in the new language aloud, as best they could, and to try to remember what the English meanings were. They were also told that they could use any strategy they wished to learn the phrases, such as repeating the phrase aloud in English or repeating the phrase more than once in the new language. All of the auditory stimuli and test items were played at a comfortable volume through noise-cancelling headphones. Participants completed a practice session with three Hungarian phrases (which were never used again) with the researcher present to answer any questions or to remind participants of the instructions. After establishing that the participant understood the instructions, the researcher went to a nearby room while the participant worked through the remainder of the experiment session by following written, on-screen instructions.

The 15-minute learning period consisted of three 5-minute aural/oral 'listen-and-repeat' learning sessions. During the first learning session, the Hungarian phrases were displayed on-screen as the 20 paired-associate phrases were presented. The written Hungarian phrases were not displayed for the second and third 5-minute learning sessions.

The Hungarian tests became progressively easier as participants proceeded through the experiment session. Participants first completed the Hungarian Production Test, followed by the English Recall Test and the Hungarian Recognition Test, and finally the Multiple-Choice Hungarian Vocabulary Post-test.

After those Hungarian tests, participants completed several measures of language learning aptitude and experience, musical ability and experience, and a brief mood post-session questionnaire (PANAS). Finally, participants attempted to have a short conversation on the Delayed-Recall Hungarian Conversation Test.

At the end of the experiment session, participants completed a 4-item debriefing questionnaire. They were also informed that Hungarian was the language they had been learning during the experiment.

3.2.7 Data analysis

Digital audio recordings were made during each experiment session. Listening to the recordings confirmed that all participants followed the instructions during the Hungarian learning sessions. Responses to the oral test items were analysed by transcribing participants' utterances on the audio recordings. The audio recordings were analysed without knowledge of the learning condition to which each participant had been assigned. This raw data was entered into a spreadsheet and scores were calculated based on the transcriptions. Responses to the web-based items were collected separately via a MySQL database.

- *Multiple-choice Hungarian Vocabulary Test.* Participants' responses to the written, web-based test items were coded in the MySQL database, with a correct response receiving one point and an incorrect answer receiving zero points. A pre-test score higher than 50% on the 20 forced-choice vocabulary questions resulted in the participant's data being excluded, due to the possibility that he or she knew some basic Hungarian phrases prior to beginning the study. The post- test Hungarian multiple-choice vocabulary scores were calculated in the same manner as for the pre-test (1 point for a correct answer and 0 points for an incorrect answer).

- *Hungarian Production Test.* All participants' utterances were transcribed from the audio recordings. One point was awarded if the participant produced the verbatim phrase in the new language correctly, with all syllables in the right order. A total score of 20 was possible. Perfect pronunciation was not required, but participants had to produce the entire phrase with all syllables in the correct order to receive one point. Following Wallace (1994), verbatim Hungarian production scores were also calculated for the accuracy of production in terms of the number of correct syllables (out of 87 possible) and the number of correct words (out of 43 possible). One point was awarded for each syllable (or for each word) that the participant produced correctly.

- *Recall Test.* Participants' English phrases spoken in response to the Hungarian prompts were transcribed from the audio recordings. One point was awarded if the participant produced the correct meaning of the phrase in English correctly. A total score of 20 was possible. A separate calculation was made for verbatim number of correct words in the English phrases produced, with a possible maximum score of 67 words.

- *Recognition Test.* These same/different judgments for spoken versions of the original Hungarian phrases were scored as 'correct' (one point) or 'incorrect' (zero points). No response was also scored as zero points.

- *Delayed-Recall Hungarian Conversation Test.* Participants' responses on this test were transcribed from the audio recordings and scored out of a possible 10 points. Two points were awarded if the participant gave an appropriate reply to the previous Hungarian statement. Responses such as 'I don't know' or 'I don't understand' received one point, while English replies, incorrect or uninterpretable Hungarian phrases earned zero points.

3.3 Results

This section reports the descriptive statistics for overall performance on the five Hungarian tests, followed by inferential statistics comparing differences in performance across the three learning conditions. The influence of individual differences on Hungarian performance is presented in section 3.5.

Results showed no ceiling effects for the Hungarian tests in any group. Many individuals received zero points for the two Hungarian tests which required participants to speak in Hungarian (the Hungarian Production Test and the Delayed-Recall Hungarian Conversation). The number of correct syllables and the number of correct words produced for the Hungarian Production Test were also calculated, following Wallace (1994), and a separate score was calculated for the total number of correct English words on the Recall Test. These additional scores showed the same overall pattern of performance as scores on the verbatim Hungarian Production Test and the English Recall Test. For simplicity, these secondary scores are not included in the statistical analyses. Table 3.3 shows the descriptive statistics for performance on the five Hungarian tests in the three learning conditions.

Table 3.3: Raw Hungarian test scores in the Speaking, Rhythmic, and Singing conditions – Study 1

Condition	**Speaking**			**Rhythmic**			**Singing**		
	N = 20			N = 20			N = 20		
Hungarian Test	*M*	*SD*	Range	*M*	*SD*	Range	*M*	*SD*	Range
Production	1.9	2.2	0 - 9	2.7	2.4	0 - 8	4.0	3.3	0 - 11
Conversation	2.3	2.6	0 - 8	1.8	1.9	0 - 6	3.7	2.2	0 - 7
English Recall	7.5	3.7	2 - 17	7.4	3.7	2 - 13	8.9	3.8	3 - 17
Recognition	16.0	1.6	14 - 19	15.0	2.5	10 - 20	16.0	1.9	13 - 20
M-C Vocabulary	11.6	2.0	9 - 16	11.1	3.4	4 - 17	12.4	3.0	6 - 18
Overall Score	39.3	8.7	27 - 62	37.9	10.7	17 - 60	44.9	11.9	28 - 71

There were two high-performing outliers, both female, in the Speaking condition for the Hungarian Production Test (one of two verbatim spoken Hungarian measures), with both falling more than 1.5 standard deviations above the group mean. The two outliers were nevertheless included in the analyses. Levene's test of homogeneity of variances was then conducted to investigate whether the groups were similar to one another in terms of the dispersion of the Hungarian test scores. None of the results showed a difference for any of the Hungarian tests (for details, see Table 3.4).

Table 3.4: Levene's test results for homogeneity of Hungarian test scores in the Speaking, Rhythmic, and Singing conditions – Study 1

Hungarian Test	df	Levene statistic	*p*-value
Production	2, 57	2.637	.08
Conversation	2, 57	0.506	.61
English Recall	2, 57	0.180	.84
Recognition	2, 57	1.823	.17
M-C Vocabulary	2, 57	1.983	.15

In general, scores on the Hungarian tests did not show a normal distribution, with only the Multiple-Choice Hungarian Vocabulary Post-test passing Shapiro-Wilk's test of normality in all three groups. In the Speaking condition, the Hungarian Production Test and the Delayed-Recall Hungarian Conversation both deviated very far from a normal distribution, both ps $< .01$, while the English Recall Test and the Hungarian Recognition Test did not pass Shapiro-Wilk's test of normality at the $p < .05$ level.[2] In the Rhythmic condition, the Hungarian Production Test and the Delayed-Recall Hungarian Conversation did not pass Shapiro-Wilk's test of normality (W = 0.89, $p < .05$ and W = 0.84, $p < .01$, respectively). In the Singing condition, the distribution of scores for the five Hungarian tests all passed the Shapiro-Wilk's test, although the Hungarian Production Test nearly did not pass ($p = .056$). However, assuming that skewness and kurtosis values less than 2 still fall into the normal range, all of the Hungarian tests were close enough to a normal distribution to permit additional statistical analyses. The histograms and density plots[3] for the results of the five Hungarian tests can be found in Appendix A.

[2]At least in part, this result was due to the two high-performing female outliers in this group.

[3]Density plots were also provided since these charts show a more fine-grained view of the distribu-

Although a normal distribution of scores on the Hungarian tests was not generally found, comparison of the Kruskal-Wallis rank sum test (for use with non-parametric data because it uses the median of two or more samples, rather than using the mean) showed a similar pattern to the ANOVA results. See Table 3.5 for the full ANOVA table and Table 3.6 for a comparison of Kruskal-Wallis and ANOVA results for the five Hungarian tests and overall scores.

Table 3.5: ANOVA for Hungarian tests in the Speaking, Rhythmic, and Singing conditions (signif. code: * 0.05) – Study 1

Hungarian Test	N	df	Sum Sq.	Mean Sq.	F-statistic	p-value
Production	60	2, 57	45.30	22.65	3.177	**.049***
Conversation	60	2, 57	40.43	20.22	4.027	**.023***
English Recall	60	2, 57	27.10	13.55	0.982	.381
Recognition	60	2, 57	14.03	7.02	1.667	.198
M-C Vocabulary	60	2, 57	17.20	8.60	1.044	.359
Overall Score	60	2, 57	557.2	278.6	2.514	.090

Table 3.6: Comparison of Kruskal-Wallis and ANOVA results for Hungarian tests in the Speaking, Rhythmic, and Singing conditions (signif. code: * 0.05) – Study 1

Test	**Kruskal-Wallis**			**ANOVA**		
(Total N = 60)	df	chi-squared	p-value	df	F-statistic	p-value
Production	2	5.666	**.058**	2, 57	3.177	**.049***
Conversation	2	7.669	**.022***	2, 57	4.027	**.023***
English Recall	2	2.395	.30	2, 57	0.982	.381
Recognition	2	2.861	.24	2, 57	1.667	.198
M-C Vocabulary	2	1.668	.43	2, 57	1.044	.359
Overall Score	2	3.611	.16	2, 57	2.514	.090

Since ANOVA/ANCOVA and MANOVA/MANCOVA are robust against violations of normality and heterogeneity of variances when the sample sizes are equal (as was the case in this study), these tests were used to compare participants' Hungarian performance scores in the three groups. More comparisons between the groups were also available for the parametric analysis of variance tests than for their non-parametric equivalents, which was useful for exploring the influence of

tion of scores, not being dependent on the number of bins included, as histograms are.

individual differences on paired-associate phrase learning (section 3.5).

As hypothesised, MANOVA[4] comparing participants' scores on the two verbatim, spoken Hungarian tests (Hungarian Production Test and Delayed-Recall Hungarian Conversation) showed a main effect for learning condition, $F(2,57) = 2.801$, $p < .05$, with the Singing condition showing the highest performance. By contrast, MANOVA comparing participants' scores on the three meaning-based Hungarian tests did not show a difference based on learning condition, $F(2,57) = 0.846$, $p = .54$. One-way ANOVA comparing average performance across all five of the Hungarian tests did show a trend approaching a main effect for condition, $F(2,57) = 2.514$, $p = .090$. With MANOVA conducted using scores on each of the five Hungarian tests, there was also a trend toward a significant group difference, $F(2,57) = 1.796$, $p = .07$.

Post-hoc analysis comparing scores on the five individual Hungarian tests revealed that the Singing condition had significantly better performance on the Hungarian Production Test compared to the Speaking condition, $p < .05$. The Singing condition had significantly better performance on the Delayed-Recall Hungarian Conversation compared to the Rhythmic condition, $p < .05$. ANOVAs revealed that the group differences on the English Recall, Hungarian Recognition, and Multiple-Choice Vocabulary tests were not statistically significant.

These experimental results show a significant effect for learning condition, with the Singing condition showing superior performance on the two verbatim, spoken Hungarian tests compared to the other two learning conditions. The individual differences measures and results are described in more detail in the next section, followed by an investigation of the influence of the ID factors on participants' Hungarian phrase learning in section 3.5.

[4] Wilks' lambda was used for the multivariate F-tests because there were more than two groups in this experiment.

3.4 Measures of individual differences

This part describes the measures of individual differences (IDs) developed and used in this study and the calculation of scores for each measure. Seven ID measures were included in this study to evaluate whether any observed differences between the randomly assigned groups could be explained in terms of IDs that could be relevant when using a musical language-learning method. It was hoped that the measures of IDs would show that the groups were well matched for these potentially confounding factors, including language learning background and experience, musical ability and training, phonological working memory, mood, handedness, dyslexia, age, and gender.

3.4.1 Language learning ability and experience

Participants' language learning ability and experience were assessed using a self-report questionnaire and two brief language learning ability tests modelled on the Modern Language Aptitude Test (Carroll & Sapon, 1959). These measures and the results are described in more detail below.

Language Experience Questionnaire. The 21-item language experience questionnaire (LEQ) was designed specifically for this study and contained questions about participants' prior first and foreign language learning experience. The first 10 items consisted of a Likert-type scale where participants were asked to indicate to what extent each statement was true of their language learning experience. The questions were loosely based on the Strategy Inventory for Language Learning (SILL) developed and validated by Oxford (1996) and a previous research study (Ludke, 2006). The second section consisted of 11 in-depth questions about participants' language learning experience, including the number of languages learned and participants' highest level of competency in a foreign language for speaking, listening, reading, and writing skills. Information was also gathered about the amount of time participants had spent in countries where English was not spoken and how long they spent learning the language before travel.

Participants' responses on the first section of the LEQ ranged from a total of 15 to 48, out of a possible low score of 10 and high score of 50 (overall $M = 31.6$, $SD = 7.8$). There was wide variability in scores between individuals, but ANOVA showed no group differences for the first LEQ sub-section ($p = .92$).

For the second section of the Language Experience Questionnaire, mean scores were similar in the different learning conditions, again with a wide range (from 11 to 35, with 6 to 42 points possible). The Rhythmic condition had the highest mean score ($M = 23.8$) and the Speaking condition had the lowest mean score ($M = 20.8$), but ANOVA showed no significant group differences on the second section of the LEQ ($p = .23$). Participants' scores on the two LEQ sub-sections showed a strong and significant positive correlation (Pearson's $r(58) = 0.72$, $p < .001$).

Total LEQ scores[5] were similar in the different groups. The Singing condition had the highest score overall ($M = 55.2$) and the Speaking condition had the lowest mean LEQ score ($M = 52.2$). No group differences were found for participants' total Language Experience Questionnaire scores ($p = .71$). Table 3.7 shows the overall LEQ scores in the three groups, with separate rows for condition and gender.

Overall, women's total scores on the Language Experience Questionnaire ($M = 56.7$, $SD = 12.1$) were higher and showed slightly less variation than the men's ($M = 51.4$, $SD = 12.4$). An overall trend approaching a difference for gender was observed ($F(1,58) = 3.04$, $p = .086$).[6] A significant gender difference was found in the

[5]The overall score for each participant was calculated by adding the two LEQ sub-section scores together.

[6]For the first section of the LEQ, $F(1,58) = 3.42$, $p = .07$, and for the second section of the LEQ, $F(1,58) = 1.68$, $p = .20$. There were no significant interactions found between gender and condition on the sub-sections or overall on the LEQ.

Table 3.7: Language Experience Questionnaire total scores – Study 1

Learning Condition	*N*	*M*	*SD*	Range
Speaking	20	52.2	9.6	36 - 71
Rhythmic	20	55.0	13.1	33 - 76
Singing	20	55.2	14.5	34 - 85
Male	30	51.4	12.4	33 - 85
Female	30	56.9	12.1	36 - 83
Speaking: Male	10	52.0	7.7	42 - 63
Speaking: Female	10	52.4	11.5	36 - 71
Rhythmic: Male	10	47.7	11.8	33 - 73
Rhythmic: Female	10	62.3	10.4	47 - 76
Singing: Male	10	54.4	16.4	34 - 85
Singing: Female	10	55.9	13.1	37 - 83
Overall (3 groups)	60	54.1	12.4	33 - 85

Rhythmic condition, with women reporting greater language learning experience than men, $p < 0.05$ on the first LEQ sub-section and $p < .01$ on the second sub-section and for the overall LEQ score. By contrast, scores for men and women were very similar on the LEQ in the other two groups (see Table 3.7). The effects of LEQ scores on Hungarian learning are explored in section 3.5.

The language ability test used in this experiment consisted of brief versions of two of the five sub-tests used in the Modern Language Aptitude Test (MLAT) created by Carroll & Sapon (1959). With permission, in this study the items provided by Gilleece (2006) were used as a model. This measure was included to provide a brief measure of participants' language learning ability. The two MLAT-style tests are described below.

Language Structure: Words-in-Sentences. This 12-item grammar test assessed participants' ability to understand the function of words and phrases in English sentences. A high score on this MLAT sub-test correlates with taking an analytical approach to foreign language learning (Skehan, 1989).

Language Memory: Number Pairs. This brief verbal memory test assessed participants' ability to memorise Finnish words paired with numbers. For 60 seconds, participants studied 12 word-number pairs on-screen and could fill in practice items. The web page was timed to automatically proceed to the first of 12 fill-in-the-blank test items when the study time had finished. A high score on this MLAT sub-test correlates highly with taking a memory-based approach to foreign language learning (Skehan, 1989).

Overall scores on the two language learning ability sub-tests were varied. Performance on the Language Structure test showed a ceiling effect, while results of the Language Memory test ranged from zero to a perfect 12 points. The Speaking condition showed the lowest mean scores on both measures, particularly for the Language Memory test, and had the smallest standard deviation of the three groups. Detailed results for the Language Structure and Language Memory sub-tests are shown in Table 3.8. No gender differences were observed on these two language ability tests (ps > .05).

Table 3.8: MLAT-style language learning test scores – Study 1

Learning Condition	**MLAT Sub-Test**	*N*	*M*	*SD*	Range
Speaking	Language Structure	20	9.10	1.37	7 - 11
Rhythmic		20	9.20	2.07	4 - 12
Singing		20	9.20	1.58	6 - 12
Overall (3 groups)		60	9.17	1.67	4 - 12
Speaking	Language Memory	20	3.7	2.89	0 - 11
Rhythmic		20	4.75	3.31	0 - 11
Singing		20	6.15	4.15	0 - 12
Overall (3 groups)		60	4.87	3.57	0 - 12

Analysis of variance (ANOVA) comparing results on the Language Structure test showed no difference between the groups on this measure ($F(2,57) = 0.002$, $p = .98$). By contrast, as previously discussed, the Language Memory test showed a trend toward a main effect for condition, $F(2,57) = 2.487$, $p = .092$, with the Singing group achieving the highest mean scores on this measure. Student's t-test was calculated to compare the difference between the Singing and Speaking groups, which showed the largest difference in scores, and results showed a significant difference, $t(38) = 2.17$, $p = .036$, with the Singing group performing higher than the Speaking group.

A large difference between groups on this measure of verbal memory was problematic because this could indicate that the groups were not well matched for this factor. Despite random group assignment and despite the LEQ scores being very similar in the three learning conditions, the results of the Hungarian tests could be invalidated due to the groups being poorly matched for language learning memory abilities. Another possibility was that participants in the Singing condition were better able to complete the language memory task due to a cognitive benefit of aural/oral singing practice during the 15-minute Hungarian learning session (e.g., the so-called 'Mozart effect' originally observed by Rauscher, Shaw, & Ky, 1993). These possibilities had to be tested to establish whether the groups were not well matched in terms of memory for language learning or whether verbal memory performance was influenced by learning condition.

Therefore, a second version of the Singing condition was conducted with 10 individuals (5 male and 5 female). The experimental procedure was identical to the first Singing condition except that the MLAT-style language learning sub-tests were brought to the start of the experiment, before participants heard any songs in the Hungarian learning sessions. The second Singing group's scores on the Language Memory test did not reach the high level attained by the original Singing group. The second Singing group's results on the Language Memory test ($M = 5.00$, $SD = 3.06$) were similar to scores in the Rhythmic group, $t(28) = 0.200$, $p = .84$, and did not differ from the Speaking condition, $t(28) = 1.14$, $p = .26$. By contrast, performance on the Hungarian tests in the second Singing condition matched extremely well with Hungarian test performance in the original Singing condition for the main study, $t(28) = .003$, $p = 1.00$.

These results suggest that the learning condition may have increased performance on the Language Memory test in the Singing condition, rather than the difference in performance stemming from a pre-existing difference between the groups. This interesting possibility is explored in another experiment, described in Chapter 4. Because the MLAT-style language ability test may be influenced by participants' learning condition instead of providing a stable measure of individual differences between learners, in further statistical analyses participants' scores on this language test (as well as the other measures of individual differences) will be controlled for.

3.4.2 Musical ability and experience

All participants completed a brief musical experience questionnaire regarding their prior training and experiences with music. They also completed a series of brief musical ability tests, divided into sub-tests of receptive and productive musical skills. These measures of individual differences and the results are described below.

Musical Experience Questionnaire. For the first section of the musical experience questionnaire (MEQ), participants were asked to indicate to what extent each of ten statements was true of their musical experience using a Likert-type scale. For this measure, the highest possible score was 50 and the lowest possible score was 10 points. The second section consisted of 13 questions about each participant's instrumental music and singing experience and training, including several fill-in-the-blank items. A total score of between 0 and 52 points was possible.

Scores on the first section of the Musical Experience Questionnaire ranged between 19 and 49 points. The highest mean score was in the Singing condition (M = 38.0) and the lowest score in the Speaking condition (M = 35.8), with the Rhythmic condition falling in between (M = 36.1). ANOVA showed no differences for scores on the first section of the MEQ in the three groups, $F(2,57) = 0.486$, $p = .62$.

For the second section of the Musical Experience Questionnaire, most participants reported that they had previous or current training on at least one musical instrument (90.0%) or voice lessons (25.0%, all of whom reported having musical instrument training as well). The Rhythmic condition had the highest level of musical instrument training and the lowest level of voice training. In addition, 45 participants (75.0%) reported that they had a piano or keyboard in their home, which someone played in 37 participants' homes (61.7%). ANOVA showed no group differences for any of the individual questions, nor for total scores on the second section of the Musical Experience Questionnaire.

Overall scores on the MEQ[7] in the three groups ranged from 25 to 86 points (a total of between 10 and 102 points were possible), with a mean score of 56.8 points. Table 3.9 shows the overall MEQ scores in the three learning conditions, also divided by gender.

Table 3.9: Musical Experience Questionnaire total scores – Study 1

Learning Condition	*N*	*M*	*SD*	*Range*
Speaking	20	54.9	15.9	25 - 80
Rhythmic	20	56.2	12.6	30 - 74
Singing	20	59.3	15.8	25 - 86
Male	30	52.6	16.2	25 - 80
Female	30	61.0	11.8	34 - 86
Speaking: Male	10	53.6	20.0	25 - 80
Speaking: Female	10	56.2	11.4	38 - 74
Rhythmic: Male	10	52.0	13.1	30 - 73
Rhythmic: Female	10	60.3	11.1	34 - 74
Singing: Male	10	52.2	16.6	25 - 73
Singing: Female	10	66.4	11.8	48 - 86
Overall (3 groups)	60	56.8	14.7	25 - 86

In general, women reported higher and less variable scores than men on the Musical Experience Questionnaire, particularly in the Singing condition (see Table 3.9). ANOVA showed a gender difference for the second sub-section of the MEQ and

[7]The overall score for each participant was calculated by adding scores on the two MEQ sub-sections together.

for the overall MEQ score, both significant at the $p < .05$ level.[8] There was no interaction between gender and condition overall on the MEQ (or on the two sub-sections). However, there were within-group gender differences for the second section of the MEQ in the Rhythmic condition, $p < .05$, with women reporting higher scores than men. In addition, there were within-group gender differences in the Singing condition for both the second MEQ section and for the overall MEQ scores, with women's reported musical experience higher than men's, significant at the $p < .05$ level.

Musical Ability Test. Participants also completed a brief musical ability test (MAT). The MAT consisted of two main sections: a receptive music test and a productive music test. The receptive music test was divided into three 8-item sub-tests: Rhythm Discrimination (same/different judgments), Pitch Discrimination (second pitch was higher/lower judgments), and Melody Discrimination (same/different judgments). For the productive music test, participants sang 'Happy Birthday' along with a pre-recorded piano tune while tapping on the desk for each syllable, providing two measures of productive musical abilities (singing and rhythmic tapping in time to the syllables). A total score out of 10 possible points was awarded for each participant's singing (out of 5) and tapping (out of 5) for the productive music test. ANOVA showed no group differences for the MAT sub-tests, all ps > .05. Tables 3.10 and 3.11 show the MAT sub-test scores.[9] Across the three conditions, a significant overall difference for gender on the 'Happy Birthday' tapping (productive rhythm in music) sub-test was found, $t(58) = 2.33$, $p = .02$, but no significant difference was found for learning condition or for the interaction between gender and learning condition.

[8]No gender difference was observed for the first section of the MEQ, $p = .15$.

[9]The influence of musical experience and ability on the Hungarian test results is presented in section 3.5.

Table 3.10: Receptive musical ability test scores in the Speaking, Rhythmic, and Singing conditions – Study 1

Learning Condition	*N*	**MAT sub-test**	*M*	*SD*	*Range*	*Possible*
Speaking	20	Rhythm Disc.	5.7	1.75	2 - 8	8
Rhythmic	20		6.5	1.28	4 - 8	8
Singing	20		6.4	1.23	4 - 8	8
Overall	60		6.2	1.46	2 - 8	8
Speaking	20	Pitch Disc.	4.6	1.19	2 - 7	8
Rhythmic	20		5.1	0.85	4 - 7	8
Singing	20		5.2	1.44	1 - 8	8
Overall	60		5.0	1.19	1 - 8	8
Speaking	20	Melody Disc.	7.2	0.77	5 - 8	8
Rhythmic	20		7.3	0.64	6 - 8	8
Singing	20		7.1	1.12	4 - 8	8
Overall	60		7.2	0.85	4 - 8	8
Speaking	20	Total Receptive	16.7	2.20	13 - 21	24
Rhythmic	20		17.9	1.89	14 - 21	24
Singing	20		17.8	1.85	14 - 20	24
Overall	60		17.5	2.00	13 - 21	24

Table 3.11: Productive musical ability test scores in the Speaking, Rhythmic, and Singing conditions – Study 1

Learning Condition	*N*	**MAT sub-test**	*M*	*SD*	*Range*	*Possible*
Speaking	20	Happy Birthday: Sing	3.3	1.22	0 - 5	5
Rhythmic	20		3.5	1.36	0 - 5	5
Singing	20		3.7	1.31	0 - 5	5
Overall	60		3.5	1.28	0 - 5	5
Speaking	20	Happy Birthday: Tap	3.7	0.81	2 - 5	5
Rhythmic	20		3.6	1.10	1 - 5	5
Singing	20		4.0	0.83	2 - 5	5
Overall	60		3.7	0.92	1 - 5	5
Speaking	20	Total Productive	7.0	1.57	4 - 10	10
Rhythmic	20		7.1	2.28	1 - 10	10
Singing	20		7.6	1.79	4 - 10	10
Overall	60		7.2	1.92	1 - 10	10

3.4.3 PANAS mood questionnaire

Information about each participant's mood was gathered at the beginning and at the end of the experiment session using the 20-item Positive and Negative Affect Scale (PANAS) self-report questionnaire (Watson, Clark, & Tellegen, 1988). Mean scores for the initial and final positive and negative affect scores in the three groups are shown in Table 3.12. ANOVA results showed no group or gender differences on the PANAS mood measure either at the beginning or at the end of the session (all *p*s > 0.05).

At the start of the experiment, the mean positive and negative mood scores in all three conditions were approximately equal. Overall for the three groups, there were significant decreases in both positive and negative affect scores from the pre-session to post-session mood questionnaire (one-tailed paired t-test for positive affect, $t(58) = 2.65$, $p < .01$ and for negative affect, $t(58) = 2.66$, $p < .01$). At the end of the experiment session, the mean positive affect score had decreased significantly in two of the three groups, while decreasing only slightly in the Singing condition where the mean decrease was 0.7 points (n.s.). By contrast, the mean negative affect score in the Speaking condition remained approximately stable at the end of the experiment compared to the beginning, while negative affect scores decreased in both the Singing condition (2 points lower, significant at $p < .01$) and the Rhythmic condition (3 points lower, significant at $p < .01$).

3.4.4 Phonological working memory

All participants completed a brief phonological working memory (WM) test, consisting of the 20 nonwords on the Children's Test of Nonword Repetition (CNRep) with a low-wordlike rating (Archibald & Gathercole, 2006, p. 514). The participant repeated each nonword aloud as best they could after hearing the researcher say it aloud. Responses were audio recorded for later scoring. Participants' scores on the phonological working memory test in the three groups ranged from a low score of 13 up to a perfect score of 20 (overall $M = 17.1$). ANOVA showed no group differences between the learning conditions, $F(2,57) = 0.453$, $p = .64$. Overall group results for the phonological WM test are available in Table 3.13. The chart also

Table 3.12: PANAS mood scores in the Speaking, Rhythmic, and Singing conditions (signif. codes: ** .01, * .05) – Study 1

Learning Condition	**Affect Measure**	*N*	*M*	*SD*	Range	Change
Speaking	Initial Positive Affect	20	19.10	6.44	9 - 30	
Rhythmic		20	19.15	6.60	8 - 33	
Singing		20	19.10	6.25	7 - 33	
Overall (3 groups)		60	19.12	6.32	7 - 33	
Speaking	Final Positive Affect	20	13.75	5.81	0 - 24	*
Rhythmic		20	16.00	8.32	3 - 28	*
Singing		20	18.40	7.79	8 - 35	
Overall (3 groups)		60	16.1	7.51	0 - 35	**
Speaking	Initial Negative Affect	20	3.65	3.51	0 - 11	
Rhythmic		20	6.40	5.58	1 - 21	
Singing		20	4.20	3.61	0 - 11	
Overall (3 groups)		60	4.75	4.43	0 - 21	
Speaking	Final Negative Affect	20	3.75	3.46	0 - 13	
Rhythmic		20	3.35	5.41	0 - 22	**
Singing		20	2.45	2.95	0 - 11	**
Overall (3 groups)		60	3.18	4.05	0 - 22	**

provides details separated by gender due to a significant interaction between condition and gender, with women performing at a higher level than men in the Rhythmic condition on this measure of phonological working memory, $p < .01$.

While there was no difference observed between groups for this measure of phonological working memory, it was still possible that this factor was influencing Hungarian language learning and performance at the individual level. Therefore, ANCOVA was conducted for the five Hungarian tests with participants' phonological working memory score included as a covariate. Results showed that phonological working memory was a better predictor of overall Hungarian performance than learning condition was; however, as reported in section 3.3, the learning condition also continued to show a significant effect for the two spoken Hungarian tests, with less of a difference for the other three Hungarian tests.

Table 3.13: Phonological working memory scores in the Speaking, Rhythmic, and Singing conditions – Study 1

Learning Condition	*N*	*M*	*SD*	*Range*
Speaking	20	17.0	1.5	14 - 20
Rhythmic	20	16.8	1.8	14 - 20
Singing	20	17.3	1.8	13 - 20
Overall (3 groups)	60	17.1	1.7	13 - 20
Male	30	16.8	2.0	13 - 20
Female	30	17.3	1.3	14 - 20
Speaking Male	10	17.7	1.5	16 - 20
Speaking Female	10	16.3	1.3	14 - 18
Rhythmic Male	10	15.8	1.6	14 - 19
Rhythmic Female	10	17.8	1.3	16 - 20
Singing Male	10	16.9	2.3	13 - 20
Singing Female	10	17.6	1.1	16 - 19

3.4.5 Handedness

There were five left-handed individuals in the Speaking condition and six in the Singing condition, with none in the Rhythmic condition. Details about the incidence of left-handed participants in the three groups are shown in Table 3.14.

In general, the differences in performance between left-handed and right-handed individuals were small and in the expected direction, with the right-handed participants outperforming the left-handers. The Hungarian Recognition Test was the only measure on which left-handers performed at a slightly higher level than right-handers in the Speaking group (L: M = 16.40 compared to R: M = 15.87, n.s.). By contrast, the left-handed individuals in the Singing condition performed at a slightly higher level than the right-handed participants in the same learning condition on the Hungarian Production Test (L: M = 4.17 compared to R: M = 3.86, n.s.), the Hungarian Recognition Test (L: M = 16.50 compared to R: M = 15.71, n.s.), and on the phonological working memory measure (L: M = 17.50 compared to R: M = 17.29, n.s.). Handedness did not show any interactions with condition for the Hungarian test results in this study. However, with such a small number of left- handed participants, these results are not conclusive and a separate study would be needed to investigate the effects of handedness on learning L2 material through this musical 'listen-and-repeat' method.

Table 3.14: Left-handed participants in the Speaking, Rhythmic, and Singing conditions – Study 1

Learning Condition	Total *N*	*N* Left-handers	% Left-handers
Speaking	20	5 (3 M / 2 F)	25%
Rhythmic	20	0	0%
Singing	20	6 (3 M / 3 F)	30%
Overall (3 groups)	60	11 (6 M / 5 F)	18.3%

3.4.6 Dyslexia

In the advertisement for this experiment, participants were asked to participate only if they were native English speakers under the age of 30, had normal (or corrected to normal) hearing and vision, and no learning difficulties (such as dyslexia). However, one male participant in the Rhythmic condition appeared to be at risk of dyslexia, with a score of 14 points (out of 20) on the Phonological Working Memory test, 4 points (out of 8) on the Rhythm Discrimination test, and 2 points (out of 5) on the 'Happy Birthday' Tapping sub-test. Difficulties with these skills are common for a sub-group of dyslexic individuals (Overy, Nicolson, Fawcett, & Clarke, 2003; Pickering, 2006). However, apart from the Hungarian Recognition Test,[10] this participant's scores on the Hungarian tests were broadly in line with performance of other participants. Because his Hungarian test scores were similar to those of other participants in the same learning condition, his data were included in the analyses. No participants appeared to be at risk of dyslexia in the Speaking or Singing conditions.

[10]This test consisted of forced-choice questions (same/different) about whether twenty Hungarian phrases were identical to those heard during the learning phase or whether two of the syllables had been swapped; he scored 12 points out of 20 possible on this test.

3.4.7 Debriefing questionnaire

Participants completed an open-ended, 4-item debriefing questionnaire at the end of this experiment before reading a debriefing sheet which informed them that the language they had heard was Hungarian and described the research questions in more detail. A summary of responses to each question are described below.

1. Before beginning this study, did you know the meaning of any of the words you heard in this new language? If so, could you say how many (or which) words you already knew?

Most participants (88.3%) reported that they did not know any of the words in the new language. However, seven participants knew a total of 17 Hungarian words prior to beginning the study. The distribution of individuals who knew a few Hungarian words was fairly well balanced between the groups (one participant in the Speaking condition, three participants in the Rhythmic condition, and three participants in the Singing condition).[11]

Participants in the Singing condition reported knowing a total of nine Hungarian words, compared to two words in the Speaking condition and six words in the Rhythmic condition. The participant (male) in the Speaking condition could say ‘good day’ (*jó napot* – participants in this study heard a polite form, *jó napot kívánok*, ‘I wish you a good day,’ although this more formal phrase was translated into English in this study as ‘good day’). In the Rhythmic condition, one participant (male) knew the word ‘no/not’ (*nem*); one participant (female) knew the word for ‘thank you’ (*köszönöm*); and one participant (male) knew four words: ‘please,’ ‘no/not,’ ‘thank you,’ and ‘goodbye’ (*kérem*, *nem*, *köszönöm*, *viszontlátásra.*). In the Singing condition, two participants (both female) reported knowing eight Hungarian words

[11]In addition, one participant in the Speaking condition reported that he knew which language was being spoken, but found it difficult to ‘place’ any of the words. He also reported that he knew one of the Finnish numbers used in the MLAT-style Language Memory test.

(three words and five words respectively), but unfortunately they did not specify which words they knew; the third participant (male) knew the word meaning 'thank you' (*köszönöm*).

Although a few participants knew between 1-5 Hungarian words before starting the experiment, the number of Hungarian words known was not very large (17 words altogether in the three groups). While it is possible that a participant's knowledge that the language was Hungarian might have had other effects on their performance (for example, if they believed that Hungarian is not an easy language to learn), the impact of this was also likely balanced between the different groups. There were no major differences in Hungarian test performance for those individuals compared to other participants in the same condition.

2. What do you think the experiment was about?

Participants' responses to this open-ended question were varied, but could be sorted into several broad categories. The most common response (36 participants or 60.0%) was that the aim of the experiment was to explore how people learn and memorise words in a new language (or language aptitude) using different language learning styles (e.g., visual vs. auditory vs. musical learners). Another common response was that the purpose was to determine whether language skills were linked to musical skills (10 participants or 16.7%). Several participants also wrote that the aim was to measure and/or improve foreign language teaching methods (15.0% or 9 participants). Nine participants made guesses that were similar to the true research question, most of whom were in the Singing condition (6 individuals, which was 30.0% of that group and 10.0% of the total).[12] Again, no major differences in Hungarian test performance were noted between the participants who guessed the

[12]The total percentage does not add up to 100% because several participants, including three individuals in the Singing condition who guessed the correct answer, had more than one guess about the purpose of the experiment.

purpose of the experiment and those who did not in the same learning condition.

3. Would you suggest any changes or ways of improving the learning experience for people participating in future studies?

Many participants (55.0% or 33 participants) chose not to answer this question. Some helpful responses were the suggestion to provide learners with the written Hungarian words more frequently (six participants), 'more time to think in between speaking parts' (four participants), and clearer instructions especially for the measures of individual differences (three participants).[13] Also, three participants wanted more time to memorise the number words (in Finnish, on the Language Memory test) and one requested more time to feel the pulse for the Rhythm Discrimination sub-test of the MAT, but changing these tests was not possible since they were designed by other researchers.

4. Do you have any other comments?

Most participants (71.7%) did not choose to answer this question, but some offered useful responses. Among these were that the experiment was a bit too long and/or difficult (nine participants), with two individuals suggesting that it might be easier if there were multiple-choice questions for the phrase meanings, especially for more visual learners. Two individuals wrote that the music parts were fun and that they would like to have had more music included, although one participant reported that the 'Happy Birthday' tune was not in her range. In addition, nine participants wrote that they felt the experiment had been good, fun and/or interesting.

[13]For this question, one participant (male, in the Rhythmic condition), wrote: 'Instructions were a bit hard, but because of my reading comprehension.' Although he had a lower than average Phonological Working Memory score (14 points out of 20), his scores on the other measures of individual differences did not suggest that he was at risk of dyslexia (see page 74 for details of how the possibility a participant had dyslexia was determined) and his Hungarian test scores were in line with other participants in the same condition, so his data were included in the analysis.

3.4.8 Age

In this study, an attempt was made to keep the variability in age of participants to a small range by including individuals only between 18 and 29 years of age. Table 3.15 shows the age of participants in the three groups, also separated by gender, none of which showed significant differences. While older participants tended to show higher performance on the Hungarian tests, age was not found to be a significant overall predictor of performance in this study.

Table 3.15: Age of participants in the Speaking, Rhythmic, and Singing conditions – Study 1

Learning Condition	*N*	*M*	*SD*	Range
Speaking	20	21.2 years	1.74	19 - 25
Rhythmic	20	22.4 years	2.66	18 - 28
Singing	20	21.7 years	2.72	18 - 29
Overall (3 groups)	60	21.8 years	2.42	18 - 29
Male	30	21.7 years	2.35	18 - 27
Female	30	21.8 years	2.54	18 - 29
Speaking Male	10	21.5 years	2.27	19 - 25
Speaking Female	10	20.9 years	0.99	19 - 22
Rhythmic Male	10	22.2 years	2.90	18 - 27
Rhythmic Female	10	22.6 years	2.55	19 - 28
Singing Male	10	21.5 years	1.96	19 - 26
Singing Female	10	21.9 years	3.41	18 - 29

3.4.9 Gender

Overall results for this study showed few differences for gender. For the individual differences measures, the only test showing a significant difference between men and women was the 'Happy Birthday' tapping test which was previously mentioned ($p < .05$), with women performing better than men on this measure. Within the Rhythmic group, women also performed better than men on the phonological working memory measure (CNRep), $p < .01$.

For the five Hungarian tests, MANOVA showed an overall trend toward a significant difference for gender (p = .073), with women performing better than men. This trend held at approximately the same level (p = .09) for both spoken, verbatim Hungarian tests (the Hungarian Production Test and the Delayed-Recall Hungarian Conversation) and for the meaning-based Hungarian tests (English Recall, Hungarian Recognition Test, and Multiple-Choice Hungarian Vocabulary Post-test). However, no individual Hungarian test showed a difference for gender and there were no significant interactions between gender and condition.

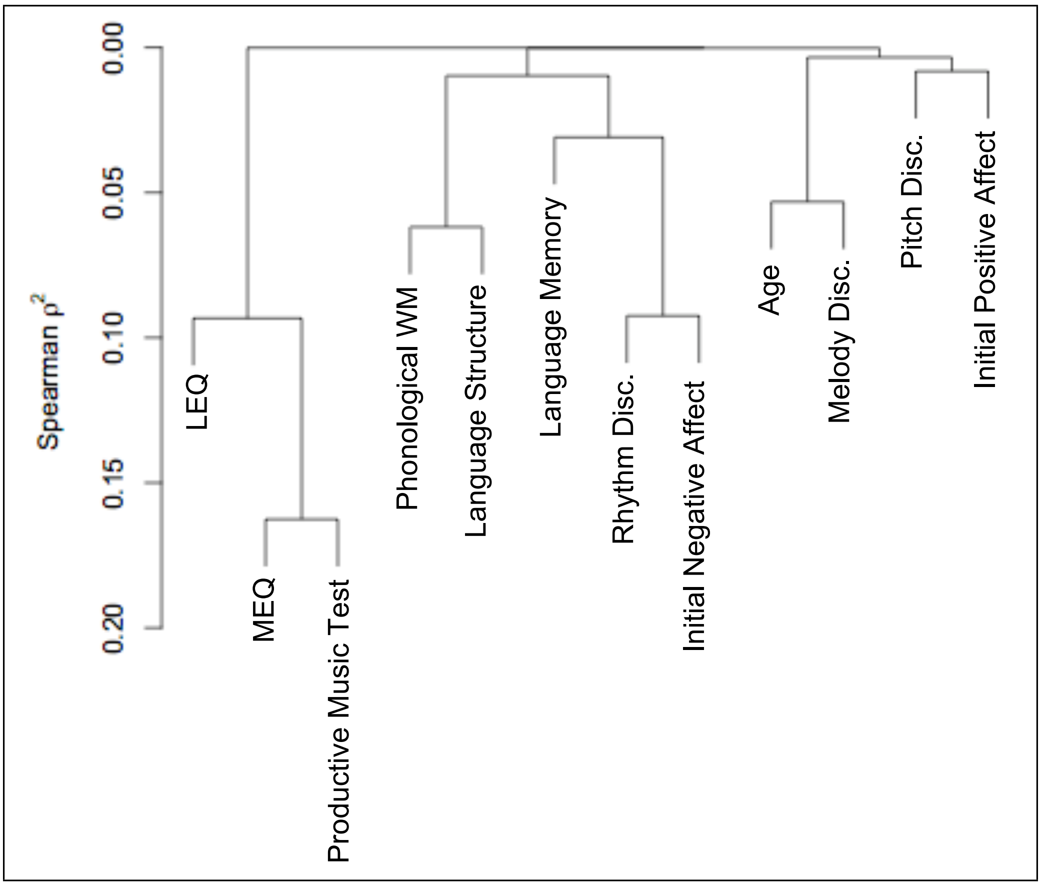
3.5 Influence of individual differences on Hungarian test performance

This section explores the effects of the measures of individual differences described in the previous section on participants' paired-associate Hungarian phrase learning. These analyses have the potential to provide an account of which of the factors measured in this study have the greatest influence on paired-associate foreign language learning, in addition to exploring whether singing can support the learning of verbal material when the influence of IDs are controlled for. This section provides an analysis of the Hungarian test results, using several statistical procedures to take into account the effects of the IDs measured in this study on participants' Hungarian test performance.

3.5.1 MANCOVA controlling for pre-existing ID factors

MANCOVA (multivariate analysis of co-variance) was conducted on the data to investigate the effects of the 12 individual differences measures (including the sub-tests) on the Hungarian test results. Since the 'Happy Birthday' Singing and Tapping test scores correlated highly ($r(58) = 0.48$, $p < .001$) compared to correlations between the other ID measures, the total Productive Music test score (calculated by adding together the two 'Happy Birthday' sub-test scores) was used in the analysis. For the same reason, the total Language Experience Questionnaire and Musical Experience Questionnaire scores were used rather than including the score for each LEQ and MEQ sub-section.[14] None of the remaining ID measures correlated strongly with one another, as shown visually in Figure 3.1. Most of the musical and language ability tests also clustered separately from one another, providing support for the assumption that the sub-tests were measuring distinct underlying abilities or constructs.

Figure 3.1: Correlations between the ID factors – Study 1



MANCOVA results including the twelve IDs as covariates plus the learning

[14]The LEQ sub-section and MEQ sub-section correlations were also significant at the $p < .001$ level; for more details see pages 62 and 67.

condition, gender, and handedness factors revealed that the overall Language Experience Questionnaire (LEQ) score was the best predictor of performance on the five Hungarian tests ($F(5,37) = 5.080$, $p = .001$) and another significant predictor was the Initial Positive Affect score ($F(5,37) = 3.395$, $p = .013$). After controlling for the effects of the ID factors, MANCOVA revealed a significant overall main effect for learning condition on the five Hungarian language tests, $F(10,74) = 2.261$, $p = .023$, partial $\eta^2 = .234$, and a marginal interaction between gender and condition ($F(10,74) = 1.778$, $p = .08$).[15] Table 3.16 shows the full MANCOVA results, including the effect size (partial η^2)[16] and observed power for each factor.[17]

Table 3.16: MANCOVA for all five Hungarian tests with ID factors as covariates (signif. codes: ** 0.01, * 0.05) – Study 1

Effect	**df**	**Wilks'**	***F*-stat.**	***p***	**Partial η^2**	**Power**
Intercept	5, 37	.817	1.653	.17	.183	.510
Lang. Exp. Quest.	5, 37	.593	5.080	**.00****	.407	.969
Music Exp. Quest.	5, 37	.920	0.645	.67	.080	.208
Phon. WM	5, 37	.785	2.027	.10	.215	.610
Rhythm Disc.	5, 37	.799	1.864	.12	.201	.568
Pitch Disc.	5, 37	.821	1.613	.18	.179	.499
Melody Disc.	5, 37	.913	0.705	.62	.087	.226
Productive Music Test	5, 37	.873	1.075	.39	.127	.338
Language Structure	5, 37	.947	0.413	.84	.053	.144
Language Memory	5, 37	.983	0.129	.99	.017	.076
Initial Positive Affect	5, 37	.686	3.395	**.01***	.314	.857
Initial Negative Affect	5, 37	.887	0.942	.47	.113	.297
Age	5, 37	.950	0.393	.85	.050	.139
Hand	5, 37	.933	0.533	.75	.067	.177
Gender	5, 37	.894	0.877	.51	.106	.277
Condition	10, 74	.587	2.261	**.02***	.234	.891
Gender x Condition	10, 74	.650	1.778	.08	.194	.786

Conducting MANCOVA for the two verbatim, spoken Hungarian tests revealed

[15]The marginal interaction was due to women in the Rhythmic condition performing at a lower level than men, whereas women had better Hungarian test scores than men in the Speaking and Singing conditions.

[16]Cohen defines partial η^2 values of .01, .06, and .14 as low, medium, and large effect sizes, respectively (Cohen, 1992).

[17]Type III MANCOVA was used because this statistical procedure controls for the effects of each of the ID measures when analysing the main effect of condition.

a highly significant main effect for learning condition, Wilks' lambda = .707, $F(4,80) = 3.790$, $p = .007$, partial $\eta^2 = .159$, power = .87. The overall LEQ score was again the best predictor of performance (Wilks' lambda = .688, $F(2,40) = 9.073$, $p = .001$, partial $\eta^2 = .312$, power = .97). The LEQ score was followed closely by the Initial Positive Affect score (Wilks' lambda = .688, $F(2,40) = 6.970$, $p = .003$, partial $\eta^2 = .258$, power = .906). The Phonological Working Memory score and the Pitch Discrimination test were also significant, as was the interaction between gender and condition (all at the $p < .05$ level). The full MANCOVA results for the verbatim, spoken Hungarian tests (Hungarian Production Test and Delayed-Recall Hungarian Conversation), are shown in Table 3.17.

Table 3.17: MANCOVA for the spoken, verbatim Hungarian tests with ID factors as covariates (signif. codes: ** 0.01, * 0.05) – Study 1

Effect	**df**	**Wilks'**	***F*-stat.**	***p***	**Partial η^2**	**Power**
Intercept	2, 40	.880	2.733	.08	.120	.510
Lang. Exp. Quest.	2, 40	.688	9.073	**.00****	.312	.965
Music Exp. Quest.	2, 40	.983	0.348	.71	.017	.102
Phon. WM	2, 40	.861	3.228	**.05***	.139	.583
Rhythm Disc.	2, 40	.913	1.896	.16	.087	.371
Pitch Disc.	2, 40	.844	3.692	**.03***	.156	.645
Melody Disc.	2, 40	.978	0.443	.65	.022	.117
Productive Music Test	2, 40	.949	1.075	.35	.051	.225
Language Structure	2, 40	.965	0.717	.49	.035	.163
Language Memory	2, 40	.994	0.130	.88	.006	.069
Initial Positive Affect	2, 40	.742	6.970	**.00****	.258	.906
Initial Negative Affect	2, 40	.913	1.896	.16	.087	.371
Age	2, 40	.968	0.654	.53	.032	.152
Hand	2, 40	.964	0.740	.48	.036	.167
Gender	2, 40	.953	0.989	.38	.047	.210
Condition	4, 80	.707	3.790	**.01****	.159	.874
Gender x Condition	4, 80	.775	2.713	**.04***	.119	.727

By contrast, MANCOVA results for the three meaning-based Hungarian tests (English Recall Test, Hungarian Recognition Test, and Multiple-Choice Hungarian Vocabulary Test) showed that the LEQ score was the only significant predictor of performance (Wilks' lambda = .731, $F(3,39) = 4.794$, $p = .006$, partial $\eta^2 = .269$, power = .871). For the meaning-based Hungarian tests, the interaction between gender and condition was again marginally significant (Wilks' lambda = .734, $F(6,78) = 2.176$, $p = .054$, partial $\eta^2 = .143$, power = .741). Similar to the MANOVA results, for the three meaning-based Hungarian tests MANCOVA revealed no main effect for learning condition (Wilks' lambda = .831, $F(6,78) = 1.264$, $p = .28$, partial $\eta^2 = .089$, power = .468).[18] Table 3.18 shows the detailed MANCOVA results for the three meaning-based Hungarian tests.

Table 3.18: MANCOVA for the meaning-based Hungarian tests with ID factors as covariates (signif. code: ** 0.01) – Study 1

Effect	**df**	**Wilks'**	***F*-stat.**	***p***	**Partial η^2**	**Power**
Intercept	3, 39	.866	2.017	.13	.134	.478
Lang. Exp. Quest.	3, 39	.731	4.794	**.01****	.269	.871
Music Exp. Quest.	3, 39	.951	0.664	.58	.049	.177
Phon. WM	3, 39	.914	1.216	.32	.086	.300
Rhythm Disc.	3, 39	.941	0.812	.50	.059	.209
Pitch Disc.	3, 39	.897	1.489	.23	.103	.362
Melody Disc.	3, 39	.978	0.290	.83	.022	.101
Productive Music Test	3, 39	.879	1.795	.16	.121	.430
Language Structure	3, 39	.982	0.245	.87	.018	.092
Language Memory	3, 39	.986	0.183	.91	.014	.081
Initial Positive Affect	3, 39	.873	1.898	.15	.127	.453
Initial Negative Affect	3, 39	.942	0.800	.50	.058	.206
Age	3, 39	.975	0.337	.80	.025	.110
Hand	3, 39	.959	0.553	.65	.041	.153
Gender	3, 39	.953	0.640	.59	.047	.172
Condition	6, 78	.831	1.264	.28	.089	.468
Gender x Condition	6, 78	.734	2.176	**.05**	.143	.741

Thus, when all of the ID factors were included as covariates using MANCOVA, results showed a pattern similar to the original MANOVA results. A significant effect for learning condition was found for the spoken Hungarian tests and overall for the five Hungarian tests, but not for the meaning-based Hungarian tests. In addition, the MANCOVA analyses have revealed that other factors also influenced the Hungarian test results, especially participants' reported prior language learning experience (as

[18] As would be expected, ANCOVA results for the individual Hungarian tests fell into a similar pattern for the covariate ID measures, with some variation for particular tests.

measured by total LEQ scores, with a large effect size and power above .8. Other individual differences factors that were predictive of Hungarian test performance were Initial Positive Affect score (overall and for the spoken, verbatim Hungarian tests at the $p < .01$ level, also with a large effect size), Phonological Working Memory score and Pitch Discrimination test score (both at the $p < .05$ level and with large effect sizes). Marginal and significant interactions were also found between learning condition and gender on the Hungarian tests (with a medium effect size and power slightly below .8), with women in the Rhythmic condition generally scoring lower than men, while women tended to outperform men in the other two conditions.

3.6 Discussion

The results of this randomised, controlled experimental study showed that participants in the Singing condition outperformed participants in the Speaking and Rhythmic conditions, on four out of the five Hungarian tests. A significant between-groups performance difference was found on the two verbatim spoken Hungarian tests, both immediately after learning and after a 20-minute delay, showing a facilitation for paired-associate foreign language learning through the sung listen-and-repeat learning method. The learning benefit observed for the Singing condition was less evident for three other foreign language tests, which had a greater focus on meaning rather than oral production. These results support findings that suggest that there are memory benefits for learning verbal material through singing in the native language (Thaut et al., 2008; Rainey & Larsen, 2002), and support the hypothesis that the benefits of music for verbal learning are most evident on verbatim memory tasks (Yalch, 1991).

The Hungarian language was completely unfamiliar to almost all participants before beginning the experiment session. Because Hungarian has a different sound system, syntactic structures, and no lexical cognates with more familiar languages or with English, use of this language provided a robust test of the memory benefits of music for foreign language learning. The benefit of learning by singing the phrases cannot be explained by the reasons proposed by Kilgour et al. (2000), as stemming from a difference in the overall duration or rate of presentation of stimuli, because the duration of phrases was the same and all stimuli were presented at an identical rate in the three learning conditions. In addition, no significant differences were found between the three learning conditions in terms of participants previous musical or language learning experience or ability, phonological working memory, age, or mood, so these factors do not explain the learning benefit found for participants in the Singing condition.

In contrast with the findings of Purnell-Webb and Speelman (2008), which indicated that the rhythm of native-language phrases was the most supportive element of a musical presentation method, in the current study participants in the Rhythmic condition did not perform at a similar level to those in the Singing condition. Indeed, participants in the Rhythmic condition often performed at a lower level than participants in the Speaking condition; this was particularly true for the women in the Rhythmic condition. One possible explanation is that the rhythmic stimuli were created by exaggerating the natural rhythms of the 20 short Hungarian phrases, and thus were similar in rhythmic structure to the spoken phrases; however, this does not explain the gender difference. Whether rhythmic speech can support L2 learning is worth further investigation, perhaps with intermediate language learners instead of beginners.

In general, women and men tended to be well matched in this study on the measures of individual differences, apart from the ‘Happy Birthday’ Tapping test (where women outperformed men). For the five Hungarian tests, MANOVA also showed an overall trend toward a significant gender difference (p = .073), with women performing better than men. However, this pattern was not true in all three learning conditions, as evidenced by the marginal and significant interactions found between learning condition and gender on the Hungarian tests using MANCOVA. It is unclear why the women in the Rhythmic condition generally showed lower performance than men, since women often tend to outperform men on language tasks.

Several of the individual differences factors were found to be predictive of Hungarian test performance in this study. Analysing the five Hungarian test scores while controlling for the individual differences factors revealed that, not surprisingly, previous language learning experience (LEQ) had the greatest overall influence on participants' Hungarian test scores. While MANCOVA analyses revealed that the LEQ score was the strongest predictor and positive mood at the start of the experiment was also a significant predictor of Hungarian test performance, both with medium to large effect sizes, these factors were closely followed by the learning condition (p = .02, effect size (partial η^2) = .23, a medium effect). Two additional individual differences factors that predicted performance on the spoken, verbatim Hungarian tests were Phonological Working Memory score and Pitch Discrimination test score, both with medium effect sizes and $p < .05$.

Consistent with previous findings, small positive correlations were found between participants' prior musical experience and language experience (Gilleece, 2006), but musicianship was not a strong predictor of Hungarian test performance in this experiment, nor did differences in musical experience or ability explain the facilitation for foreign language learning in the Singing condition. Participants' moods at the start of the experiment session also had a significant influence on performance for some of the Hungarian measures, as previously shown by Schellenberg (2006). However, this factor alone does not explain the facilitation for language learning in the Singing condition, because the learning condition was also a significant predictor of Hungarian test scores. In addition, the present results suggest that singing the foreign language phrases during the learning process was extremely beneficial for later production of spoken phrases in the new language – even though participants in the Singing condition reported the lowest positive mood at the start of the session.

The finding that the MLAT-style Language Memory test may have been influenced by the learning condition to which participants had been randomly assigned is an interesting possibility; however, this measure was not a significant predictor of performance on the Hungarian tests, suggesting that memorising written words may be a distinct skill from learning through an aural/oral listen-and-repeat procedure. That the musical tests were also influenced by the learning condition did not appear likely, since none of the differences between the other ID measures were statistically significant in the three learning conditions. The possibility that language memory and/or musical abilities might be enhanced by listening to and singing along with foreign language phrases is explored in the experimental study described in the next chapter.

Conclusion

The results of the present study support the claim that music can provide an effective memory aid for verbatim verbal learning tasks, but that the difference between spoken and sung presentation methods will decrease as the memory tests become easier (Yalch, 1991). Since the overall duration and rate of L2 phrase presentation was equal across the three learning conditions, this benefit of learning through singing cannot be explained by a difference in stimuli durations, as proposed by Kilgour et al. (2000). Perhaps the null result found for condition in that study might have been due to the use of an easy, meaning-based language test, since learning benefits for the Singing condition were not found for similar tests in the present study. MANCOVA results showed that previous language learning experience and positive mood at the start of the experiment session were even more significant factors influencing overall Hungarian test performance than the learning condition was, both showing medium to large effect sizes, but the results of this study show that a singing ‘listen-and-repeat’ method for learning paired-associate phrases in a foreign language can provide powerful support for verbal memory, both immediately after learning and after a 20-minute delay, and particularly for verbatim, spoken tests in the new language.

Chapter 4

Effects of Active Learning Compared to Passive Learning of Spoken or Sung Paired-Associate Foreign Language Phrases

4.1 Introduction

The previous chapter provided experimental evidence in support of the claim that singing short phrases in a new language can support foreign language learning over a short time period. Listening and repeating sung foreign language phrases for 15 minutes was shown to facilitate verbal learning and memory, especially for the adult participants' speaking skills in the new language. The current study explores a related question that is important for researchers and for modern language teachers: is singing the mechanism that supports foreign language learning, or will learning benefits appear when listening to sung material in a new language? Classroom-based research using musical mnemonics to support spelling skills (in the native language) suggests that it is important to repeat the material during learning, at least sub-vocally (Gfeller, 1983). Memory research on the importance of retrieval attempts as part of the learning process (Karpicke & Roediger III, 2008; Roediger III & Karpicke, 2006) is also relevant because these findings suggest that passive listening may be less effective for paired-associate learning than active aural-and-oral practice.

An experiment investigated the effects of passive vs. active learning through a learning method that incorporated both singing and signing (using American Sign Language) of 15 words in an unfamiliar foreign language, Japanese (Iwata, 2005). In the active learning condition, participants learned the phrases by singing and signing with the experimenter, while the participants in the passive learning condition watched the experimenter teaching the participant who was in the active learning condition. Results showed a significant advantage for the active learning condition in terms of pronunciation and vocabulary on two post-tests, although scores in both groups improved from the first post-test to the second post-test. Participants in the active learning condition also reported more positive mood than participants in the passive learning condition. In this study, there was no comparison of the singing method to a speaking method for the foreign language phrases, and all participants also observed ASL signs during the learning process. Thus, it remains unclear to what extent the differences in Japanese learning observed between the active and passive learning conditions in Iwata's study might be due to the singing auditory method, because she combined singing with movements (ASL signs) in the learning procedure. A similar pre/post-test study was conducted with young ESL learners (Schunk, 1999), which also showed that vocabulary learning was greatest for instruction through singing and ASL signs, followed by speaking with ASL signs. The singing-without-signs condition also showed significantly higher performance at post-test than the speaking-without-signs condition. To date, the question of whether there are performance differences for active vs. passive paired-associate L2 learning through spoken vs. sung phrases (without including movement) has not been explored under controlled conditions.

The present chapter describes an experimental study that investigates whether there is a difference in verbal memory for participants who passively listen to spoken or sung foreign language phrases during the learning process (listen-and-learn), compared to actively learning phrases under conditions similar to those of the first experiment (listen-and-repeat speaking or singing the Hungarian phrases). It also explores whether there are any differences for type of stimulus (spoken vs. sung foreign language phrases), which might appear for particular foreign language tests but not others. The inclusion of two of the three active learning conditions from the previous study (Speaking and Singing) aims to address some of the outstanding questions from the previous experiment; in particular, whether singing phrases in a foreign language for a short period of time might have a positive influence on participants' scores on a subsequent MLAT-style Language Memory task. In addition to the ID measures used in the previous experiment, the current study examines participants' nonverbal reasoning abilities to discourage participants from guessing that the research question was related to music at the start of the experiment, to ensure that the groups were well matched for nonverbal reasoning, and to help identify participants with undiagnosed dyslexic tendencies. The experiment also explores whether there are any effects on Hungarian test performance of participants' motivation to learn the paired-associate foreign language phrases and/or participants' self-reported success at learning the phrases.

Research questions

1. Is there a difference for an active (listen-and-repeat) versus a passive (listening only) learning method?
2. Is auditory paired-associate foreign language learning facilitated if new material is presented through sung compared to spoken stimuli?
3. Do individual differences between learners have an important influence on participants' Hungarian test performance?
4. To what extent are the Hungarian test results influenced by participants'

motivation and/or self-perceived success at learning the phrases?

4.2 Method

This section describes the design of a second experimental study, including the results of a pilot study conducted before the main study began. Details are also provided about the participants' background, a brief description of the Hungarian tests and the scoring procedure for each measure, and the experimental procedure.

4.2.1 Design

In this study, participants were randomly assigned to a learning condition in which they heard 20 of the paired-associate English-Hungarian phrases used in the previous experimental study (see Chapter 3). Four learning conditions were developed for the present study to compare the relative effects of learning material through an active learning procedure (listen and repeat phrases in the new language) compared to a passive learning procedure (only listen to the phrases) together with the effects of listening to spoken vs. sung stimuli. This design will also enable exploration of any interactions between type of learning condition and type of stimulus for particular Hungarian tests.

During each 5-minute learning session, participants either listened to 20 paired-associate English-Hungarian phrases (passive learning) or listened to and repeated the Hungarian phrases aloud as best they could (active learning). The stimuli were presented as either spoken or sung Hungarian phrases. The learning sessions were followed by a series of production, recall, recognition, and vocabulary tests for the English-Hungarian pairs.

Based on the results of the previous experiment and research findings for verbal material in the native language, it was predicted that learning the foreign language phrases by singing along with a melody would provide a significant learning advantage for the verbatim spoken language tests (Hungarian Production Test and Delayed-Recall Hungarian Conversation) compared to repeating a spoken version of the Hungarian phrases, but that the benefit for the Singing condition would be smaller for the language tests which did not require speaking in the new language (the English Recall and Hungarian Recognition tests). In addition, it was predicted that participants in the two passive learning conditions (Listen to Speech and Listen to Singing) would perform at a lower level than those in the two active learning (listen-and-repeat Speaking and Singing) conditions because actively repeating phrases in a foreign language during the learning process may facilitate learning compared to passively listening to the same phrases, at least when the melodies are combined with ASL signs (Iwata, 2005). Therefore, performance on the Hungarian tests was predicted to follow this pattern in the four groups: Listen to Speech < Listen to Singing < listen-and-repeat Speaking < listen-and-repeat Singing, with statistically significant differences in performance found for the verbatim spoken Hungarian tests.

In addition, it was predicted that individual differences between learners might influence participants' Hungarian test scores, but that IDs alone would not explain the potential benefits for the two active learning conditions or for the two sung stimuli conditions.

4.2.2 Pilot study

A pilot study was conducted with 16 participants who were randomly assigned to one of four learning conditions, for a total of 4 participants in each group: Listen to Speech, Listen to Singing, listen-and-repeat Speaking, and listen-and-repeat Singing. The participants in the two active, listen-and-repeat learning conditions followed the same learning procedure as the previous experiment (listen to the English phrase followed by hearing the Hungarian phrase twice, and then attempt to repeat the Hungarian phrase once aloud), whereas participants in the two passive learning conditions were asked only to listen to the list of phrases (with each Hungarian phrase heard three times during each learning session).

The stimuli, measures, and procedure were similar to the previous experimental study, except that all of the individual differences measures were implemented using a pre/post-test design for this pilot study, apart from the phonological working memory test (CNRep) and the Language and Musical Experience Questionnaires. Pre-tests and post-tests for the ID measures were included to clarify whether the marginal group difference observed in the previous experiment for the MLAT-style Language Memory test was due to a pre-existing difference between the participants who were randomly assigned to the three groups, or instead due to a short-term memory effect of singing the Hungarian phrases for 15 minutes before completing the ID measures in that experiment – possibly similar to the short-term 'Mozart effect' observed for spatio-temporal IQ (Rauscher et al., 1993). The individual differences pre/post-tests in this pilot study permitted exploration of whether scores on any of the other ID measures might be influenced by singing the L2 phrases.

However, it was important to prevent participants from guessing at the start of the experiment that music was an important part of the research question, because this knowledge might influence participants' performance. Thus, two additional ID measures were included in this pilot study, a Paper Folding and Cutting test (which was administered as a pre/post-test) and the Nonverbal Reasoning sub-test of the Dyslexia Adult Screening Test (administered at the start of the experiment session).[1] The pilot study sessions took between 75-90 minutes for participants to complete. The pilot session structure is shown in Figure 4.1.

[1]These tests are described in more detail in section 4.4.

Table 4.1: Structure of the pilot study sessions – Study 2

<table>
<tr><th colspan="6">Pre-Learning Phase / Individual Differences Pre-tests</th></tr>
<tr><td>10-15 min</td><td>Instructions, consent form</td><td>Phon. WM (CNRep)</td><td>Nonverbal Reasoning</td><td>Mood pre-test (PANAS)</td><td>Paper fold & cut pre-test</td></tr>
<tr><th colspan="6">Individual Differences Pre-tests</th></tr>
<tr><td>15 min</td><td colspan="2">MLAT-style Lang. pre-test (Structure & Memory)</td><td colspan="2">Musical Ability pretest (Receptive & Productive)</td><td>M-C Hung. vocab. pre-test</td></tr>
<tr><th colspan="6">Learning Phase</th></tr>
<tr><td>20 min</td><td colspan="2">3 practice phrases (2x)</td><td>Learning 1</td><td>Learning 2</td><td>Learning 3</td></tr>
<tr><th colspan="6">Testing Phase</th></tr>
<tr><td>10 min</td><td>Hungarian Production</td><td>English Recall</td><td>Hungarian Recognition</td><td colspan="2">Multiple-Choice Hungarian vocabulary posttest</td></tr>
<tr><th colspan="6">Individual Differences Posttests</th></tr>
<tr><td>10-15 min</td><td>LEQ & MEQ</td><td>Mood post-test (PANAS)</td><td>Delay Recall Hung. Conv.</td><td colspan="2">MLAT-style Lang. posttest (Structure & Memory)</td></tr>
<tr><td>10-15 min</td><td>Paper fold & cut posttest</td><td colspan="2">Musical Ability posttest (Receptive & Productive)</td><td>Debriefing Questions</td><td>Debriefing Sheet</td></tr>
</table>

Results of the pilot study were informative, but should be interpreted with caution due to the small number of participants in each condition, coupled with the wide variations in performance between individuals that was observed in the previous experiment described in Chapter 3. Another difficulty when interpreting the pilot study results was that 6 of the 8 participants in the two passive learning groups considered using the strategy of repeating the phrases aloud to help them learn the phrases, and 4 participants actually tried this strategy (speaking and whispering could be heard on the audio recordings). In addition, one pilot study participant in the Listen to Speech group repeated the phrases in a sing-song manner, exaggerating the intonation contour as he attempted to remember the phrases. Thus, the results of the pilot study should not be considered indicative of performance in the passive learning procedure as it was designed. Since nearly everyone in the two passive learning groups considered repeating the phrases aloud as a learning strategy, instructions for the final study explicitly forbade repetition.[2]

[2]The full script used in the final design for this study is available in Appendix B on page 368.

Pilot study results for the Hungarian tests

In the pilot study, the raw average Hungarian test scores in the two active learning conditions (listen-and-repeat Speaking and Singing) were considerably higher than the corresponding scores in the previous version of the experiment (overall M = 8.1 in the previous study compared to M = 12.6 in the current pilot results). There was higher overall performance in this pilot study for both the Speaking condition (M = 14.9) and the Singing condition (M = 10.1), which had a low-performing outlier and the lowest average performance. The Listen to Speech condition had the second-best overall raw Hungarian performance score (M = 13.4), with performance in the Listen to Singing condition somewhat lower (M = 11.9), but also higher than the Singing condition in this pilot study.

In the pilot study, a two-way multivariate analysis of variance (MANOVA) comparing Hungarian test scores for type of stimulus (spoken or sung phrases) and type of learning condition (the two passive learning conditions, Listen to Speech and Listen to Singing, compared to the two active learning conditions, listen-and-repeat Speaking and Singing) showed a significant difference in performance for the Hungarian Recognition Test, $p < .01$, with the two passive conditions showing higher performance. A likely explanation for this result is that participants in the two passive learning conditions heard the Hungarian phrases a total of 9 times, while participants in the active learning conditions heard the correct Hungarian phrases a total of 6 times and then heard themselves repeat the phrases once (perhaps inaccurately) in Hungarian during each of the three learning sessions. No group differences between the active and passive learning conditions were found for the other four Hungarian tests in the pilot study (all ps = n.s.). There were no significant differences for type of stimulus (spoken or sung stimuli), nor any interactions between type of stimulus and type of learning condition on Hungarian test performance.

Since performance on the two spoken Hungarian tests remained quite low, with many pilot study participants earning zero correct, the decision was made to add one 5-minute learning session to the final study design in an attempt to improve spoken Hungarian test scores and increase participants' self-perceived success at learning the phrases. Thus, in the final study design, participants had a total of four 5-minute learning sessions.

Pilot study results for the individual differences measures

A few of the measures of individual differences showed significant group differences in the pilot study. Two-way analysis of variance (ANOVA) that included type of learning condition (active and passive) and type of stimulus (spoken and sung phrases) revealed that the passive learning groups had higher performance, p = .045, on the pre-test of productive musical ability for the sub-test that measured participants' ability to tap to the syllables of 'Happy Birthday.' However, no difference was observed for the MAT Happy Birthday: Tapping post-test. There were no significant interactions between type of learning and type of stimulus for this measure. The Singing condition had the lowest mean scores on the MAT Happy Birthday: Tapping sub-test (M = 3.0 out of 5); the Listen to Speech condition had the highest mean scores (M = 4.0 out of 5).

Two-way ANOVA results also showed a significant difference for type of stimulus (spoken or sung stimuli) on the Negative Affect Pre-test (p = .016) and Positive Affect Pre-test (n.s.), with the two sung groups reporting a more negative mood at the start of the experiment (mean pre-test mood scores in the listen-and-repeat Singing condition were the most negative). There was no interaction between type of stimulus and type of learning on the negative or positive mood pre-test measure, nor were there any group differences or interactions for negative or positive mood post-test scores at the end of the pilot experiment sessions.

On the Debriefing questionnaire at the end of the experiment, a significant group difference was found for type of learning condition, $p = .049$, with the active learning conditions reporting lower overall motivation to learn the phrases in this pilot study. There was also a significant interaction between type of learning condition and type of stimulus on this question, $p = .007$, with the Singing condition reporting the lowest motivation and the Listen to Singing condition reporting the highest motivation, while mean scores in the two spoken conditions were approximately equal for Motivation. Also on the Debriefing questionnaire, there was a marginal effect for self-reported success at learning the phrases ($p = .082$), with the two spoken groups reporting more successful learning. No interaction was found between type of learning condition and type of stimulus for self-reported success at learning the phrases.

The pilot study results showed a few significant group differences for the ID measures, but a consistent pattern of differences was not observed for particular types of ID measures (e.g., language learning or musical abilities). However, the Singing condition had the lowest scores on many ID measures in this pilot study, particularly for the Happy Birthday: Tapping sub-test of the MAT, mood, motivation, and success at learning the phrases. Taken together, this combination may in part explain the lower overall Hungarian test performance in the Singing group compared to the other three groups in this pilot study. Indeed, the pilot Singing condition performed at a lower level on all Hungarian and ID measures than the Singing condition in the previous study, apart from the Multiple-Choice Hungarian Vocabulary Post-test and the MLAT-style language test. Significant group differences between this pilot Singing condition and the Singing condition in the previous study were found for negative mood (both pre and post) and total musical aptitude test score at the $p < .01$ level, and for positive mood and receptive musical skills (combining scores for the rhythm, pitch, and melody discrimination sub-tests) at the $p < .05$ level, with the pilot Singing condition showing lower and more negative scores.[3]

[3]The t-test compared the musical ability post-test scores, because the ID measures were administered only as post-tests in the previous study (apart from mood).

By contrast, the pilot Speaking condition compared to the Speaking condition in the previous study showed an overall pattern of higher scores, with significant group differences found for the MLAT-style Language Memory test and overall MLAT-style language test score (comparison of post-test scores only), $p < .001$, with the pilot Speaking condition showing higher performance. Significant group differences in the opposite direction were found for the Hungarian Recognition Test and for receptive musical skills (rhythm, pitch, and melody discrimination) at the $p < .05$ level, with the pilot Speaking group showing lower scores on these measures.[4]

Pilot study repeated-measures ANOVA results for ID measures

Repeated-measures ANOVA comparing scores on the ID measures was conducted to explore whether any changes in performance from pre-test to post-test had been influenced by participation in the active vs. passive learning procedure or from hearing a different type of stimuli (spoken vs. sung phrases). A problematic finding was that there were significant improvements from pre-test to post-test for many of the ID measures that were designed to measure underlying skills (such as musical ability or language learning ability) and which, at least in theory, should be stable attributes over such a short period of time. The improvement in scores on these measures may be explained, in part, by the practice effect, where improvement in performance often occurs after a particular test has been re-administered, even if different items were used (Kaufman, 1994).

In this pilot study, repeated-measures ANOVA showed few significant group differences or interactions on the ID measures based on type of learning condition (active listen-and-repeat vs. passive listening only) or type of stimulus (spoken vs. sung Hungarian phrases). However, a significant main effect was found for type of learning condition on the MAT Pitch Discrimination sub-test, $p = .012$, with the two

[4]The t-test comparisons were for the language and musical ability post-test scores, because the ID measures were administered only as post-tests in the previous study (apart from mood).

active conditions showing a decrease in performance, and no interaction with type of stimulus. In addition, there was a significant main effect for type of stimulus on the Positive Affect measure, $p = .026$, with the two sung conditions showing a decrease in positive mood from pre- to post-session; there was no interaction with type of learning condition.

No significant group differences were found when looking at the ID results in the four learning conditions separately. Table 4.2 shows the repeated-measures ANOVA results for the ID measures in the four learning conditions (Listen to Speech, Listen to Singing, Speaking, and Singing). The table shows that there were no significant group differences between any of the learning conditions on the ID measures from pre-test to post-test, indicating that the improvements in performance from pre-test to post-test were not significantly different due to an influence of the learning condition to which participants had been assigned. Although there were marginal main effects on both the Pitch Discrimination test and the Positive Affect score, the *p*-values did not reach statistical significance ($p = .06$ and $p = .08$, respectively).

Table 4.2: Pilot study repeated-measures ANOVA for ID measures in the four learning conditions (signif. codes: *** 0.001, ** 0.01, * 0.05) – Study 2

ID Measure	*N*	df	**Pre to Post**		**By Condition**	
			F-stat.	*p*-value	*F*-stat.	*p*-value
Paper Folding & Cutting Test	16	3, 12	1.827	.20	0.323	.81
MLAT-style Language Test	16	3, 12	20.85	**.001** **	1.281	.33
Language Structure Test	16	3, 12	7.410	**.019** *	0.231	.87
Language Memory Test	16	3, 12	6.218	**.028** *	0.536	.67
Musical Ability Test (MAT)	16	3, 12	10.41	**.007** **	0.627	.61
Receptive Music	16	3, 12	24.42	**.000** ***	1.380	.30
Rhythm Discrimination	16	3, 12	2.028	.18	0.598	.63
Pitch Discrimination	16	3, 12	8.727	**.012** *	0.750	.06
Melody Discrimination	16	3, 12	6.193	**.029** *	0.541	.66
Productive Music	16	3, 12	10.71	**.007** **	1.571	.25
Happy Birthday: Sing	16	3, 12	2.000	.18	0.667	.59
Happy Birthday: Tap	16	3, 12	5.400	**.039** *	0.600	.63
Positive Mood	16	3, 12	8.022	**.015** *	2.926	.08
Negative Mood	16	3, 12	0.890	.36	0.235	.87

Final study design

The next few paragraphs summarise the final design of the experiment sessions in the main study, briefly describing the changes that were made on the basis of the pilot study results.

The final study was conducted with participants randomly assigned to one of four learning conditions using a controlled, experimental design: Listen to Speech, Listen to Singing, Speaking, and Singing. For the most part, the stimuli, the measures, and the procedure for teaching the Hungarian phrases were very similar to those used in the previous experimental study and are briefly described in section 4.2.5.

Many participants in all of the learning conditions performed at a very low level on the verbatim, spoken Hungarian tests, which were the two tests that showed a significant difference between the Singing condition and the other two conditions in the previous experiment, resulting in data sets that were not normally distributed. While making the spoken Hungarian tests easier might reduce the overall difference in scores in the four groups, it was important not to overwhelm participants during the learning process to such an extent that they gave up trying to learn the phrases. In an attempt to increase performance on the Hungarian tests, the learning period was increased to four learning sessions rather than only three learning sessions, resulting in a 20-minute learning period rather than the 15-minute period in this pilot study and in the previous experiment.

In addition, the order of the Hungarian tests was rearranged to help increase participants' feelings of success in learning the phrases. The English Recall Test was moved to the start of the Hungarian tests, followed by the more difficult Hungarian Production Test, and finally the Hungarian Recognition Test. The Hungarian Multiple- Choice Vocabulary Post-test was removed because it did not show any group differences in the pilot study results or in the previous experiment. Thus, the

Delayed- Recall Hungarian Conversation was the final Hungarian test used in the data analysis. However, a new Hungarian vocabulary test was included at the very end of the experiment session, after the Delayed-Recall Hungarian Conversation, which was designed to be shorter and easier than the Multiple-Choice Hungarian Vocabulary Post-test. This final test was intended to help participants realise that they had successfully learned a few phrases in the new language, even if they had not managed to learn all 20 phrases that had been presented.[5]

The wording of the Debriefing sheet that participants read at the end of the experiment session was also changed to make it clear that the Hungarian tests were very difficult for everyone, with a reported average score of only 20%. It also stated that Hungarian is a very challenging language to learn and that the task was designed to be very difficult. The hope was that this would encourage more participants to report that they felt at least 'somewhat successful' in learning the Hungarian phrases and the English meanings on the Debriefing Questionnaire.

The group difference in performance for the pilot study on the Hungarian Recognition Test had to be addressed for the final study design, since the passive learning groups' significantly higher performance on this measure was likely due to having heard the correct Hungarian phrases three times in a row rather than only twice. The decision was made to play each English phrase followed by the Hungarian translation three times in all four learning conditions. For the two active learning conditions (which had heard the Hungarian phrases only twice), participants repeated each Hungarian phrase aloud the third time they heard it played during each of the four learning sessions.[6]

Also based on the pilot study results, the decision was made to remove a few ID post-test measures for the final version of the experiment. Since none of the IDs post-

[5]This vocabulary test is described in more detail in section 4.2.5.

[6]Because half of the participants in the two 'passive learning' conditions in the pilot study did this and had high mean Hungarian test scores, it was assumed that this small adjustment to the learning procedure would not be difficult to do or have a detrimental effect on learning.

tests appeared to have been influenced differentially by the learning procedure in the four different groups, most of the post-test measures that had been administered in the pilot study as IDs pre/post-tests could be removed from the final study design because there were no significant group differences. A reduction in the number of IDs post-tests was important because this allowed a time reduction of approximately 20 minutes for each experiment session in the final study design. This was necessary due to the inclusion of the fourth 5-minute learning session that was added to improve Hungarian test performance. The aim was to reduce the total duration of the experiment sessions to approximately 75 minutes, to reduce participant fatigue.

While repeated-measures ANOVA did show a significant group difference for type of stimulus ($p < .01$) on the Pitch Discrimination test, this measure was not included as a pre/post-test in the final study because each Pitch Discrimination test consisted of only 4 items and therefore it was difficult to have a great deal of confidence in the significant result. The brief Paper Folding and Cutting test was also removed because although there may have been an interesting trend of improvement for participants in the listen-and-repeat Singing condition (pre-test $M = 2.0$ to post-test $M = 3.3$ out of 4 possible, $p =$ n.s., but with much smaller improvements in scores observed in the other three groups), the effects of singing on nonverbal reasoning was not one of the primary research questions for the current study and including more than four items on the Paper Folding and Cutting pre-test and post-test would have required considerably more time.

Thus, the ID measures that were retained as pre/post-tests for the main study included the positive and negative mood (PANAS) questionnaire (significant group difference for type of stimulus, $p < .05$) and the MLAT-style language test. The latter did not show significant group differences, but there was a trend for an overall difference on this measure for the active vs. the passive groups ($p = .09$). The inclusion of an MLAT-style pre/post-test for the final study was also important for ascertaining the reason for the marginal group difference found in the previous experiment on the Language Memory test between the Speaking and Singing conditions, since a definitive answer to this outstanding question could not be expected in the pilot study which had only four participants in each group. A chart outlining the final study procedure is shown in subsection 4.2.6 (see page 111).

4.2.3 Participants

Data collected for this study included 80 adult student participants who were randomly assigned to one of four learning conditions, with 20 participants in each group (10 male and 10 female). Four participants were excluded from the data set due to scoring higher than 50% on the Hungarian vocabulary pre-test and two participants were excluded on the basis of age (mature students aged 30 years or older).

Unfortunately, there were technical difficulties with the website for this experiment which resulted in missing data on the Hungarian Production Test for all 20 participants who were assigned to the passive Listen to Speech condition. It was not considered feasible to exclude the Hungarian Production Test from the analysis because this test was one of two verbatim, spoken Hungarian tests that showed a significant benefit for the Singing condition in the previous experimental study. Thus, an additional 20 participants were recruited to take part in the Listen to Speech condition of the experiment. Most of the participants in the second Listen to Speech condition took part after the exam period had ended, which may have influenced their performance.

The mean age for participants in the four groups was 21.6 years. The listen-and-repeat Singing group had the lowest mean age (21.0 years) and the smallest age range. The Listen to Speech group had the highest mean age (22.2 years), with the mean age for the other two groups falling in between. Further details are shown in Table 4.3. ANOVA showed a marginal group difference for type of stimulus, $p = .078$, with the two sung conditions younger overall than the two spoken conditions.

Table 4.3: Age of participants – Study 2

Group	*N*	*M*	*SD*	Range
Listen to Speech	20	22.2 years	2.24	18 - 28
Listen to Singing	20	21.4 years	2.64	18 - 28
Speaking	20	22.0 years	2.59	18 - 28
Singing	20	21.0 years	1.76	18 - 25
Active	40	21.5 years	2.34	18 - 28
Passive	40	21.8 years	2.46	18 - 28
Spoken stimuli	40	22.1 years	2.48	18 - 28
Sung stimuli	40	21.2 years	2.23	18 - 28
Overall (4 groups)	80	21.6 years	2.34	18 - 28

Participants' performance on the measures of individual differences showed that for several of these pre-existing factors, the groups were not particularly well matched in terms of type of stimulus (spoken or sung stimuli) and type of learning (active listen-and-repeat or passive listening). A consistent pattern was found for the two sung stimuli conditions to perform at a higher level than the two spoken stimuli conditions on the language-related ID measures, while the two spoken conditions showed higher overall performance on the music-related ID measures. Although in most cases the differences in scores were quite small, it is possible that this overall pattern could have an impact on the Hungarian test results. Consistent with the pilot study results, repeated-measures ANOVA calculated for the positive and negative mood pre/post-session scores and for the MLAT-style language pre/post-tests showed significant changes from pre- to post-session, but there were no significant group differences or interactions. The ID results for type of stimulus and type of learning condition are discussed in more depth in section 4.4.

One-way ANOVA results[7] showed no significant differences between the four groups for phonological working memory, nonverbal reasoning, mood, language learning ability or experience, musical experience, motivation to learn the phrases, or

[7]Results of two-way ANOVAs for the ID measures are described in section 4.4, but for simplicity those details are not discussed here.

for self-perceived success at learning the phrases (all ps = n.s.). However, one-way ANOVA also showed that the listen-and-repeat Singing condition had significantly lower performance than the other three groups for productive musical skills and in particular for singing 'Happy Birthday,' both at the $p < .05$ level. Table 4.4 shows an ANOVA table for participants' ID test scores in the four groups.

Table 4.4: ANOVA for ID measures in the four learning conditions (signif. code: * .05) – Study 2

ID Measure	*N*	df	Sum Sq.	Mean Sq.	*F*-stat.	*p*-value
Age	80	3, 76	20.1	6.7	1.179	.32
Nonverbal Reasoning (DAST)	80	3, 76	3.1	1.0	0.594	.62
Phonological WM	80	3, 76	15.8	5.3	2.104	.11
Language Experience Quest.	80	3, 76	187.3	62.4	1.597	.20
MLAT-style Pre-test Total	80	3, 76	66.7	22.2	1.161	.33
MLAT-style Pre-test Structure	80	3, 76	1.1	0.4	0.179	.91
MLAT-style Pre-test Memory	80	3, 76	51.5	17.2	1.227	.31
MLAT-style Post-test Total	80	3, 76	31.8	10.6	0.622	.60
MLAT-style Post-test Structure	80	3, 76	8.7	2.9	1.681	.18
MLAT-style Post-test Memory	80	3, 76	11.5	3.8	0.308	.82
Musical Experience Quest.	80	3, 76	13.7	4.5	0.105	.96
Musical Ability Test Total	80	3, 76	40.0	13.3	1.945	.13
MAT Receptive	80	3, 76	4.5	1.5	0.369	.78
MAT Rhythm Discrimination	80	3, 76	1.5	0.5	0.301	.83
MAT Pitch Discrimination	80	3, 76	4.3	1.4	1.803	.15
MAT Melody Discrimination	80	3, 76	0.3	0.1	0.150	.93
MAT Productive	80	3, 76	22.2	7.4	3.671	**.016 ***
Happy Birthday: Sing	80	3, 76	2.000	0.18	0.667	**.035 ***
Happy Birthday: Tap	80	3, 76	5.0	1.7	1.998	.12
Positive Mood Pre-session	80	3, 76	173.4	57.8	1.510	.22
Negative Mood Pre-session	80	3, 76	19.5	6.5	0.492	.69
Positive Mood Post-session	80	3, 76	93.4	31.1	0.593	.62
Negative Mood Post-session	80	3, 76	0.8	0.5	0.012	.99
Motivation to learn phrases	80	3, 76	1.6	0.5	0.714	.55
Change in Motivation	80	3, 76	1.34	0.45	0.761	.52
Success: Hungarian phrases	80	3, 76	1.9	0.6	1.238	.30
Success: English phrases	80	3, 76	1.9	0.6	1.061	.37

4.2.4 Stimuli

The stimuli used in this study were identical to those used in the first experimental study. The 20 Hungarian phrases were chosen for inclusion based on frequency of use and utility in conversation (e.g., 'Good day,' 'I don't understand,' 'Could you repeat that, please?') to enhance participants' desire to learn the phrases. Repetition of certain words within the phrases permitted the analysis of which English words corresponded to those words in Hungarian.[8] For reference, the full list of phrases is available in Appendix A on page 345.

The mean duration of the two types of Hungarian stimuli for this study (spoken and sung phrases) was approximately equal (M = 2.2 seconds), with the shortest, 2-syllable phrases lasting one second and the longest, 8-syllable phrases lasting four seconds. Since duration of the Hungarian phrases was an important factor to control experimentally (Kilgour et al., 2000), Student's t-test was conducted to compare the stimuli durations (in milliseconds) for the two types of Hungarian stimuli. Results showed a close relationship between the phrase durations for the spoken and sung stimuli, p = .80.[9]

To control for rate of presentation, the auditory stimuli were again presented at the same rate in all four learning conditions: English phrase, 1s pause, Hungarian phrase, 1s pause, Hungarian phrase, 1s pause, Hungarian phrase during which participants in the two passive learning groups listened to the phrase a third time, while the two active learning groups repeated the phrase aloud as they listened to the phrase. After a 2s pause, the next English phrase was presented, and so on.

[8] See page 50 in Chapter 3 for more details about how the stimuli were designed and recorded.

[9] Although this p-value shows more variation in the phrase durations than when the rhythmically spoken stimuli used in the previous study were also included (ANOVA showed p = .97), this was considered to be an acceptable level of similarity for the stimuli durations, particularly because it was not the case that all of the spoken phrases were shorter in duration than the sung phrases.

4.2.5 Measures and data analysis

Hungarian language tests

Six Hungarian language tests, five of which were developed for the previous study, were administered to assess participants' learning of the paired-associate English-Hungarian phrases in the pilot study and the main study reported in this chapter. However, only four of the six Hungarian tests were included in the statistical analyses, as described below. The scoring procedure for each Hungarian measure is also outlined in this section.[10]

- *Multiple-Choice Hungarian Vocabulary Pre-test.* Twenty forced-choice multiple-choice pre-test questions assessed whether participants had prior knowledge of basic words in Hungarian. A score higher than 50% resulted in the participant's data being excluded, due to the possibility that the participant knew some Hungarian words prior to beginning the study (four participants were removed for this reason). The vocabulary post-test was not included in the final study because it never showed any performance differences between the groups. Participants received one point for each correct answer, with a total of 20 points possible.

- *English Recall Test.* Participants heard the 20 Hungarian phrases as prompts – presented in a different, randomised order – and attempted to remember and say the English translation of the Hungarian phrase. Participants' responses were transcribed from the audio recordings and one point was awarded for each answer which had the correct English meaning, for a total of 20 possible points. A separate score was also calculated for the number of correct English words on this measure, but since this score was very similar to the overall score, it is not reported in the results.

[10]For reference, a complete list of the Hungarian test items is available in Appendix A.

- *Hungarian Production Test.* Participants heard the 20 English prompts from the learning sessions – presented in a different, randomised order – and attempted to remember and say the corresponding Hungarian phrase aloud verbatim, as best they could. Participants received one point for each verbatim Hungarian response which had all of the correct syllables, for a total of 20 possible points. Separate scores were also calculated for the number of correct syllables and the number of correct words spoken in Hungarian, but these scores showed similar results to the verbatim Hungarian Production Test score and therefore those additional scores are not reported in the results.

- *Hungarian Recognition Test.* Participants made same/different judgments for spoken versions of the 20 Hungarian phrases they had heard. Ten phrases were presented with all syllables in the original order, while the remaining 10 items swapped two syllables within the phrases. This resulted in 10 'different' phrases which all contained exactly the same syllables, but presented in a different order from the originals. Participants received one point for each correct answer, with a total of 20 possible points.

- *Delayed-Recall Hungarian Conversation.* After a 20-minute delay during which several ID measures were completed, participants attempted to have a brief conversation in Hungarian. Participants heard a Hungarian phrase on the audio recording and tried to reply using a logical response using any Hungarian words they knew. Participants were instructed to guess or to say 'I don't know' or 'I don't understand' in Hungarian if they were unsure how to reply. The recording consisted of 5 simple Hungarian phrases, separated by 8s pauses, which functioned as one side of a short conversation. Participants' responses were transcribed from the audio recordings and scored out of a possible 10 points, with two points awarded if the participant gave an appropriate reply to the previous Hungarian statement, while responses such as 'I don't understand' or ‘I don’t know’ received one point, and English replies, incorrect or unintelligible Hungarian phrases earned zero points.

- *Simple Hungarian Vocabulary Test.* This 5-item test showed participants a Hungarian phrase displayed above a photograph of a person, with the English translation of that phrase written below the picture. Participants had to choose one of two written Hungarian phrases as a response. This simple test was designed to improve participants’ perception of how well they had learned the phrases. Each Hungarian phrase had a preferred answer, but there was no ‘incorrect’ response for two of the five items. Thus, the results of this final vocabulary test were not included in the overall Hungarian test results.

4.2.6 Procedure

This study was designed to investigate a new research question regarding the effect of active vs. passive paired-associate foreign language learning through spoken vs. sung stimuli, while still providing a near-replication of the previous experimental study. Thus, the experimental procedure for the final study was only slightly different from that of the previous experiment and the pilot study. In this study, there were a total of four massed practice Hungarian learning sessions rather than three. In addition, most of the ID measures were only administered as pre-tests, and the nonverbal reasoning sub-test of the Dyslexia Adult Screening Test (Fawcett & Nicolson, 1998), which was used in the pilot study, was also administered as a pre-test. Only the MLAT-style Language Structure and Language Memory sub-tests and the PANAS mood questionnaire were administered as pre/post-tests in this study.

All participants were treated according to the ethical research standards published by the American Psychological Association (2002). Sessions were completed on an individual basis, with each participant taking approximately 75 minutes to complete the experiment. Table 4.5 shows a pictorial view of the procedure for the final experiment sessions. Participants were compensated £7 for their time.

Table 4.5: Structure of the experiment sessions – Study 2

Pre-Learning Phase / Individual Differences Pretests				
10 min	Instructions, consent form	Phon. WM (CNRep)	DAST Nonverb. Reasoning	Mood pre-test (PANAS)
Individual Differences Pretests				
10-15 min	MLAT-style Lang. pretest (Structure & Memory)		Musical Ability Test (Receptive & Productive)	
Learning Phase				
25 min	M-C Hung. vocab. pretest	3 practice phrases (2x)	Learning Sessions 1-4 (20 minutes)	
Testing Phase				
10 min	English Recall	Hungarian Production		Hung. Recog.
Individual Differences Posttests				
10 min	LEQ & MEQ	MLAT-style Lang. posttest (Structure & Memory)		Mood post-test (PANAS)
Final Hungarian Tests and Debriefing				
5-10 min	Delayed Recall Hung. Conv.	Simple Hung. Vocab. Test	Debriefing Sheet	Debriefing Questions

At the start of the session, each participant read a description about the experiment and signed an informed consent form.[11] Participants were told that the most important task was to learn 20 phrases in a foreign language, as well as the English meanings, as best they could. Participants then completed the 20 low-wordlike items of the CNRep (Archibald & Gathercole, 2006, p. 514) and the 8-item nonverbal reasoning sub-test of the Dyslexia Adult Screening Test (Fawcett & Nicolson, 1998) with the researcher. These measures were followed by brief background questions regarding the participant's age and gender, the pre-session PANAS mood questionnaire, MLAT-style language pre-tests and musical ability tests, which were presented through a PC desktop computer and noise-cancelling headphones.

Each participant then completed the 20-item Hungarian Multiple-Choice Vocabulary Pre-test before reading the on-screen instructions for the paired-associate

[11]Please see Appendix B for a copy of these materials.

foreign language phrase learning section. The participant then heard, or heard and repeated, three practice phrases in Hungarian (which were never used again) while the researcher was present and able to answer any questions. When the participant was confident that they understood the instructions, the researcher left the room while the participant completed four 5-minute learning sessions and the English Recall Test, the Hungarian Production Test, and the Hungarian Recognition Test by following written on-screen instructions.

The Hungarian tests were followed by the Language Experience Questionnaire, the Musical Experience Questionnaire, and the MLAT-style language post-tests before participants completed the post-session PANAS mood questionnaire. Finally, participants attempted to respond to the Delayed-Recall Hungarian Conversation test and completed the simple Hungarian Vocabulary test. When the session was complete, the researcher asked participants to read a debriefing sheet that described the study in more detail and then filled in a 9-item Debriefing questionnaire.[12]

4.3 Results

In this section, descriptive and inferential statistics exploring the main research question are reported. The prediction was that a listen-and-repeat learning procedure with sung phrases (Singing group) would be the most beneficial for paired-associate foreign language learning and that a passive, listening-only learning procedure with spoken phrases (Listen to Speech group) would show the lowest performance. No significant interactions between type of learning condition (active listen-and-repeat vs. passive listening only) and type of stimulus (spoken vs. sung phrases) were expected, because the two active learning conditions were both expected to perform at a higher level than the two passive listening conditions on the Hungarian measures.

[12]Please see Appendix B on page 371 for a copy of these materials.

Descriptive statistics

Raw scores on the Hungarian tests showed a different pattern than predicted. Scores were higher for the passive (listening only) groups in comparison to the active (listen-and-repeat) learning conditions, apart from the Delayed-Recall Hungarian Conversation (M = 2.6 points for both groups). The two spoken stimuli groups also had higher scores than the two sung stimuli groups, again apart from the Delayed-Recall Hungarian Conversation (M = 2.6 points for both groups). Comparing results in the four conditions showed that the passive Listen to Speech group had the highest performance on all Hungarian measures, rather than the lowest performance as was predicted. Table 4.6 shows the raw Hungarian test scores in the four groups, as well as active and passive learning conditions and spoken and sung stimuli conditions.

Normality testing

No outliers falling more than 1.5 standard deviations from the group mean were found in the different groups. Levene's test of homogeneity of variances confirmed that the groups were similar to one another in terms of the dispersion of Hungarian test scores, with none of the results showing a significant difference for any of the Hungarian tests. However, participants' scores on the Hungarian tests in the four groups did not generally show a normal distribution and in the current study, none of the Hungarian tests passed Shapiro-Wilk's test of normality in all four groups.

Type of Learning Condition. For the two active learning conditions, the Hungarian Production Test did not pass Shapiro-Wilk's test of normality at the $p < .001$ level, while the English Recall Test and the Delayed-Recall Hungarian Conversation did not pass Shapiro-Wilk's test of normality at the $p < .01$ level. By contrast, the Hungarian Recognition Test and the overall raw Hungarian test scores did pass Shapiro-Wilk's test of normality for the two active groups. For the two passive learning conditions, only the English Recall Test and the overall raw

Table 4.6: Raw Hungarian test results – Study 2

Group	**Listen to Speech**			**Listen to Singing**		
	N = 20			*N* = 20		
Hungarian Test	*M*	*SD*	Range	*M*	*SD*	Range
English Recall	9.5	3.5	4 - 15	6.7	3.7	1 - 17
Production	3.75	2.5	0 - 7	2.1	2.1	0 - 6
Recognition	15.9	2.0	13 - 20	14.6	2.0	10 - 19
Conversation	3.0	2.2	0 - 7	2.2	2.1	0 - 6
Overall Score	32.1	6.7	21 - 43	25.5	6.0	16 - 41

Group	**Speaking**			**Singing**		
	N = 20			*N* = 20		
Hungarian Test	*M*	*SD*	Range	*M*	*SD*	Range
English Recall	6.0	2.6	3 - 13	6.7	3.5	1 - 13
Production	1.6	1.8	0 - 8	2.4	2.2	0 - 7
Recognition	15.1	2.4	10 - 19	14.5	2.2	11 - 18
Conversation	2.3	2.1	0 - 6	2.9	1.9	0 - 6
Overall Score	24.8	5.5	17 - 41	26.5	7.6	14 - 36

Group	**Active (listen-and-repeat)**			**Passive (listen only)**		
	N = 40			*N* = 40		
Hungarian Test	*M*	*SD*	Range	*M*	*SD*	Range
English Recall	6.3	3.0	1 - 13	8.1	3.8	1 - 17
Production	2.0	2.0	0 - 8	2.9	2.4	0 - 7
Recognition	14.8	2.3	10 - 19	15.2	2.0	10 - 20
Conversation	2.6	2.0	0 - 6	2.6	2.1	0 - 7
Overall Score	25.7	6.6	14 - 41	28.8	7.0	16 - 43

Group	**Spoken stimuli**			**Sung stimuli**		
	N = 40			*N* = 40		
Hungarian Test	*M*	*SD*	Range	*M*	*SD*	Range
English Recall	7.7	3.5	3 - 15	7.4	3.7	1 - 17
Production	2.7	2.4	0 - 8	2.2	2.1	0 - 7
Recognition	15.5	2.2	10 - 20	14.5	2.0	10 - 19
Conversation	2.6	2.2	0 - 7	2.6	2.0	0 - 6
Overall Score	28.5	7.1	17 - 43	26.0	6.8	14 - 41

Hungarian test scores passed Shapiro-Wilk's test of normality. The other three Hungarian tests did not pass Shapiro-Wilk's test in the passive learning conditions as follows: Hungarian Production Test at the $p < .001$ level, Delayed-Recall Hungarian Conversation at the $p < .01$ level, and Hungarian Recognition Test at the $p < .05$ level. However, in both the active and passive learning groups, all of the Hungarian test scores had skewness and kurtosis values of less than 2, indicating that the deviance from normality was at an acceptable level and therefore permitting the use of ANOVA and MANOVA to analyse the data.

Type of Stimulus. Dividing the groups by type of stimulus, in the two spoken groups and in the two sung groups, the Hungarian Production Test did not pass Shapiro-Wilk's test of normality at the $p < .001$ level, and the Delayed-Recall Hungarian Conversation did not pass at the $p < .01$ level. In addition, the two spoken stimuli groups did not pass Shapiro-Wilk's test for the English Recall Test or for the overall mean Hungarian test score, both at the $p < .01$ level. Both the English Recall Test and the overall mean Hungarian test score passed Shapiro-Wilk's test of normality in the two sung stimuli groups. In addition, the Hungarian Recognition Test passed Shapiro-Wilk's test in both the spoken and sung stimuli groups. Because all of the Hungarian test scores had skewness and kurtosis values of less than 2 for the spoken and sung stimuli conditions, the deviance from normality was at an acceptable level so that ANOVA and MANOVA could be conducted on the data.

Four Conditions. When Hungarian test scores in the four learning conditions were investigated separately, the Hungarian Production Test barely passed Shapiro-Wilk's test of normality in the Listen to Speech and Singing conditions ($p = .069$ and $p = .079$, respectively) while the test did not pass Shapiro-Wilk's test at the $p < .01$ level in the Listen to Singing condition and the $p < .001$ level in the Speaking condition. The Speaking condition also did not pass Shapiro-Wilk's test for the English Recall Test at the $p < .01$ level. For the Hungarian Recognition Test, the

Listen to Singing condition did not pass Shapiro-Wilk's test of normality at the $p <$ .01 level. For the Delayed- Recall Hungarian Conversation, both the Listen to Singing and the Speaking conditions deviated from a normal distribution, both *p*s < .05. For the overall raw Hungarian scores, the listen-and-repeat Speaking and Singing conditions did not pass Shapiro-Wilk's test of normality at the $p < .05$ level. Table 4.7 shows a summary of the results of the Shapiro-Wilk's tests of normality.

Table 4.7: Results of Shapiro-Wilk's tests of normality on the Hungarian tests – Study 2

Group	**Listen to Speech**		**Listen to Singing**	
Hungarian Test	*N* = 20		*N* = 20	
English Recall	W = 0.94	*p* = .210	W = 0.94	*p* = .280
Production	W = 0.91	*p* = .069	W = 0.80	*p* = .001 ***
Recognition	W = 0.85	*p* = .005 **	W = 0.85	*p* = .006 **
Conversation	W = 0.92	*p* = .115	W = 0.87	*p* = .014 *
Overall Score	W = 0.93	*p* = .132	W = 0.97	*p* = .658

Group	**Speaking**		**Singing**	
Hungarian Test	*N* = 20		*N* = 20	
English Recall	W = 0.87	*p* = .010 *	W = 0.94	*p* = .280
Production	W = 0.72	*p* = .000 ***	W = 0.88	*p* = .016 *
Recognition	W = 0.79	*p* = .001 ***	W = 0.94	*p* = .204
Conversation	W = 0.87	*p* = .011 *	W = 0.92	*p* = .090
Overall Score	W = 0.85	*p* = .005 **	W = 0.93	*p* = .155

Group	**Active (listen-and-repeat)**		**Passive (listen only)**	
Hungarian Test	*N* = 40		*N* = 40	
English Recall	W = 0.92	*p* = .009 **	W = 0.97	*p* = .334
Production	W = 0.86	*p* = .000 ***	W = 0.89	*p* = .001 **
Recognition	W = 0.97	*p* = .354	W = 0.94	*p* = .040 *
Conversation	W = 0.90	*p* = .002 **	W = 0.90	*p* = .002 **
Overall Score	W = 0.99	*p* = .873	W = 0.97	*p* = .249

Group	**Spoken stimuli**		**Sung stimuli**	
Hungarian Test	*N* = 40		*N* = 40	
English Recall	W = 0.91	*p* = .003 **	W = 0.97	*p* = .130
Production	W = 0.88	*p* = .000 ***	W = 0.89	*p* = .001 **
Recognition	W = 0.96	*p* = .228	W = 0.97	*p* = .260
Conversation	W = 0.90	*p* = .003 **	W = 0.91	*p* = .003 **
Overall Score	W = 0.92	*p* = .009 **	W = 0.97	*p* = .299

The tests of normality also revealed an even more problematic finding. The Hungarian Production Test showed both kurtosis and skewness values of more than 2 in the Speaking condition (the Shapiro-Wilk's test was significant at the $p < .001$ level) and a kurtosis level greater than 2 for the overall raw Hungarian test score (significant at the $p < .05$ level). In the Listen to Singing condition, the Hungarian Recognition Test showed a kurtosis level higher than 2 (the Shapiro-Wilk's test was significant at the $p < .001$ level). This indicated that the results of ANOVA and MANOVA calculations on these data might not be informative, despite having equal numbers of participants in each group.

ANOVA vs. Kruskal-Wallis for each Hungarian test

Although the Hungarian test scores did not generally fit a normal distribution, comparison of the non-parametric Kruskal-Wallis rank sum test for each Hungarian test showed a similar pattern to the two-way ANOVA results for type of learning condition (active or passive) and for type of stimulus (spoken or sung stimuli). Similar values for the two statistical procedures were also found for one-way ANOVA in the four learning conditions.

Type of Learning Condition. A significant group difference was found for type of learning condition on the English Recall Test (both ANOVA and Kruskal-Wallis at the $p < .05$ level), with the passive groups showing higher performance. A marginal difference was found on the Hungarian Production Test (ANOVA $p = .065$, with the Kruskal-Wallis less significant, $p = .10$), also with the passive groups scoring higher. For the overall raw Hungarian test scores, ANOVA showed a significant difference at the $p < .05$ level while the Kruskal-Wallis result was marginal, $p = .06$; again, the passive learning group had higher performance than the active learning group. No significant differences for type of learning condition were found for the Hungarian Recognition Test or the Delayed-Recall Hungarian Conversation Test.

Type of Stimulus. A marginal difference was found for type of stimulus on the Hungarian Recognition Test (MANOVA $p = .055$ and $p = .062$ for Kruskal-Wallis), with the two spoken stimuli groups showing higher performance on this measure. None of the other Hungarian tests showed significant differences for type of stimulus, nor for the overall Hungarian test scores ($p = .12$).

Interaction between Type of Learning Condition and Type of Stimulus. The two-way ANOVA for each of the four Hungarian tests revealed significant interactions between type of stimulus and type of learning condition for the Hungarian Production Test and the overall raw Hungarian test score at the $p < .01$ level, and for the English Recall Test at the $p < .05$ level.

Four Conditions. Because significant interactions between the groups were found between type of stimulus and type of learning condition on several of the Hungarian tests, it was important to determine how each of the groups performed on these measures because this might help explain the interaction. One-way ANOVA and Kruskal-Wallis test results were compared for the Hungarian test scores in the four conditions, with the two statistical analyses showing similar values. Table 4.8 shows the ANOVA table and Table 4.9 shows a side-by-side comparison of the Kruskal-Wallis and ANOVA results.

Table 4.8: One-way ANOVA for Hungarian tests in the four groups (signif. codes: ** 0.01, * 0.05) – Study 2

Hungarian Test	*N*	df	Sum Sq.	Mean Sq.	*F*-statistic	*p*-value
Production	80	3, 76	53.2	17.7	3.915	.012 *
English Recall	80	3, 76	147.0	49.0	4.380	.007 **
Recognition	80	3, 76	23.5	7.8	1.726	.169
Conversation	80	3, 76	10.6	3.5	0.820	.487
Overall Score	80	3, 76	663.0	221.0	5.260	.002 **

Table 4.9: Comparison of Kruskal-Wallis and one-way ANOVA results for Hungarian tests in the four groups (signif. codes: ** 0.01, * 0.05) – Study 2

Test	**Kruskal-Wallis**			**ANOVA**		
(Total N = 80)	df	chi-squared	p-value	df	F-statistic	p-value
Production	3	9.574	.023 *	3, 76	3.915	.012 *
English Recall	3	11.44	.010 *	3, 76	4.380	.007 **
Recognition	3	4.945	.18	3, 76	1.726	.17
Conversation	3	2.374	.50	3, 76	0.820	.49
Overall Score	3	12.91	.005 **	3, 76	5.260	.002 **

MANOVA test results for the spoken, verbatim Hungarian tests

Two-way factorial MANOVA for the spoken, verbatim Hungarian measures (Hungarian Production Test and Delayed-Recall Hungarian Conversation) comparing the effects of type of learning condition and type of stimulus showed no significant main effects for learning type or stimulus type, but there was a significant overall interaction, $p < .05$. Post-hoc tests showed a significant interaction between type of stimulus and type of learning condition for the Hungarian Production Test, $p = .009$, with the passive Listen to Speech condition showing significantly higher performance than the listen-and-repeat Speaking condition.

MANOVA test results for the meaning-based Hungarian tests

Two-way factorial MANOVA for the English Recall and Hungarian Recognition tests showed a marginal trend for a group difference in terms of learning type (p = .06), stimulus type ($p = .09$), and for the interaction ($p = .06$). Post-hoc tests showed a significant group difference for learning type on the English Recall Test, $p < .05$, with the passive conditions showing higher scores, and a significant interaction was also found for this test, $p < .05$. Post-hoc analysis showed that the largest difference in performance was between the passive Listen to Speech condition (with the highest mean score) and the active, listen-and-repeat Speaking condition. In addition, a trend was found for type of stimulus on the Hungarian Recognition Test, $p = .056$, with the two spoken conditions showing higher performance; no interaction between type of learning condition and type of stimulus was observed on this measure.

MANOVA test results for all four Hungarian tests

Two-way factorial MANOVA for all of the Hungarian measures comparing the effects of type of learning condition (active or passive) and type of stimulus (spoken or sung) showed no significant main effects for learning type or stimulus type, but there was a trend for an overall interaction ($p = .074$). Post-hoc results showed a significant difference for learning type on the English Recall Test and for the overall Hungarian test score at the $p < .05$ level, and a marginal difference for the Hungarian Production Test, $p = .056$, with the passive groups showing higher performance. For type of stimulus, there was a marginal group difference for the Hungarian Recall Test, $p = .056$, with the spoken conditions showing higher performance on this measure. However, the overall advantage for the passive groups was primarily due to much higher mean scores in the passive Listen to Speech group, since significant interactions were found between type of learning condition and type of stimulus for the Hungarian Production Test and for the overall raw Hungarian test score, both at the $p < .01$ level, and for the English Recall Test at the $p < .05$ level (whereas the passive Listen to Singing group tended to have scores that were more in line with the two active learning groups). Graphs of the mean Hungarian test scores in the different groups are shown in Figures 4.1 to 4.5.

Figure 4.1: Mean scores for Hungarian Production Test (20 points possible) – Study 2

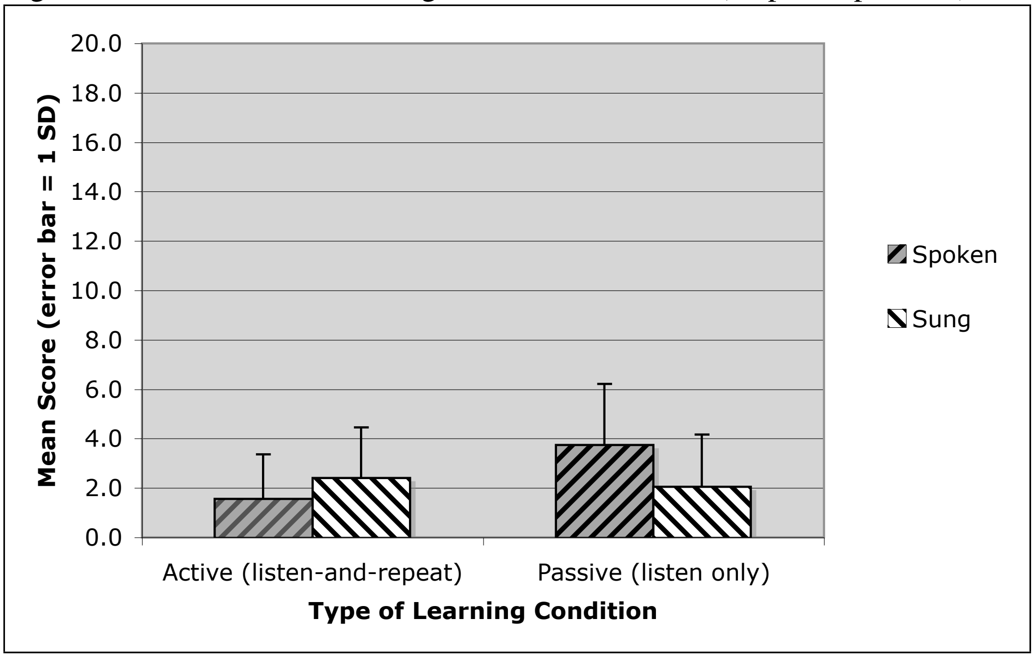


Figure 4.2: Mean scores for English Recall Test (20 points possible) – Study 2

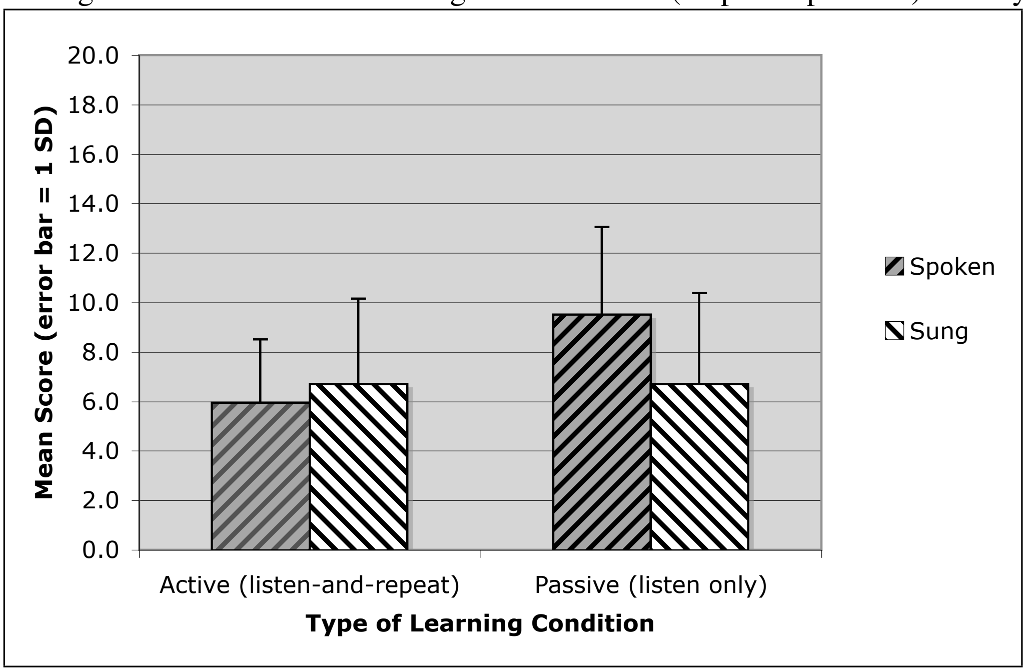


Figure 4.3: Mean scores for Hungarian Recognition Test (20 points possible) – Study 2

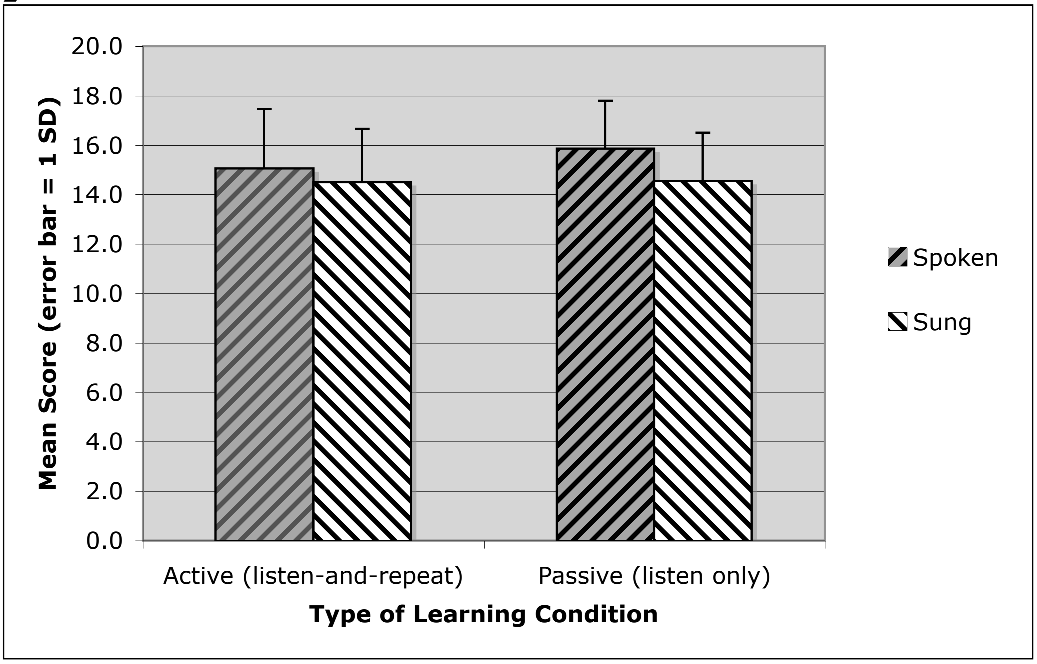


Figure 4.4: Mean scores for Delayed-Recall Hungarian Conversation (10 points possible) – Study 2

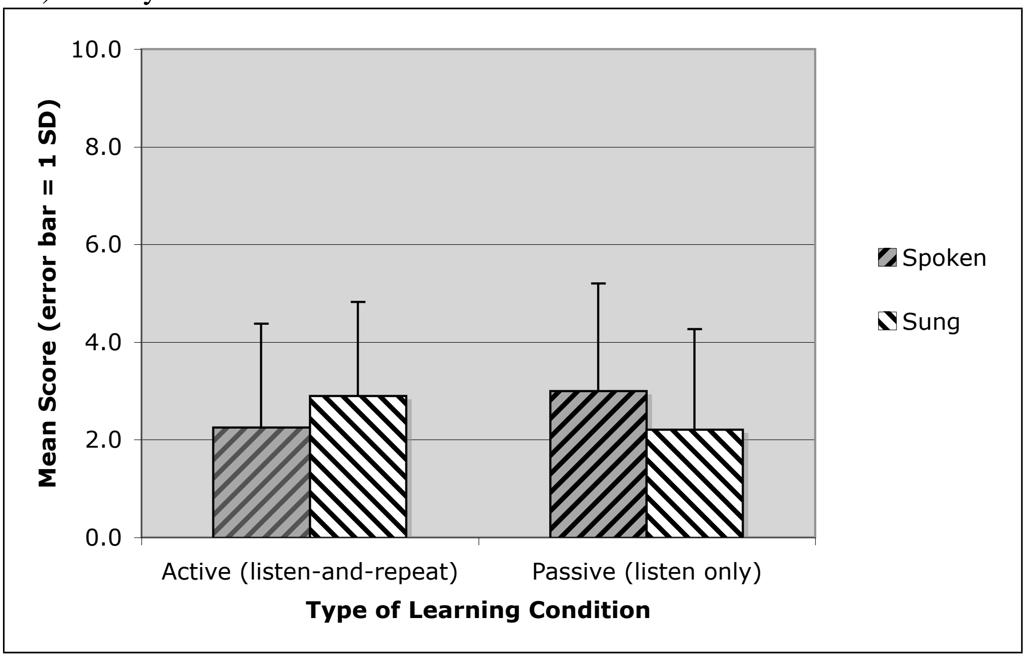
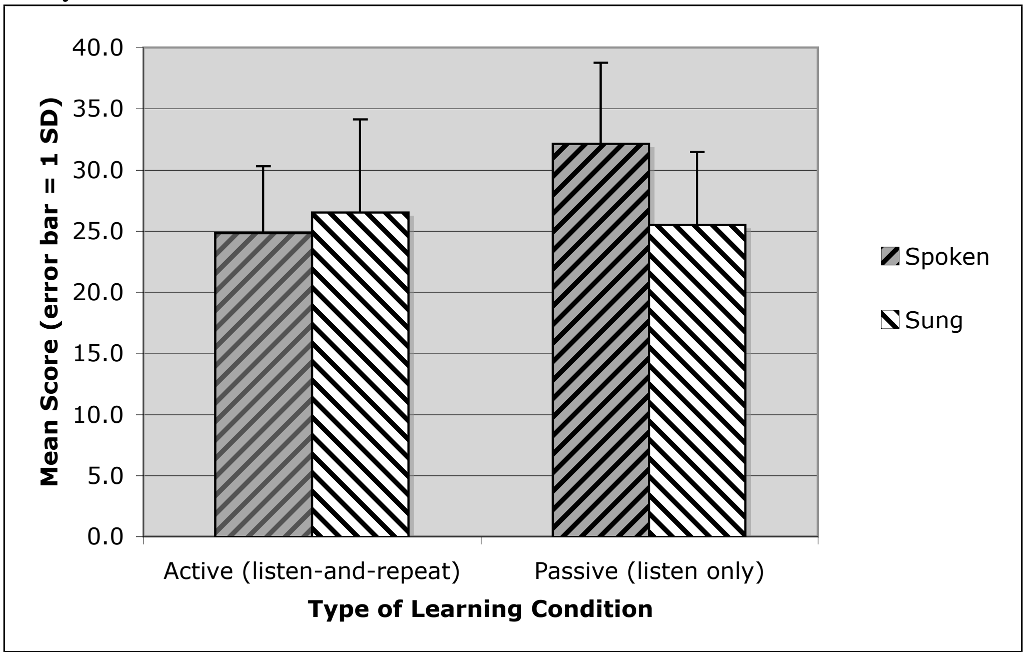


Figure 4.5: Mean scores for overall raw Hungarian test score (70 points possible) – Study 2



Thus far, the results of this experiment have shown a significant main effect for learning condition, but not in the direction that was predicted. The passive Listen to Speech condition showed the highest performance on all of the Hungarian measures even though it was expected to perform at the lowest level. A marginally significant difference was found on the Hungarian Recognition Test, with the passive learning groups and the spoken stimuli groups showing higher performance. Finding a significant difference for MANOVA on the English Recall and the Hungarian Recognition tests was unexpected because the previous study showed no group differences on those measures. In this study, the largest group difference was found between the two spoken stimuli conditions and in the opposite direction than was predicted, with the passive Listen to Speech condition showing the highest performance on all Hungarian measures.[13] By contrast, the sung stimuli groups and the active learning groups tended to have similar, lower performance on the Hungarian tests in this experiment.

[13]This pattern was true apart from the Hungarian Recognition Test, on which participants in the listen-and-repeat Speaking group performed at a similar level to the passive Listen to Speech group, with both groups scoring higher than the sung stimuli conditions.

The individual differences measures and results are described in more depth in the next section, while section 4.5 investigates the influence of the ID factors on participants' Hungarian phrase learning.

4.4 Measures of individual differences

All participants in this study completed several measures of individual differences, which were administered to establish whether the groups were well matched in terms of potentially confounding factors such as age, language learning experience and ability, musical experience and ability, mood, phonological working memory, nonverbal reasoning, and handedness. This section briefly describes the measures of individual differences[14] and presents the results in more detail than was outlined earlier (in the Participants subsection 4.2.3). This section also explores whether there were any differences or interactions for gender, which was balanced in the four conditions, and investigates the possibility that a few individuals with undiagnosed dyslexia took part in the study. Responses to several debriefing questions that were developed to provide important information about participants' motivation to learn and self-perceived success at learning the foreign language phrases are also discussed. As mentioned previously, a few pre-existing group differences were found for the potentially confounding ID factors, as discussed in more depth below.

4.4.1 Language learning ability and experience

Participants completed a brief language learning experience questionnaire and a brief language aptitude pre/post-test, adapted from previous research (Oxford, 1996; Ludke, 2006; Gilleece, 2006) for use in this study, to assess whether participants in

[14]Please see Chapter 3 for a full description of these individual differences measures and the scoring procedures.

the four learning conditions were matched in terms of previous language learning experience.[15]

Language Experience Questionnaire (LEQ). Participants' scores on the first, Likert-style LEQ section fell into a normal distribution and ranged from 17 to 44 (the lowest possible score was 10 and the highest possible score was 50), with an overall mean of 31.1 and standard deviation of 6.3. The Listen to Singing group had the highest mean score on this measure and the Listen to Speech group had the lowest mean score. A trend for a main effect for type of stimulus was observed with two-way ANOVA, $p = .07$, with the two sung stimuli conditions showing higher mean scores on the first LEQ section.

Scores on the second section of the LEQ were also variable, ranging from 11 to 35 with an overall mean of 22.3 points and a standard deviation of 5.9 (between 6 and 42 points were possible). The Listen to Speech condition had the lowest mean score on this measure, while the listen-and-repeat Speaking condition had the highest mean. However, no main effects or interactions between type of stimulus and type of learning condition were observed on the second sub-section of the LEQ.

Participants' scores on the two sub-sections of the LEQ showed a significant positive correlation (Pearson's $r(78) = 0.61$, $p < .001$), but no main effects or interactions were observed when using two-way ANOVA to compare the total overall LEQ scores[16] in the different groups.

No main effect for gender was observed on the Language Experience Questionnaire. In the two spoken stimuli groups, women tended to have higher scores than men overall on the LEQ, and there was about the same level of variability for both genders (women's $M = 54.9$, $SD = 11.4$ compared to men's $M = 49.6$, $SD =$

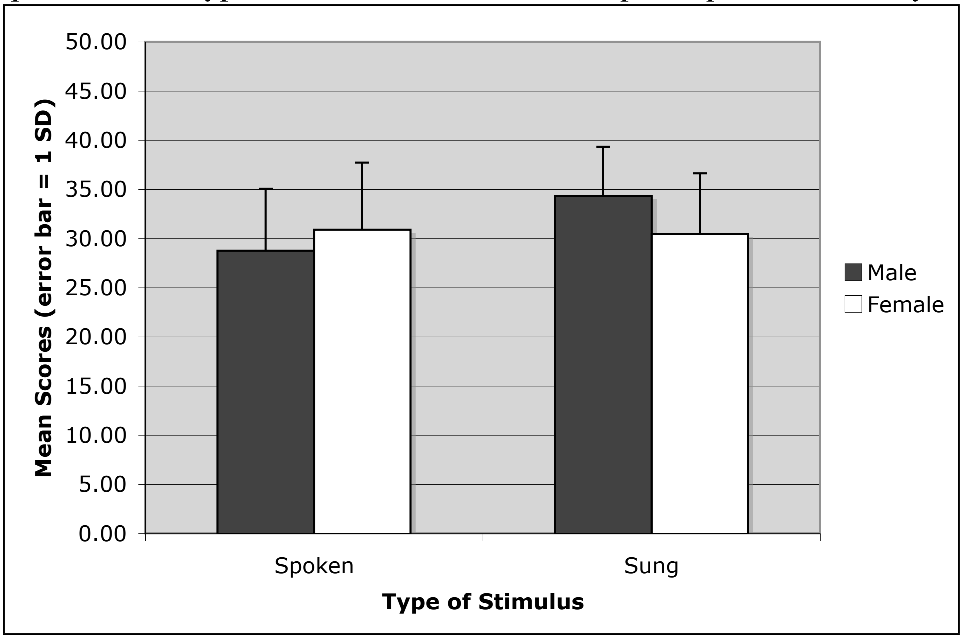
[15]Please see page 62 in the previous chapter for a full description and the scoring procedure for the LEQ.
[16]This total was calculated by adding together the scores from both LEQ sub-sections.

11.5). By contrast, women in the two sung stimuli conditions had lower and more variable scores (M = 53.1, SD = 11.4) than men (M = 56.7, SD = 8.9). Three-factor ANOVA showed a trend toward an interaction between gender and type of stimulus for the LEQ, p = .09, and separating scores for the two LEQ sub-sections revealed a significant interaction between gender and type of stimulus for the Likert- style LEQ questions, p = .033, with men in the two sung stimuli groups reporting greater language learning experience than women. No effects or interactions for gender were observed on the second LEQ sub-section. Table 4.10 shows the descriptive statistics for Language Experience Questionnaire scores in the different learning conditions. Figure 4.6 shows the mean scores in the spoken and sung stimuli groups for both genders on the first LEQ section.

Table 4.10: Language Experience Questionnaire total scores – Study 2

Group	*N*	*M*	*SD*	Range
Listen to Speech	20	50.1	10.7	32 - 69
Listen to Singing	20	56.3	8.8	40 - 72
Speaking	20	53.8	12.3	39 - 79
Singing	20	53.6	11.6	32 - 75
Active (listen-and-repeat)	40	53.6	11.8	32 - 79
Passive (listening only)	40	53.2	10.2	32 - 72
Spoken stimuli	40	51.9	11.5	32 - 79
Sung stimuli	40	54.9	10.3	32 - 75
Overall (4 groups)	80	53.4	10.8	32 - 79
Overall: Male	40	53.1	11.3	32 - 72
Overall: Female	40	53.7	11.0	32 - 79

Figure 4.6: Mean scores for the Language Experience Questionnaire (Likert-style questions) for Type of Stimulus and Gender (50 points possible) – Study 2



Language Ability Pre/Post-test. All participants also completed a brief language ability test which consisted of brief versions of two sub-tests of the Modern Language Aptitude Test (MLAT) created by Carroll and Sapon (1959). With permission, the items provided by Gilleece (2006) were used as a model and additional items were created by the researcher. This test provided a brief measure of participants' language learning ability and investigated whether language memory performance at post-test had been influenced by participation in the musical learning conditions.[17]

Language Structure Pre/Post-test: Words-in-Sentences. The multiple-choice language structure pre-test and post-test each consisted of 12 sentences with one word highlighted. In a different sentence, participants chose one option (of 5 possibilities) that served the same grammatical function as the highlighted word in the first sentence (adjective, adverb, and so on).

Language Memory Pre/Post-test: Number Pairs. The Language Memory pre-test consisted of the 12 Finnish numeral words used in the previous study. The post-test

[17]Please see page 64 in the previous chapter for a description of the scoring procedure for the Language Structure and the Language Memory pre-tests.

used number words from the Tlingit language, which is spoken in southeastern Alaska and western Canada. Several of the Tlingit words were modified by the researcher to keep the same number of syllables in both the Finnish and Tlingit number lists. The same arabic numerals were used for both the pre-test and post-test because the predictable structure in the original 12 Finnish number words could support learning during the 60-second learning phase. Memory interference due to memorising number words for the same numerals on both tests would be the same in all four learning conditions, so using the same numerals at post-test was not expected to show differential effects for learning condition.

In general, pre-test and post-test scores on the Language Structure and Language Memory tests did not fit a normal distribution using Shapiro-Wilk's test of normality, with scores tending to show a left-skew towards the higher end of performance due to a possible ceiling effect. However, since all skewness and kurtosis values fell into the acceptable range of less than 2, analysis of variance could be conducted.

Similar to the results of the pilot study, repeated-measures ANOVA showed significant improvements for the Language Structure and the Language Memory scores from pre-test to post-test, significant at the $p < .05$ level for the Language Structure test and at the $p < .01$ level for the Language Memory test. This significant improvement in scores was slightly problematic because the tests were designed to assess a theoretically stable measure of underlying language learning ability; however, the overall improvement in post-test scores can be explained by the practice effect (Kaufman, 1994). Tables 4.11 and 4.12 show the descriptive statistics of participants' raw scores on the MLAT-style language ability pre-tests and post-tests.

The listen-and-repeat Singing condition had the highest mean scores on both the Language Structure and Language Memory pre-tests and post-tests, while the listen-and-repeat Speaking condition had the lowest performance. Two-way ANOVA comparing results on the language ability tests for type of stimulus and type of

Table 4.11: MLAT-style Language Structure test scores – Study 2

Group	**Structure Pre-test**	*N*	*M*	*SD*	Range
Listen to Speech		20	9.1	1.4	6 - 11
Listen to Singing		20	9.2	1.2	7 - 11
Speaking		20	9.0	1.8	4 - 11
Singing		20	9.3	1.3	7 - 11
Active (listen-and-repeat)		40	9.1	1.6	4 - 11
Passive (listening only)		40	9.1	1.3	6 - 11
Spoken stimuli		40	9.0	1.6	4 - 11
Sung stimuli		40	9.2	1.3	7 - 11
Overall (4 groups)		80	9.1	1.4	4 - 11
Overall: Male		40	9.3	1.3	7 - 11
Overall: Female		40	9.0	1.6	4 - 11

Group	**Structure Post-test**	*N*	*M*	*SD*	Range
Listen to Speech		20	9.1	1.7	4 - 11
Listen to Singing		20	9.9	1.1	7 - 11
Speaking		20	9.2	1.2	7 - 11
Singing		20	9.7	1.2	7 - 11
Active (listen-and-repeat)		40	9.4	1.2	7 - 11
Passive (listening only)		40	9.5	1.5	4 - 11
Spoken stimuli		40	9.1	1.4	4 - 11
Sung stimuli		40	9.8	1.1	7 - 11
Overall (4 groups)		80	9.5	1.3	4 - 11
Overall: Male		40	9.6	1.4	4 - 11
Overall: Female		40	9.3	1.3	7 - 11

learning condition showed no differences for the Language Memory Pre-test or Post-test or for the Language Structure Pre-test. However, a main effect was found for type of stimulus on the Language Structure Post-test, p = .03, with the two sung conditions scoring higher than the two spoken conditions. No significant interactions were found between type of stimulus and type of learning condition on the language ability tests.

No main effects or interactions for gender were observed for the language ability measures, although men tended to outperform women on both tests by a small margin, except in the Singing condition on the Language Memory test. In the Singing

Table 4.12: MLAT-style Language Memory test scores – Study 2

Group	Memory Pre-test	*N*	*M*	*SD*	Range
Listen to Speech		20	7.1	3.7	0 - 11
Listen to Singing		20	7.6	4.1	0 - 12
Speaking		20	6.7	3.6	1 - 12
Singing		20	8.8	3.5	0 - 12
Active (listen-and-repeat)		40	7.7	3.7	0 - 12
Passive (listening only)		40	7.4	3.9	0 - 12
Spoken stimuli		40	6.9	3.6	0 - 12
Sung stimuli		40	8.2	3.8	0 - 12
Overall (4 groups)		80	7.5	3.8	0 - 12
Overall: Male		40	7.9	3.8	0 - 12
Overall: Female		40	7.2	3.7	0 - 12

Group	Memory Post-test	*N*	*M*	*SD*	Range
Listen to Speech		20	8.2	4.0	0 - 12
Listen to Singing		20	8.4	4.1	0 - 12
Speaking		20	8.4	2.9	1 - 12
Singing		20	9.2	2.9	1 - 12
Active (listen-and-repeat)		40	8.8	2.9	1 - 12
Passive (listening only)		40	8.3	4.0	0 - 12
Spoken stimuli		40	8.3	3.5	0 - 12
Sung stimuli		40	8.8	3.5	0 - 12
Overall (4 groups)		80	8.5	3.5	0 - 12
Overall: Male		40	8.8	3.2	1 - 12
Overall: Female		40	8.3	3.8	0 - 12

condition, men's scores were lower than women's on the Language Memory Pre-test (men's M = 7.7 vs. women's M = 9.9) and on the Post-test (men's M = 8.8 vs. women's M = 9.5).

In summary, the two sung stimuli conditions performed at a higher level than the two spoken conditions on all of the language-related ID measures, with a significant difference found on the Language Structure Post-test, $p < .05$. However, repeated-measures ANOVA analysing the changes in performance from pre-test to post-test on the language ability tests did not show significant group differences for type of stimulus or for type of learning condition, nor for the interaction.

A significant interaction between gender and type of stimulus (spoken or sung phrases) was observed on the first sub-section of the Language Experience Questionnaire, with men reporting higher scores than women in the sung learning conditions ($p < .05$). In the previous study, of all the language ability and experience measures, the LEQ score was the best predictor of participants' Hungarian test performance. Thus, the significant interaction between type of stimulus and gender for the LEQ could indicate that the men in the two sung conditions were predisposed to be more successful in learning the paired-associate English-Hungarian phrases compared to the women.

It was unexpected to find that the two sung conditions had slightly higher scores on the language ability and experience measures, because the Singing and Listen to Singing conditions showed lower performance on the Hungarian tests than the Listen to Speech condition. The influence of the language ability and experience measures on Hungarian test performance in this study are explored in section 4.5.

4.4.2 Musical ability and experience

Participants responded to a musical experience questionnaire (MEQ), identical to the one used in the previous study (see page 67), to assess whether participants in the four learning conditions were matched in terms of prior musical experience. Participants also completed a brief musical ability pre-test[18] which assessed receptive and productive musical skills. All sounds were played at a comfortable volume through noise-cancelling headphones.

Musical Experience Questionnaire. For the first section of the Musical Experience Questionnaire, mean scores were approximately the same in the different

[18] In the pilot study, participants also completed a post-test for this measure, which had the same number of items as the pre-test, but with different test items.

learning conditions.[19] The Speaking condition had the highest mean score (M = 37.8) and the Singing condition showed the lowest mean score (M = 36.7). Two-way ANOVA showed no significant group differences or interactions between type of stimulus and type of learning condition on the first section of the MEQ.

For the second part of the Musical Experience Questionnaire, scores ranged from 3 to 43 points (scores between 0 and 53 points were possible). No main effects or interactions between type of stimulus and type of learning condition were observed with two-way ANOVA on the second part of the MEQ. Participants' scores on the two MEQ sub-sections showed a small, but significant, positive correlation (Pearson's $r(78) = 0.40$, $p < .001$).

Overall scores on the MEQ[20] were quite similar in the different groups, but the active Speaking condition had the highest scores and also the most variability (M = 59.4, SD = 14.5). No group differences or interactions between type of stimulus and type of learning condition were found for participants' total Musical Experience Questionnaire scores. Table 4.13 shows the descriptive statistics in the different groups and by gender on the MEQ.

Table 4.13: Musical Experience Questionnaire total scores – Study 2

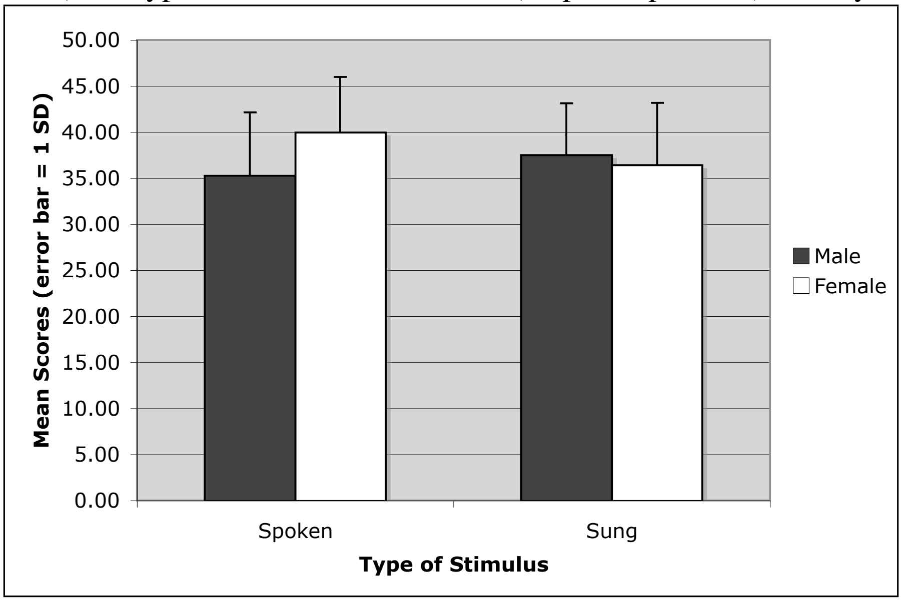
Group	*N*	*M*	*SD*	Range
Listen to Speech	20	55.6	10.5	35 - 76
Listen to Singing	20	57.8	13.7	29 - 76
Speaking	20	59.4	14.5	31 - 88
Singing	20	56.5	10.7	34 - 74
Active (listen-and-repeat)	40	57.9	12.7	31 - 88
Passive (listening only)	40	56.7	12.1	29 - 76
Spoken stimuli	40	57.5	12.6	31 - 88
Sung stimuli	40	57.1	12.1	29 - 76
Overall (4 groups)	80	57.3	12.3	29 - 88
Overall: Male	40	57.8	12.6	31 - 80
Overall: Female	40	56.8	12.2	29 - 88

[19] Scores on the first MEQ section did not show a very normal distribution based on Levene's test of equality of error variances, p = .013, but since ANOVA is robust against this violation when the group sizes are equal (as they were in this experiment), this is not of great concern.

[20] The total score was calculated by adding scores on the two MEQ sub-sections together.

No main effect for gender was found for the Musical Experience Questionnaire in this study, $p = .70$. However, three-factor ANOVA did show a significant interaction between type of stimulus and gender for the first section of the MEQ, $p = .048$, with the women in the two spoken conditions showing higher scores than the men (see Figure 4.7).[21] As shown on the next page in Figure 4.8, the largest gender difference was found in the Speaking condition, with women reporting more musical training and experience than men in that group; but the interaction between gender and the four learning conditions was not significant on the first section of the MEQ ($p = .18$) or overall ($p = .46$).

Figure 4.7: Mean scores on the Musical Experience Questionnaire (Likert-style questions) for Type of Stimulus and Gender (50 points possible) – Study 2



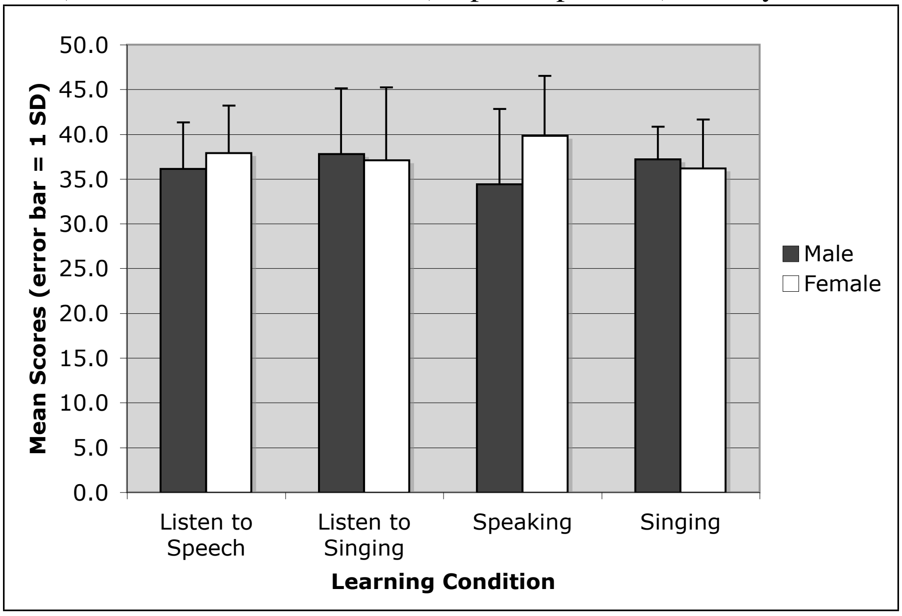
Receptive Musical Ability Test. The measure of receptive musical abilities consisted of three brief tests that were used in the previous experimental study: Rhythm Discrimination, Pitch Discrimination, and Melody Discrimination.[22] These receptive music tests were only administered as pre-tests in the final study.[23] Results

[21]For the overall MEQ scores, the interaction between type of stimulus and gender was not significant, $p = .12$, and for the second section of the MEQ, $p = .40$.

[22]Please see page 3.4.2 in the previous chapter for a description.

[23]By contrast, participants in the pilot study completed both a pre-test and post-test, which contained

Figure 4.8: Mean scores on the Musical Experience Questionnaire (Likert-style questions) for Condition and Gender (50 points possible) – Study 2



are presented for each musical sub-test, followed by a summary of the total receptive musical ability test scores.

Rhythm Discrimination. Scores on the Rhythm Discrimination test were very similar in the different groups, as shown in Table 4.14. Of the four groups, the Speaking condition had the highest mean score on the Rhythm Discrimination test and the Listen to Singing group had the lowest mean score. Two-way ANOVA with type of learning condition and type of stimulus showed no significant main effects or interaction on the Rhythm Discrimination Test.

In this study, a main effect for gender was found for the Rhythm Discrimination test, $p = .013$, with men showing higher performance on this measure. In addition, a significant three-way interaction between type of stimulus, type of learning condition, and gender was found for the Rhythm Discrimination test, at the $p < .001$ level. As shown in Figure 4.9, performance by women in the Listen to Singing and in the Speaking conditions was much lower than men's scores in those groups, while scores

different items assessing their receptive musical abilities.

Table 4.14: Rhythm Discrimination test scores – Study 2

Group	*N*	*M*	*SD*	*Range*
Listen to Speech	20	5.9	1.2	3 - 7
Listen to Singing	20	5.8	1.5	3 - 8
Speaking	20	6.2	1.2	4 - 8
Singing	20	6.0	1.3	3 - 8
Active (listen-and-repeat)	40	6.1	1.2	3 - 8
Passive (listening only)	40	5.8	1.3	3 - 8
Spoken stimuli	40	6.0	1.2	3 - 8
Sung stimuli	40	5.9	1.4	3 - 8
Overall (4 groups)	80	6.0	1.3	3 - 8
Overall: Male	40	6.3	1.2	3 - 8
Overall: Female	40	5.6	1.3	3 - 8

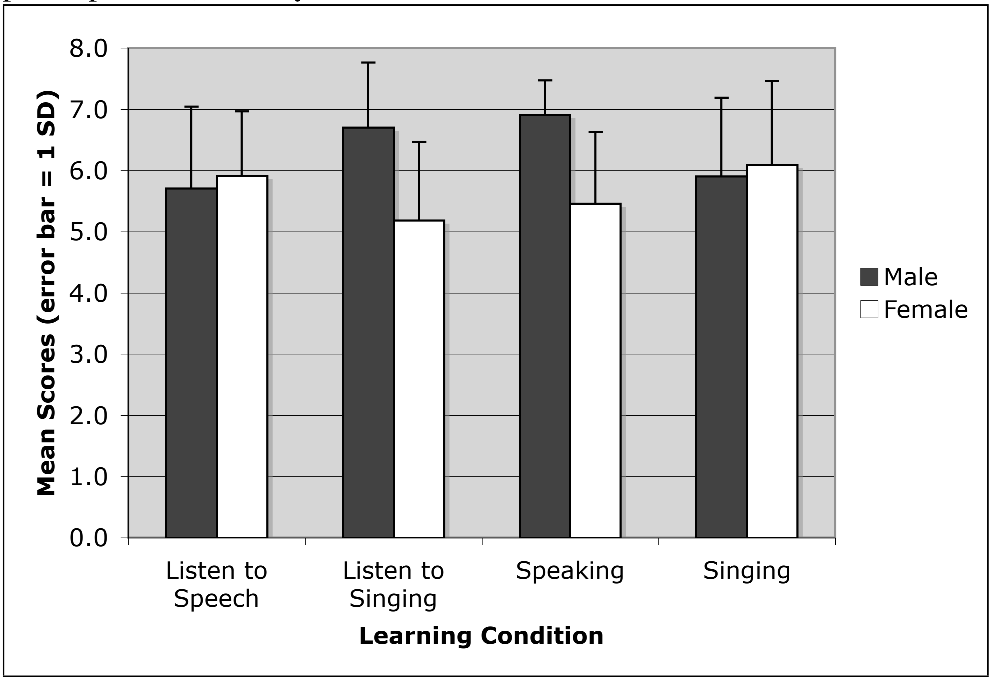
were approximately equal for women and men in the Listen to Speech and the Singing groups.

Pitch Discrimination. Mean scores on the Pitch Discrimination pre-test were similar in the different groups, as shown in Table 4.15. The Listen to Singing condition had the highest mean score (M = 3.4) and the Singing group had the lowest mean score (M = 2.7) and the largest range (from 0 to 4 points out of 4 possible) on the Pitch Discrimination test.

Table 4.15: Pitch Discrimination test scores – Study 2

Group	*N*	*M*	*SD*	*Range*
Listen to Speech	20	3.1	0.9	2 - 4
Listen to Singing	20	3.4	0.7	2 - 4
Speaking	20	3.0	0.6	2 - 4
Singing	20	2.7	1.2	0 - 4
Active (listen-and-repeat)	40	2.9	0.9	0 - 4
Passive (listening only)	40	3.2	0.8	2 - 4
Spoken stimuli	40	3.0	0.8	2 - 4
Sung stimuli	40	3.0	1.0	0 - 4
Overall (4 groups)	80	3.0	0.9	0 - 4
Overall: Male	40	3.1	1.0	0 - 4
Overall: Female	40	3.0	0.8	1 - 4

Figure 4.9: Mean scores on Rhythm Discrimination test for Condition and Gender (8 points possible) – Study 2



Although the two passive groups performed at a higher level than the two active groups on the Pitch Discrimination test, two-way ANOVA showed no significant differences for type of learning condition (p = .084) or for type of stimulus (p = .99), nor was the interaction significant. In addition, no gender differences or interactions were observed on the Pitch Discrimination test, with scores approximately equal for both genders in the different groups.

Melody Discrimination. Performance on the Melody Discrimination test showed very high and very similar mean scores in the different groups, as shown in Table 4.16. The Listen to Speech and Speaking conditions showed the highest performance (M = 7.3), while the listen-and-repeat Singing condition had the lowest mean score on the Melody Discrimination test (M = 7.1).

Two-way ANOVA showed no group differences or interaction for type of stimulus or for type of learning. No main effect for gender was observed (p = .78), but a significant three-way interaction was found between type of stimulus, type of

Table 4.16: Melody Discrimination test scores – Study 2

Group	*N*	*M*	*SD*	*Range*
Listen to Speech	20	7.3	0.8	6 - 8
Listen to Singing	20	7.2	0.9	5 - 8
Speaking	20	7.3	0.8	5 - 8
Singing	20	7.1	0.8	5 - 8
Active (listen-and-repeat)	40	7.2	0.8	5 - 8
Passive (listening only)	40	7.2	0.8	5 - 8
Spoken stimuli	40	7.3	0.8	5 - 8
Sung stimuli	40	7.2	0.8	5 - 8
Overall (4 groups)	80	7.2	0.8	5 - 8
Overall: Male	40	7.2	0.8	5 - 8
Overall: Female	40	7.2	0.8	5 - 8

learning condition, and gender (p = .028). Figure 4.10 is a graph of the mean scores on the Melody Discrimination test in the four learning conditions. Men in the Listen to Singing condition had higher performance than women on the Melody Discrimination test, while women in the Listen to Speech condition had higher performance than the men. In the two active learning conditions (Speaking and Singing), performance on the Melody Discrimination test was approximately equal for men and women.

Total Receptive Musical Ability. A total score was calculated for each participant on the receptive musical ability tests by adding together the participant's scores on the Rhythm, Pitch, and Melody Discrimination tests. Overall, the Singing group had the lowest receptive musical ability score (M = 15.8) and the Listen to Singing and Speaking groups had the highest score (M = 16.4).

Two-way ANOVA showed that overall receptive musical performance scores in the different groups were approximately equal and there was no interaction for type of stimulus (spoken or sung phrases) or type of learning condition (active or passive learning). Table 4.17 shows the descriptive statistics for the different groups.

A marginal effect for gender was found for the receptive musical ability test total

Figure 4.10: Mean scores on Melody Discrimination test for Condition and Gender (8 points possible) – Study 2

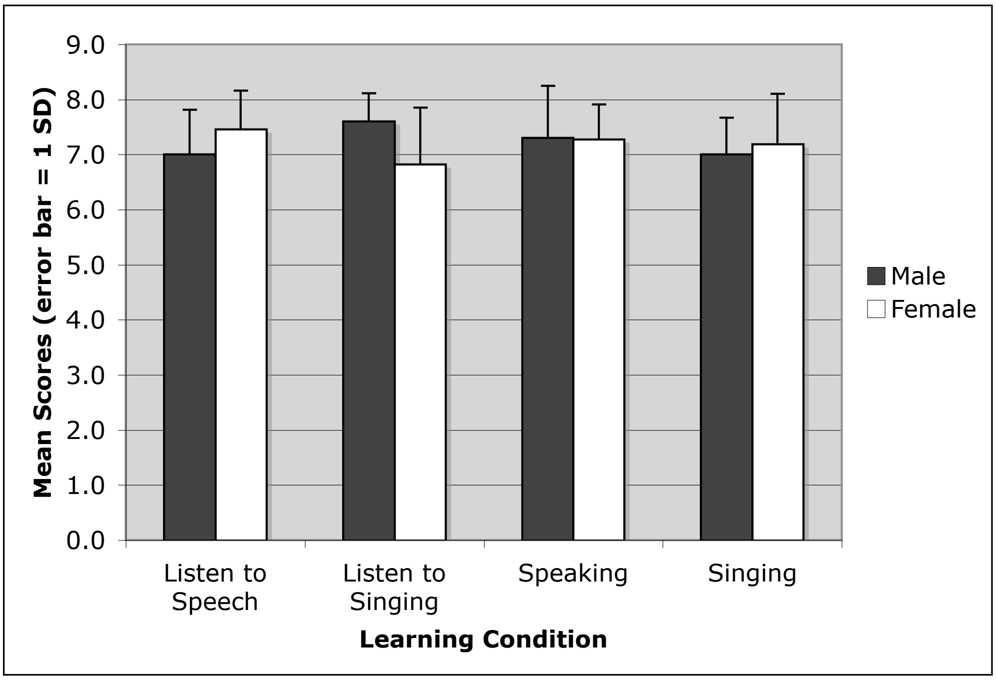


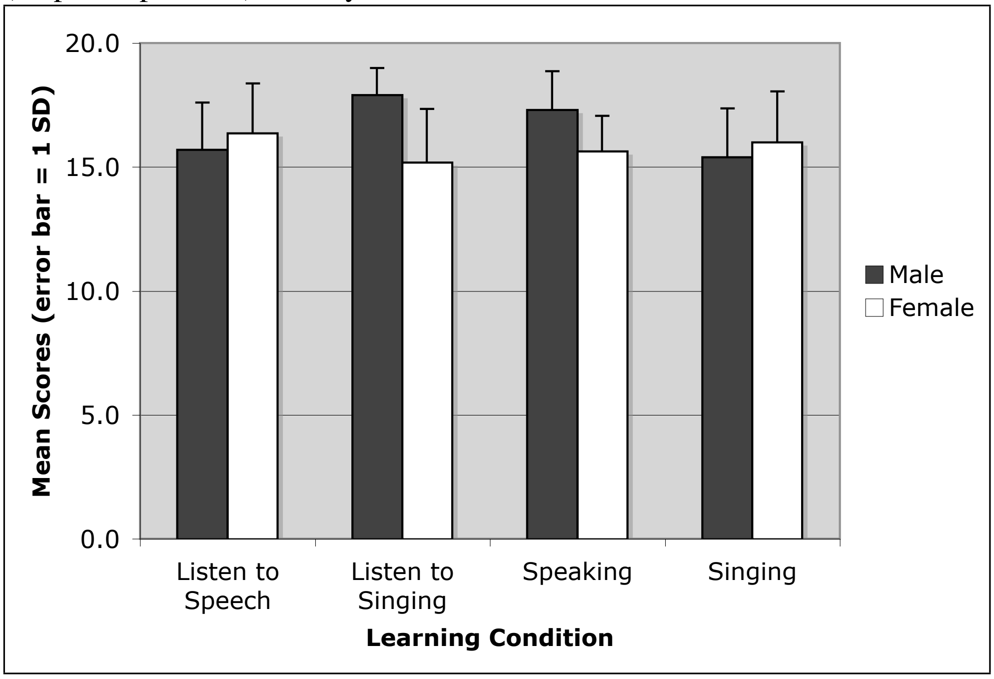
Table 4.17: Total Receptive Musical Ability Test scores – Study 2

Group	*N*	*M*	*SD*	*Range*
Listen to Speech	20	16.2	2.0	13 - 19
Listen to Singing	20	16.4	2.3	11 - 19
Speaking	20	16.4	1.7	13 - 19
Singing	20	15.8	2.0	13 - 20
Active (listen-and-repeat)	40	16.1	1.9	13 - 20
Passive (listening only)	40	16.3	2.1	11 - 19
Spoken stimuli	40	16.3	1.8	13 - 19
Sung stimuli	40	16.1	2.1	11 - 20
Overall (4 groups)	80	16.2	2.0	11 - 20
Overall: Male	40	16.6	3.7	13 - 20
Overall: Female	40	15.8	3.9	11 - 20

score, $p = .070$, with men showing higher mean scores ($M = 16.6$) than women ($M = 15.8$). Three-way factorial ANOVA showed an even greater gender difference when type of stimulus and type of learning condition were included in the calculation, $p = .051$, and the three-way interaction of gender, type of stimulus, and type of learning condition was significant at the $p < .001$ level, with men in the Listen to Singing and

Speaking conditions showing higher performance than women in those conditions (see Figure 4.11).

Figure 4.11: Mean Receptive Musical Ability Test scores for Condition and Gender (20 points possible) – Study 2



Productive Musical Ability Test. For this test, participants were asked to sing and tap along to each syllable as they heard the song ‘Happy Birthday’ played through headphones. Participants’ production was audio recorded for later scoring for accuracy in terms of singing (out of 5) and tapping (out of 5). Results for the two sub-scores on the Productive MAT are outlined below.

Happy Birthday: Singing. Scores on the ‘Happy Birthday’ singing test ranged between 2 and 5, with the highest performance in the Speaking condition (M = 3.5) and the lowest performance in the Singing condition (M = 2.9). Descriptive statistics for this measure are shown in Table 4.18.

Two-way ANOVA showed a significant difference on the singing sub-test of the Productive Musical Ability Test for type of stimulus, p = .042, with the spoken stimuli conditions showing higher performance; a marginal interaction between type

Table 4.18: Happy Birthday: Singing test scores – Study 2

Group	*N*	*M*	*SD*	*Range*
Listen to Speech	20	3.4	0.6	2 - 4
Listen to Singing	20	3.4	0.6	2 - 4
Speaking	20	3.5	0.7	2 - 5
Singing	20	2.9	0.5	2 - 4
Active (listen-and-repeat)	40	3.2	0.7	2 - 5
Passive (listening only)	40	3.4	0.6	2 - 4
Spoken stimuli	40	3.4	0.6	2 - 5
Sung stimuli	40	3.1	0.7	2 - 4
Overall (4 groups)	80	3.3	0.7	2 - 5
Overall: Male	40	3.2	0.7	2 - 5
Overall: Female	40	3.4	0.4	2 - 4

of learning condition and type of stimulus was also observed (p = .089). No main effect for gender was found (p = .32) nor were any interactions between type of learning condition, type of stimulus, and gender observed on the 'Happy Birthday' singing test, with performance approximately equal for both genders in the different groups.

Happy Birthday: Tapping. For the 'Happy Birthday' tapping test, scores ranged from 1 to 5 (out of 5 possible), with an overall average of 3.4 points. The highest performance was in the Listen to Speech condition (M = 3.7) and the lowest performance was in the Singing and Speaking conditions (M = 3.2). Descriptive statistics for the Productive Musical Ability tapping sub-test in the different groups and overall are shown in Table 4.19.

For the 'Happy Birthday' tapping test, two-way ANOVA showed a main effect for type of learning condition, p = .016, with the two passive learning conditions showing higher performance. There was no interaction between type of stimulus and type of learning condition. In addition, no main effect for gender was observed on this measure (p = .81), and no significant interactions between gender, type of stimulus, and type of learning condition were found. Results showed the widest gender gap

Table 4.19: Happy Birthday: Tapping test scores – Study 2

Group	*N*	*M*	*SD*	*Range*
Listen to Speech	20	3.7	0.8	2 - 5
Listen to Singing	20	3.6	0.9	2 - 5
Speaking	20	3.2	0.9	1 - 5
Singing	20	3.2	0.9	2 - 5
Active (listen-and-repeat)	40	3.2	0.9	1 - 5
Passive (listening only)	40	3.7	0.9	2 - 5
Spoken stimuli	40	3.4	0.9	1 - 5
Sung stimuli	40	3.4	0.9	2 - 5
Overall (4 groups)	80	3.4	0.9	1 - 5
Overall: Male	40	3.4	0.8	2 - 5
Overall: Female	40	3.4	1.0	1 - 5

between high-performing women (M = 3.6) and low-performing men (M = 2.7) in the Singing condition; in the other three conditions, men's scores were slightly higher than women's scores.

Total Productive Musical Ability. An overall Productive Musical Ability Test score was calculated by adding together the Happy Birthday: Singing and Happy Birthday: Tapping scores, for a total of 10 points possible. Results showed that the Singing condition had the lowest overall score on this measure (M = 6.1), while the Listen to Speech condition had the highest score (M = 7.1). Table 4.20 shows the descriptive statistics for the Productive Musical Ability Test scores in the different groups and overall.

Two-way ANOVA revealed a significant main effect for type of learning condition, $p < .01$, with the two passive learning groups showing higher Productive MAT performance. There was no interaction between type of stimulus and type of learning condition, $p = .40$.

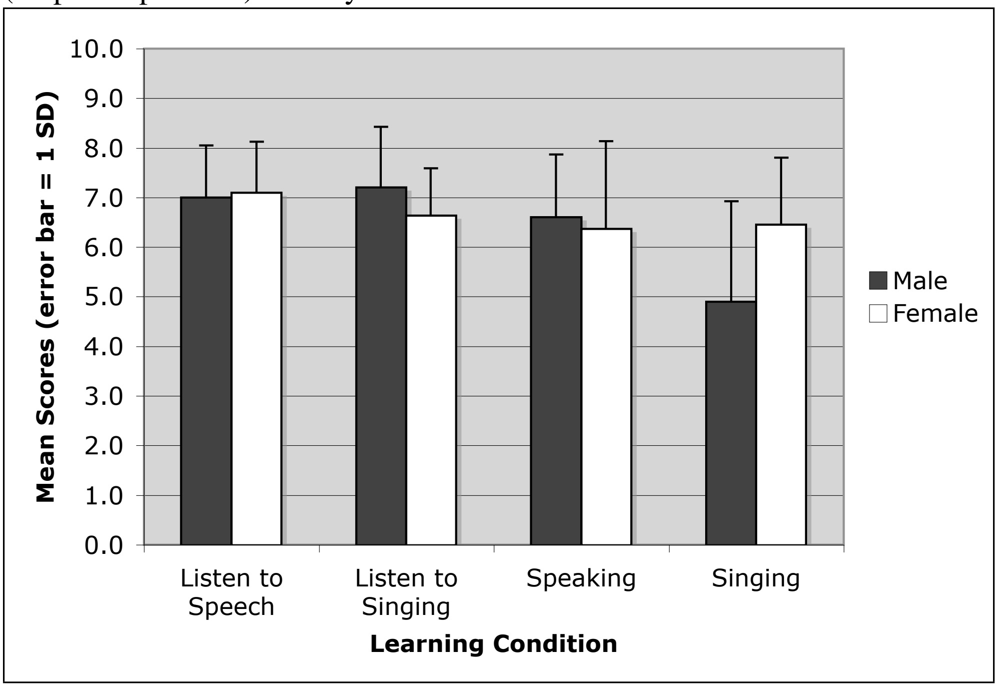
No main effect for gender was observed for the Productive Musical Ability test ($p = .41$). However, a three-way interaction between gender, type of stimulus, and type of learning condition was found ($p = .032$), with women showing higher

Table 4.20: Total Productive Musical Ability Test scores – Study 2

Group	*N*	*M*	*SD*	*Range*
Listen to Speech	20	7.1	1.0	6 - 9
Listen to Singing	20	7.0	1.1	5 - 9
Speaking	20	6.6	1.3	5 - 9
Singing	20	6.1	1.3	4 - 8
Active (listen-and-repeat)	40	6.3	1.3	4 - 9
Passive (listening only)	40	7.0	1.0	5 - 9
Spoken stimuli	40	6.9	1.2	5 - 9
Sung stimuli	40	6.5	2.1	4 - 9
Overall (4 groups)	80	6.7	1.2	4 - 9
Overall: Male	40	6.6	1.3	4 - 9
Overall: Female	40	6.8	1.2	5 - 9

performance than men in the Singing and Listen to Speech conditions while men scored slightly higher than women in the Speaking and in the Listen to Singing conditions (see Figure 4.12).

Figure 4.12: Mean Productive Musical Ability Test scores for Condition and Gender (10 points possible) – Study 2



Overall Musical Ability Test. A total MAT score was calculated for each participant by adding together the musical ability sub-test scores (a total between 0 and 30 points were possible). Table 4.21 shows the descriptive statistics in the different groups. Mean scores were similar overall on the Musical Ability test for type of learning condition and type of stimulus. The Singing condition had the lowest overall Musical Ability score (*M* = 21.6) while the highest performance was found in the Listen to Speech and Listen to Singing groups (*M* = 23.3).

Table 4.21: Overall Musical Ability Test scores – Study 2

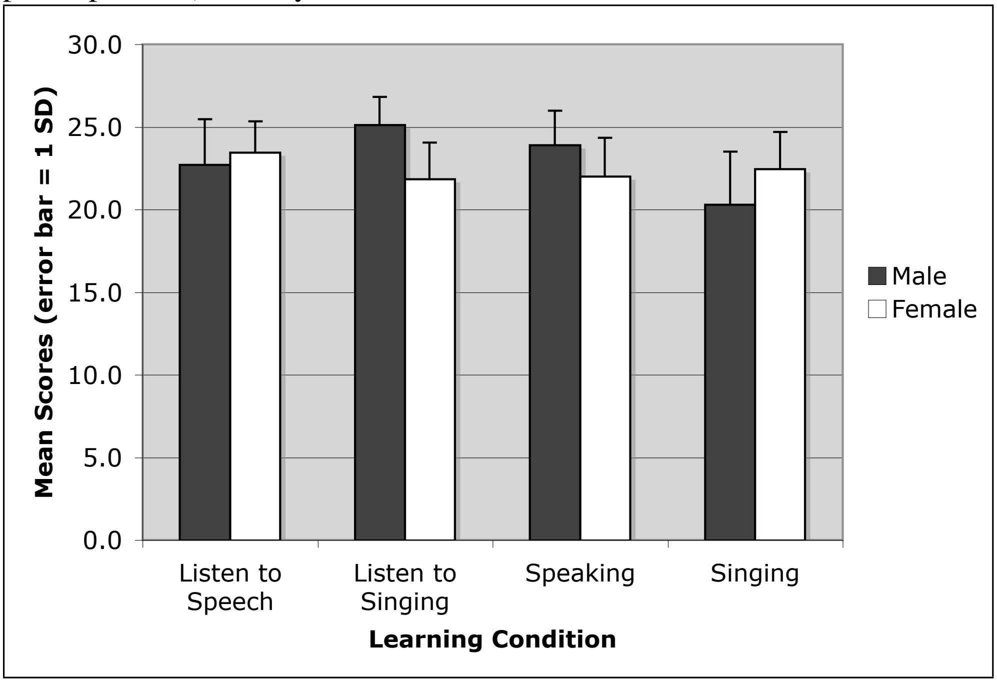
Group	*N*	*M*	*SD*	*Range*
Listen to Speech	20	23.3	2.4	19 - 27
Listen to Singing	20	23.3	2.7	18 - 27
Speaking	20	22.9	2.4	18 - 26
Singing	20	21.6	3.0	16 - 27
Active (listen-and-repeat)	40	22.2	2.8	16 - 27
Passive (listening only)	40	23.3	2.5	18 - 27
Spoken stimuli	40	23.1	2.4	18 - 27
Sung stimuli	40	22.4	2.9	16 - 27
Overall (4 groups)	80	22.7	2.7	16 - 27
Overall: Male	40	23.0	3.0	16 - 27
Overall: Female	40	22.5	2.3	18 - 27

Two-way ANOVA showed a marginal difference for type of learning condition (active or passive learning), $p = .07$, with the active learning groups showing lower performance; no difference was found for type of stimulus ($p = .29$), nor was there an interaction between type of stimulus and type of learning condition ($p = .25$).

Men had slightly higher scores than women overall on the MAT, but there was no main effect for gender, $p = .38$. However, a highly significant three-way interaction was observed between gender, type of learning condition, and type of stimulus (at the $p < .001$ level). Figure 4.13 illustrates the mean scores for overall Musical Ability in the four groups and shows that women's performance in the Listen to Speech and in the Singing groups was higher than men's scores, whereas the opposite pattern was found in the Listen to Singing and in the Speaking groups.

To summarise the results of the musical ability and experience measures, which

Figure 4.13: Mean overall Musical Ability Test scores for Condition and Gender (30 points possible) – Study 2



were administered to establish whether the groups were well matched for musical skills, unfortunately several group differences were observed. In general, the lowest scores on the measures of musical ability and experience were found in the Singing group, while the Listen to Speech group tended to have the highest scores. Two-way ANOVA showed a significant difference for type of stimulus (with the two spoken stimuli groups showing higher scores) on the Happy Birthday: Singing test and for type of learning condition (with the two passive groups showing higher performance) on the Happy Birthday: Tapping test, both at the $p < .05$ level. There was also an effect for type of learning condition on the Productive Musical Ability test, $p = .01$, with the two passive groups showing higher performance, and a marginal effect for the overall Musical Ability Test score ($p = .07$, with the two passive groups scoring higher).

In terms of gender differences, men tended to have higher scores than women on the musical ability and experience measures; however, only one significant main effect for gender was found, on the Rhythm Discrimination test at the $p < .05$ level.

Also, significant interactions between gender and learning condition were found, with men in the Listen to Singing and Speaking conditions showing higher performance than women, whereas women performed at a slightly higher level than men in the Singing and Listen to Speech conditions. In addition, significant three-way interactions were found between type of learning condition, type of stimulus, and gender for the Rhythm Discrimination test, Melody Discrimination test, the Receptive Musical Ability test score, and the overall Musical Ability Test score.

The results of the musical ability tests indicate that participants in the different groups were not very well matched for musical skills. However, the overall pattern of participants' pre-existing musical skills is complex, and the effects of musical ability on Hungarian test scores are explored using MANCOVA in section 4.5. It is hoped that the group differences in musical ability observed in this study will not prove problematic for the generalisability of results and the exploration of the effects of the IDs on the Hungarian test scores.

4.4.3 PANAS mood questionnaire

Participant completed the 20-item self-report Positive and Negative Affect Scale (PANAS) mood questionnaire (Watson, Clark, & Tellegen, 1988) as a pre/post-test. Separate scores were calculated for overall positive and negative affect at the beginning and at the end of the experiment session; descriptive statistics on the PANAS mood questionnaires are shown in Tables 4.22 and 4.23.

Positive Affect. At the start of the experiment session, the Listen to Speech condition reported the highest positive mood scores on the PANAS (M = 20.1), while the Speaking and Singing conditions reported the least positive mood (M = 17.6). The two sung stimuli conditions also had a wider range of scores for positive mood at pre-test, with a 35-point range compared to a 19-point range in the two spoken stimuli conditions. At the end of the experiment session, positive mood scores had decreased

Table 4.22: PANAS positive mood scores – Study 2

Group	**Measure**	*N*	*M*	*SD*	*Range*
Listen to Speech	Initial Positive Affect	20	21.2	5.7	10 - 29
Listen to Singing		20	19.0	7.1	8 - 40
Speaking		20	17.6	4.6	10 - 27
Singing		20	17.6	7.0	5 - 31
Active (listen-and-repeat)		40	17.6	5.8	5 - 31
Passive (listening only)		40	20.1	6.5	8 - 40
Spoken stimuli		40	19.4	5.4	10 - 29
Sung stimuli		40	18.3	7.0	5 - 40
Overall (4 groups)		80	18.9	6.2	5 - 40
Overall: Male		40	20.1	6.5	9 - 40
Overall: Female		40	17.6	5.8	5 - 29
Listen to Speech	Final Positive Affect	20	15.9	6.2	4 - 29
Listen to Singing		20	16.7	7.4	4 - 34
Speaking		20	14.8	6.8	2 - 25
Singing		20	13.8	8.4	2 - 32
Active (listen-and-repeat)		40	14.3	7.6	2 - 32
Passive (listening only)		40	16.3	6.7	4 - 34
Spoken stimuli		40	15.3	6.4	2 - 29
Sung stimuli		40	15.2	8.0	2 - 34
Overall (4 groups)		80	15.3	7.2	2 - 34
Overall: Male		40	16.7	6.8	2 - 34
Overall: Female		40	13.8	7.4	2 - 32

in all groups, by a mean of 5.3 points in the Listen to Speech condition, while positive mood decreased the least in the Listen to Singing condition (by only 2.3 points).

Two-way repeated-measures ANOVA showed a significant overall decrease in Positive Affect scores from pre-test to post-test, $p < .001$, but no significant group differences or interactions were found for the change in Positive Affect scores.

In addition, two-way ANOVA showed a marginal difference for Positive Affect at pre-test ($p = .075$) for type of learning condition, with the two active (listen-and-repeat) groups tending to report lower Positive Affect scores than the passive (listen only) groups. There was no interaction between type of stimulus and type of learning condition, p = .43. For Positive Affect at post-test, two-way ANOVA showed the marginal difference at pre-test for type of learning condition was no longer evident (p = .23). No main effect for type of stimulus was observed, nor was there a significant interaction between type of learning condition and type of stimulus for Positive Affect at the end of the experiment sessions.

In this study, men had marginally higher Positive Affect scores on the PANAS than women, both at the start (p = .079) and at the end of the session (p = .074); there was no interaction between gender and change in Positive Affect (p = .73). Three-factor repeated-measures ANOVA exploring the change in Positive Affect from pre-test to post-test showed no main effects or interactions between gender, type of stimulus, and type of learning condition.

Negative Affect. At the start of the experiment session, the lowest mean Negative Affect score was found in the Listen to Singing condition (M = 3.7) and the Speaking condition had the highest mean Negative Affect score (M = 4.9). At post-test, the Negative Affect scores were even more similar between the four groups, with a mean difference of only 0.3 points (see Table 4.23).

From pre-test to post-test, Negative Affect scores decreased slightly in the two active learning conditions (Speaking and Singing), while Negative Affect scores increased slightly in the two passive learning conditions (Listen to Speech and Listen to Singing). However, two-way repeated-measures ANOVA did not show a large change in Negative Affect scores from pre-test to post-test (p = .91), and no significant group differences or interactions between type of learning condition and type of stimulus were observed on this measure.

Two-way ANOVA did not show any group differences or interactions for Negative Affect at pre-test for type of learning condition and type of stimulus. At the end of the experiment session, two-way ANOVA also showed no differences or

Table 4.23: PANAS negative mood scores – Study 2

Group	**Measure**	*N*	*M*	*SD*	*Range*
Listen to Speech	Initial Negative Affect	20	4.1	3.3	0 - 10
Listen to Singing		20	3.7	3.6	0 - 12
Speaking		20	4.9	3.5	1 - 16
Singing		20	4.7	4.2	0 - 15
Active (listen-and-repeat)		40	4.8	3.8	0 - 16
Passive (listening only)		40	3.9	3.4	0 - 12
Spoken stimuli		40	4.5	3.4	0 - 16
Sung stimuli		40	4.2	3.9	0 - 15
Overall (4 groups)		80	4.3	3.6	0 - 16
Overall: Male		40	4.5	3.9	0 - 16
Overall: Female		40	4.2	3.3	0 - 12

Listen to Speech	Final Negative Affect	20	4.2	5.0	0 - 16
Listen to Singing		20	4.2	5.3	0 - 18
Speaking		20	4.5	4.0	0 - 18
Singing		20	4.3	4.6	0 - 14
Active (listen-and-repeat)		40	4.4	4.2	0 - 18
Passive (listening only)		40	4.2	5.1	0 - 18
Spoken stimuli		40	4.3	4.5	0 - 18
Sung stimuli		40	4.3	4.9	0 - 18
Overall (4 groups)		80	4.3	4.6	0 - 18
Overall: Male		40	4.4	5.2	0 - 18
Overall: Female		40	4.2	4.1	0 - 16

interaction between type of learning condition and type of stimulus for the Negative Affect post-test measure.

On the PANAS, men had slightly more negative mood scores than women; however, the gender difference for pre-test, post-test, and change in Negative Affect scores was not significant. No significant interactions were found between gender, type of stimulus, and type of learning condition for Negative Affect, either.

Positive and Negative Affect. Repeated-measures factorial MANOVA showed no interactions between the change in Positive and Negative Affect scores for type of stimulus or for gender. However, a marginal effect was found for type of learning condition, $p = .069$, for change in Positive Affect, with scores in the two passive groups decreasing more than scores in the two active groups; with gender also included, the difference was similar, $p = .071$. There was also a highly significant interaction between change in Positive Affect and change in Negative Affect scores, $p < .001$, although this was unsurprising.

In summary, a marginal difference for Positive Affect at pre-test ($p = .075$) was found for type of learning condition, with the active groups reporting lower positive mood scores than the passive groups. In addition, men had higher Positive Affect scores than women at both pre-test and post-test, although the differences were marginal. These pre-existing differences between the groups are somewhat problematic because mood has been shown to affect learning and memory (Hulse et al., 2007). In section 4.5, PANAS mood scores at the start of the session are controlled for, to establish whether the marginal differences observed for positive mood scores (for type of learning condition and for gender) might have had a differential impact on Hungarian test scores.

4.4.4 Phonological working memory

Performance on the 20 low-wordlike items from the Children's Test of Nonword Repetition (Archibald & Gathercole, 2006, p. 514) showed that the two active groups had lower mean scores ($M = 17.0$) on the Phonological Working Memory measure than the two passive groups ($M = 17.8$). The Listen to Speech condition had the highest mean scores ($M = 17.9$) and the Speaking condition had the lowest mean scores ($M = 16.8$) on the Phonological Working Memory test. Table 4.24 shows the descriptive statistics for this measure.

Two-way ANOVA showed a significant difference for type of learning condition on the Phonological Working Memory test, $p = .022$, with the two passive learning

Table 4.24: Phonological Working Memory test scores – Study 2

Group	*N*	*M*	*SD*	*Range*
Active (listen-and-repeat)	40	17.0	1.7	13 - 20
Passive (listening only)	40	17.8	1.4	14 - 20
Spoken stimuli	40	17.3	1.8	13 - 20
Sung stimuli	40	17.4	1.4	14 - 20
Listen to Speech	20	17.9	1.5	14 - 20
Listen to Singing	20	17.7	1.3	14 - 20
Speaking	20	16.8	2.0	13 - 20
Singing	20	17.2	1.4	14 - 19
Overall (4 groups)	80	17.4	1.6	13 - 20
Overall: Male	40	17.3	1.6	14 - 20
Overall: Female	40	17.5	1.6	13 - 20

groups showing higher performance than the two active learning groups. There was no effect for type of stimulus, nor was there an interaction between type of learning condition and type of stimulus.

In terms of gender differences, women tended to score higher than men, except in the Speaking condition where women performed slightly worse (M = 16.5) than men (M = 17.0). ANOVA showed that scores on the Phonological Working Memory measure were similar for both genders, p = .54, and factorial ANOVA did not show any significant interactions between gender, type of stimulus, and type of learning condition.

The group difference observed for type of learning condition (with the active groups performing worse) on the Phonological Working Memory score was not ideal because phonological working memory correlates with language learning ability and verbal IQ (Alloway & Alloway, 2010; Gathercole, 2006; Gilleece, 2006). This pre-existing difference between the active and passive groups could have influenced participants' Hungarian test performance, so this factor will also need to be controlled for in section 4.5.

4.4.5 Handedness

Altogether, eight left-handed participants took part in this study, with one in the Listen to Speech condition, three in the Listen to Singing condition, two in the Speaking condition, and two in the Singing condition. More details about the incidence of left-handed individuals are shown in Table 4.25. Since only a small number and proportion of participants in this study were left-handed, the following statistical analyses should be not considered strong evidence regarding the effects of handedness on learning new language phrases through different auditory methods.

Table 4.25: Left-handed participants – Study 2

Group	Total N	N Left-handers	% Left-handers
Listen to Speech	20	1 (1 M)	5%
Listen to Singing	20	3 (2 M / 1 F)	15%
Speaking	20	2 (1 M / 1 F)	10%
Singing	20	2 (1 M / 1 F)	10%
Active (listen-and-repeat)	40	4 (2 M / 2 F)	10%
Passive (listening only)	40	4 (3 M / 1 F)	10%
Spoken stimuli	40	3 (2 M / 1 F)	7.5%
Sung stimuli	40	5 (3 M / 2 F)	12.5%
Overall (4 groups)	80	8 (5 M / 3 F)	10%
Overall: Male	40	5	12.5%
Overall: Female	40	3	7.5%

For handedness, overall differences in Hungarian test scores were in the expected direction, with right-handers outperforming left-handed participants. One-way MANOVA for the four Hungarian tests showed an overall trend for handedness, $p = .06$, with post-hoc tests showing that right-handed individuals performed at a significantly higher level than left-handers on the Hungarian Production Test ($p = .02$) and on the English Recall Test ($p = .01$), and a trend on the Delayed-Recall Hungarian Conversation ($p = .05$). No main effect for handedness was found on the Hungarian Recognition Test ($p = .40$).

When comparing Hungarian test performance including handedness along with type of learning condition and type of stimulus, overall 2x3 MANOVA results showed a marginal main effect for handedness, $p = .07$, and for type of stimulus, $p = .05$, but no significant group difference for type of learning condition. There was a marginal three-way interaction between type of stimulus, type of learning condition, and handedness on the four Hungarian tests, $p = .08$.

No interactions between gender and handedness were observed for any of the four Hungarian tests. With handedness, gender, type of stimulus and type of learning condition included, MANOVA results showed a main effect for type of stimulus ($p = .02$) and a significant interaction between type of stimulus and handedness ($p = .04$); post-hoc tests revealed that the group difference and interaction were found only for the Hungarian Recognition Test, with the spoken stimuli groups showing higher performance. For the interaction, left-handers in the sung stimuli groups had lower performance than right-handers on the Hungarian Recognition Test, whereas left-handed participants in the spoken stimuli groups had higher performance than right-handers.

Again, it is important to remember that a separate study would be needed to properly investigate the effects of handedness (and any interactions with gender and other factors) when learning L2 material using this auditory learning procedure.

4.4.6 Nonverbal reasoning

Two nonverbal reasoning tests were administered for the pilot study, the Paper Folding and Cutting test and the nonverbal reasoning sub-test of the Dyslexia Adult Screening Test. The Paper Folding and Cutting test was removed from the main study design, as explained in more detail below.

Nonverbal reasoning: Paper Folding and Cutting test. This multiple-choice spatial IQ sub-test of the Stanford-Binet Intelligence Scale (Thorndike et al., 1986) was administered as a pre/post-test measure for the pilot study only, with a total of eight questions (four items as a pre-test and four items as a post-test, with two simple and two complex in each) and participants received one point for each correct answer. This Stanford-Binet sub-test was removed from the final study design due to time constraints in conjunction with the measure's lack of relevance to the study's main research questions.[24]

Nonverbal reasoning: DAST. The 8-item Nonverbal Reasoning sub-test of the Dyslexia Adult Screening Test created by Fawcett and Nicolson (1998) was administered to all participants at the start of the experiment sessions. The first three items required the participant to choose the shape that would follow next in the pattern, the second three items assessed the participant's ability to identify analogies (e.g., A is to B as C is to which of these shapes?), and the final two items required participants to identify the natural pair and the group of three from a group of five shapes. Participants received one point for each correct answer, with a total of 8 possible.

Results showed that performance on the Nonverbal Reasoning test was approximately equal in the different groups, with an overall mean of 6.0. The Singing condition had the highest mean score (M = 6.3) while the Listen to Singing condition had the lowest mean score (M = 5.7). The descriptive statistics in the different groups on the Nonverbal Reasoning test are available in Table 4.26. Two-way ANOVA showed no significant group differences for the Nonverbal Reasoning measure for type of stimulus or type of learning condition, nor for the interaction.

Although men outperformed women on the Nonverbal Reasoning test (apart from the women in the Singing condition, who performed slightly better than the men), no gender difference was observed overall. ANOVA showed no differences or

[24]More details about the pilot study results on the Paper Folding and Cutting pre-test and post-test were described in section 4.2.2 on page 100.

Table 4.26: Nonverbal Reasoning test scores – Study 2

Group	*N*	*M*	*SD*	*Range*
Listen to Speech	20	6.0	1.6	3 - 8
Listen to Singing	20	5.7	1.1	4 - 8
Speaking	20	5.9	1.3	3 - 8
Singing	20	6.3	1.3	3 - 8
Active (listen-and-repeat)	40	6.1	1.3	3 - 8
Passive (listening only)	40	5.9	1.4	3 - 8
Spoken stimuli	40	6.0	1.4	3 - 8
Sung stimuli	40	6.0	1.2	3 - 8
Overall (4 groups)	80	6.0	1.3	3 - 8
Overall: Male	40	6.1	1.3	3 - 8
Overall: Female	40	5.8	1.4	3 - 8

interaction between gender, type of stimulus and type of learning condition on this nonverbal reasoning test.

4.4.7 Dyslexia

Similar to the last experiment, the online advertisement for this study requested volunteers who were native English-speaking participants, without any learning difficulties, and who were under the age of 30. Participants with undiagnosed dyslexia, unevenly balanced between the different groups, could influence the Hungarian test results.

The assessment of possible dyslexia was made by screening for participants whose performance on at least two of three ID measures had: (1) a score below 15 (out of 20) on the Phonological Working Memory test (CNRep); (2) a score of 4 or less (at chance or below) on the musical ability test's Rhythm Discrimination test; and (3) a score less than the mean (3 out of 5) on the Happy Birthday: Tapping sub-test. These particular tests were chosen because dyslexic individuals often have greater difficulty with Phonological Working Memory and rhythm tasks than individuals without learning difficulties (Overy et al., 2003). The Nonverbal Reasoning test was also checked to ensure that the scores of the individuals identified as being at risk of dyslexia were not lower than the overall mean, since dyslexic individuals tend to perform at a similar level to non-dyslexic individuals on nonverbal reasoning tasks (Pickering, 2006).

Results identified five individuals who were at risk of dyslexia, and these participants were not evenly distributed between the four groups. One individual was in the Speaking condition, one individual was in the Listen to Singing condition, and three individuals were in the Singing condition, while no participants at risk of dyslexia were found in the Listen to Speech condition. The influence of this undiagnosed learning disability on the different groups' Hungarian test results is explored in more detail in section 4.5.

4.4.8 Debriefing questionnaire

Each participant completed a 9-item debriefing questionnaire at the end of the experiment, after reading a Debriefing sheet that explained the study's research questions in more detail. Participants' responses to the debriefing questions were then coded to explore whether there were any differences between the groups. Three questions were Likert-style items and responses to these questions were assigned a value from 0 to 5 points. The other six questions were open-ended and were coded based on patterns within the responses. The results are described below.

1. Before beginning this study, did you know the meaning of any of the words you heard in this new language? If so, could you say how many (or which) words you already knew?

Most participants (93.75%) reported that they did not previously know any of the words in the new language. However, five participants (6.25%) knew the meaning of a total of seven Hungarian words prior to beginning the study. One participant (female) in the Listen to Speech condition knew three words: 'yes,' 'no,' and '(I) thank you' (*igen*, *nem*, *köszönöm*); one participant (male) in the Listen to Singing condition knew one practice item, 'hi' (*szia*); two participants (both male) in the Speaking condition knew one word each, 'cost' and 'no/not' (*kerül* and *nem*), and one participant (male) in the Singing condition knew how to say 'good day' (*jó napot*).[25]

Although a few participants knew between 1-3 Hungarian words before starting the experiment, both the number of words and the distribution of individuals who knew those Hungarian words were small and fairly well balanced between the groups. While a participant's knowledge that they were learning Hungarian phrases might have influenced their performance in other ways (for example, the belief that Hungarian is not an easy language to learn might have decreased their motivation), it is likely that this effect was fairly balanced between the different groups.

2. What do you think the experiment was about?

Participants read a short description about the study's research questions on the Debriefing Sheet before answering this question, so the responses did not provide very useful information. The majority of the responses (73.75%) were general statements regarding the aim of the experiment (e.g., 'About how people learn a foreign language and what the most effective way to learn is,' 'To investigate language learning through listening'). Four participants (5.0%) wrote that the study's aim was to investigate the relationship between language learning and other factors such as musical skills, the ability to identify patterns, and so on (e.g., 'Exploring the ways we learn languages, links to music & auditory perception,' 'What outside factors can influence auditory memory of a new language e.g. musical background, mood, personality'). In total, 17 participants (21.25%) wrote a close approximation of what the research question was

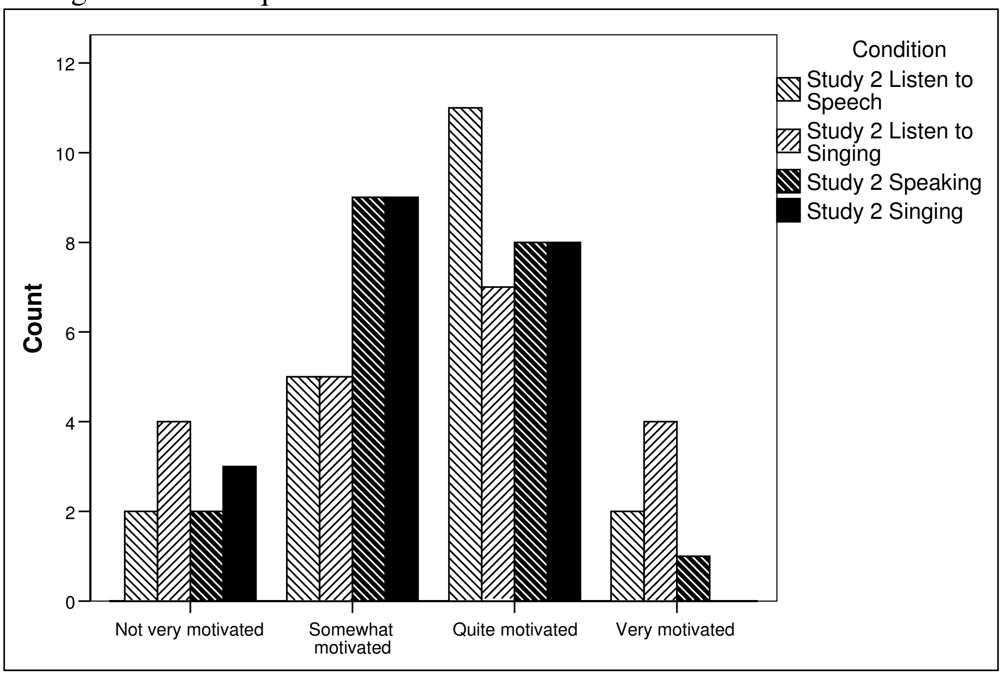
[25] As previously noted, participants in this study heard a polite form, *jó napot kívánok*, 'I wish you a good day,' although this more formal phrase was translated into English in this study as 'good day'.

(e.g., ‘Comparing how melody might aid people’s learning of new languages’).

3. Overall, how motivated did you feel to learn the phrases in this new language? (Please circle one and/or write a few words describing your level of motivation.)

Participants chose from six options on this Likert-style question: Very motivated (5 points), Quite motivated (4 points), Somewhat motivated (3 points), Not very motivated (2 points), Not at all motivated (1 point), and Unsure (0 points). Responses were coded and then scores in the different groups were compared using parametric (ANOVA) and non-parametric (Kruskal-Wallis) statistical tests. Figure 4.14 shows the frequencies for each response in the four groups and Table 4.27 shows the descriptive statistics for participants’ self-reported motivation to learn the phrases. No participants chose ‘Not at all motivated’ or ‘Unsure’ for this question.

Figure 4.14: Frequencies for Overall Motivation to Learn the Phrases – Study 2



Overall scores were similar in the different groups, with the Listen to Speech condition reporting the highest mean motivation (M = 3.7) while the Singing condition had the lowest motivation (M = 3.3) and the smallest range, with no

Table 4.27: Overall Motivation to Learn the Phrases – Study 2

Group	*N*	*Mean*	*Median*	*SD*	*Range*
Listen to Speech	20	3.7	4.0	0.8	2 - 5
Listen to Singing	20	3.6	4.0	1.1	2 - 5
Speaking	20	3.4	3.0	0.8	2 - 5
Singing	20	3.3	3.0	0.7	2 - 4
Active (listen-and-repeat)	40	3.3	3.0	0.7	2 - 5
Passive (listening only)	40	3.6	4.0	0.9	2 - 5
Spoken stimuli	40	3.5	4.0	0.8	2 - 5
Sung stimuli	40	3.4	3.0	0.9	2 - 5
Overall (4 groups)	80	3.5	4.0	0.8	2 - 5
Overall: Male	40	3.4	3.0	0.9	2 - 5
Overall: Female	40	3.6	4.0	0.8	2 - 5

participants in this condition reporting that they felt ‘Very motivated’ to learn the phrases. Although the two active learning conditions had slightly lower motivation scores (M = 3.3) than the two passive learning conditions had (M = 3.6), two-way ANOVA did not show group differences for type of stimulus or for type of learning condition, nor was there a significant interaction between these factors (p = .90). The Kruskal-Wallis test for type of stimulus (spoken or sung phrases) showed p = .50, very similar to the two-way ANOVA value, p = .51. The Kruskal-Wallis test for type of learning condition (active or passive learning) showed p = .12, which was also similar to the two-way ANOVA results, p = .15. Thus, in this study the active learning groups reported less motivation to learn the phrases than the passive learning groups, but the difference was not significant.

No main effect for gender was found for participants’ motivation to learn the phrases (p = .36). Two by three factorial ANOVA including gender, type of stimulus, and type of learning condition showed no main effects, but a marginal interaction between type of stimulus and gender was found, p = .086, with men in the two sung stimuli conditions reporting lower motivation to learn the phrases than women (see Figure 4.15); Figure 4.16 shows that men had higher motivation scores than women in the Listen to Speech condition; otherwise, women had higher scores than men.

Figure 4.15: Mean scores for Motivation to Learn the Phrases for Gender and Type of Stimulus – Study 2

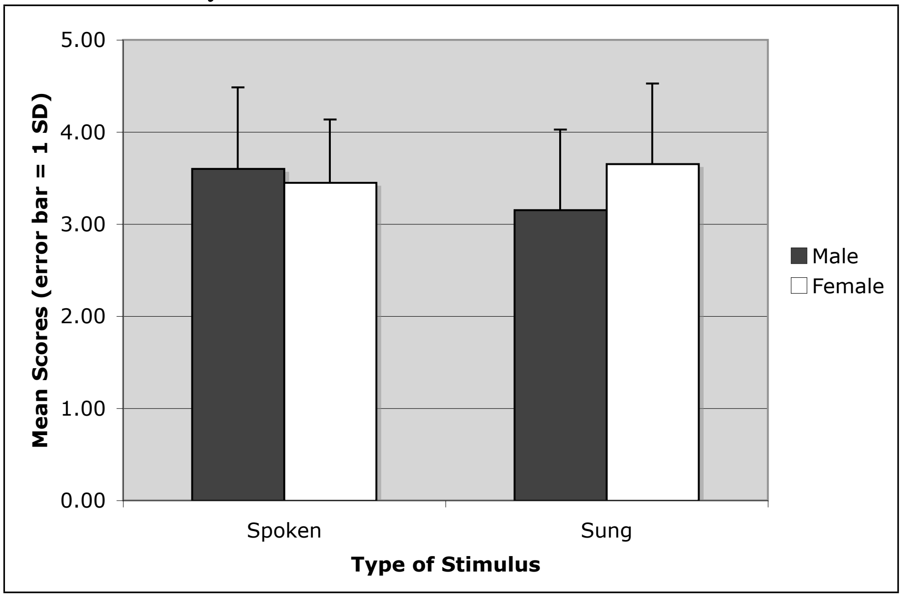
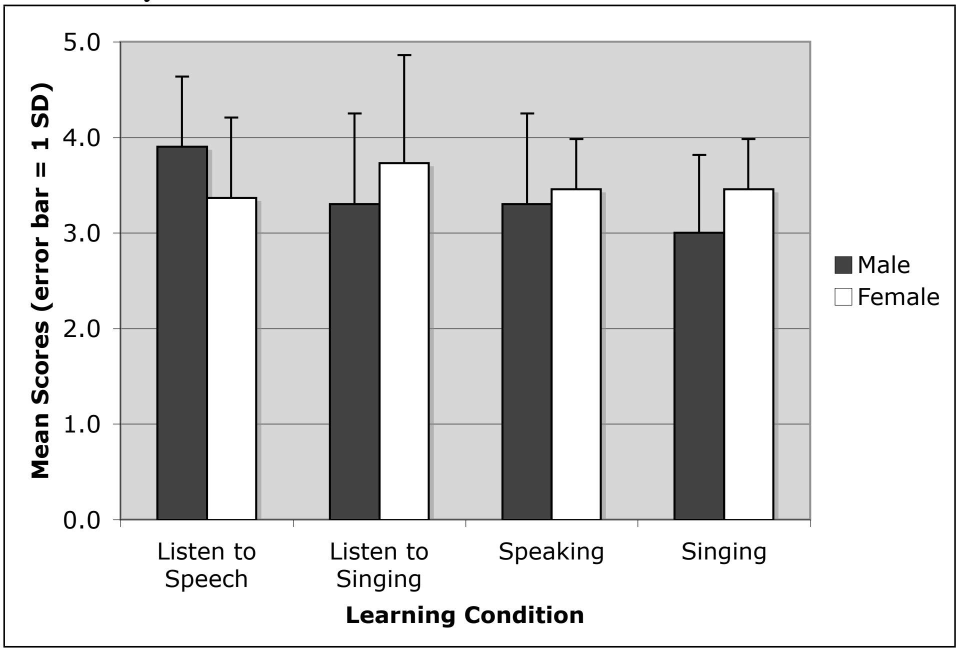


Figure 4.16: Mean scores for Motivation to Learn the Phrases for Gender and Condition – Study 2



In summary, group differences in participants' self-reported overall motivation to learn the phrases were not significant. The Listen to Speech condition reported the highest motivation to learn the phrases and the Singing condition reported the lowest motivation. This pattern is interesting because the Listen to Speech condition showed the highest performance on the Hungarian tests, while the Singing condition tended to perform second-best, and yet scores were much lower than in the Listen to Speech condition (apart from the Delayed-Recall Hungarian Conversation). This pattern of results on the Hungarian tests suggests that motivation to learn the phrases may indeed be an important factor that influenced participants' Hungarian test performance. The fourth question on the Debriefing Questionnaire also concerns motivation.

4. Did you notice any change in your motivation to learn the phrases at the beginning compared to your motivation in the middle or at the end of the learning phase? If so, could you describe the difference?

All participants chose to answer this open-ended question on the Debriefing Questionnaire. On the whole, most responses had a negative quality, with many participants describing a loss of motivation and even feelings of hopelessness when trying to learn so many difficult L2 phrases at once. The open-ended responses were simplified for categorisation and coding as either an increase in motivation (1 point), no change in motivation (0 points), or a decrease in motivation to learn the phrases (-1 point). Figure 4.17 shows the response frequencies in the four conditions and Table 4.28 shows the descriptive statistics for change in motivation score in the different groups. Scores were then compared using both parametric (ANOVA) and non-parametric (Kruskal-Wallis) statistical tests.

Scores for this question were higher in the two active learning groups, and only three participants in the two passive, listening only groups reported an increase in motivation (7.5% compared to 11 participants, or 27.5%, in the two active groups). In addition, scores were lower in the two sung stimuli groups (M = -0.60) compared to the two spoken stimuli groups (M = -0.28). In the four learning conditions, the listen-and-repeat Speaking condition showed the smallest decrease in motivation to learn the phrases (M = -0.10), while motivation decreased most in the Listen to Singing condition (M = -0.65).

Figure 4.17: Frequencies for Change in Motivation to Learn the Phrases – Study 2

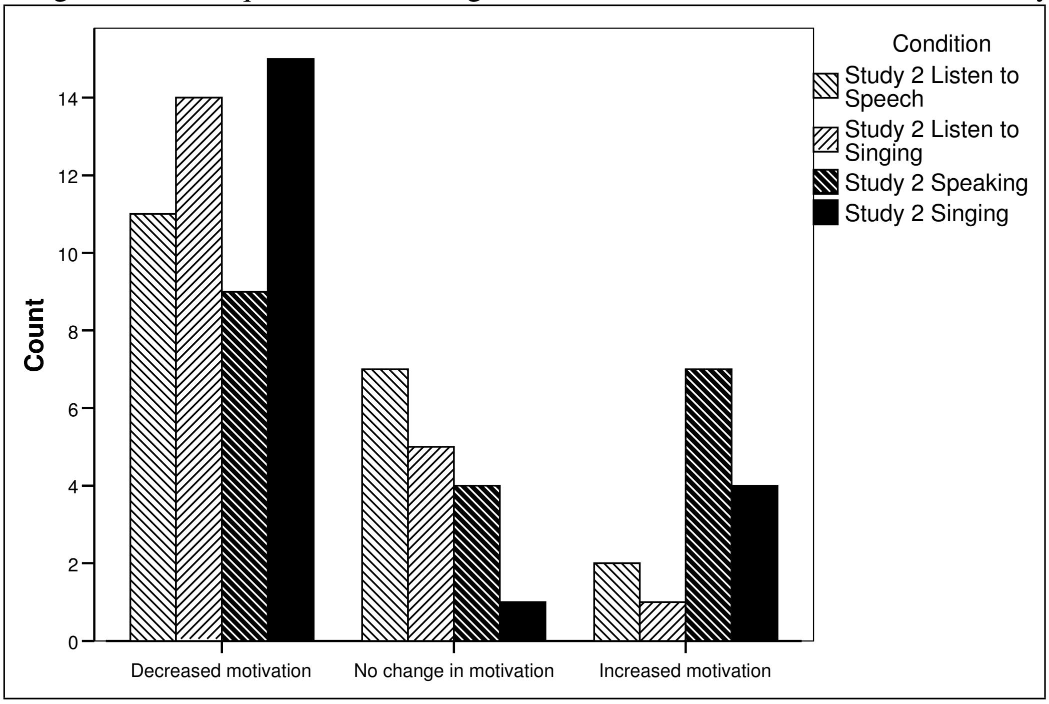


Table 4.28: Change in Motivation to Learn the Phrases – Study 2

Group	*N*	*Mean*	*Median*	*SD*	*Range*
Listen to Speech	20	-0.45	-1.0	0.7	-1 to +1
Listen to Singing	20	-0.65	-1.0	0.6	-1 to +1
Speaking	20	-0.10	0.0	0.9	-1 to +1
Singing	20	-0.55	-1.0	0.8	-1 to +1
Active (listen-and-repeat)	40	-0.33	-1.0	0.9	-1 to +1
Passive (listening only)	40	-0.55	-1.0	0.9	-1 to +1
Spoken stimuli	40	-0.28	-0.50	0.8	-1 to +1
Sung stimuli	40	-0.60	-1.0	0.7	-1 to +1
Overall (4 groups)	80	-0.44	-1.0	0.8	-1 to +1
Overall: Male	40	-0.23	0.0	0.8	-1 to +1
Overall: Female	40	-0.65	-1.0	0.7	-1 to +1

Two-way ANOVA showed no main effect for type of learning condition, $p = .19$. However, a trend was found for type of stimulus, $p = .061$, with the sung stimuli groups reporting a larger decrease in motivation to learn the phrases compared to the spoken stimuli groups. The interaction between type of learning condition and type of stimulus was not significant, $p = .47$. The Kruskal-Wallis test results were somewhat similar to the ANOVA, with no difference for type of learning condition (active or passive learning) on the Kruskal-Wallis test, $p = .39$, while for type of stimulus, the Kruskal-Wallis test was significant, $p = .047$, with participants' motivation to learn the phrases decreasing more in the sung stimuli conditions.

A main effect for gender was found for the change in motivation scores, $p = .012$, with women reporting a larger decrease in motivation than men in this study. Two by three ANOVA comparing gender, type of stimulus, and type of learning condition showed no significant differences or interactions. Later, section 4.5 explores the influence of motivation and any changes in motivation to learn the phrases on participants' Hungarian test scores.

5. How successful do you think you were at learning to say the Hungarian phrases? (Please circle one and/or write a few words describing your overall opinion.)

Participants chose from six options for this Likert-style question and their responses were coded as follows: Very successful (5 points), Quite successful (4 points), Somewhat successful (3 points), Not very successful (2 points), Not at all successful (1 point), and Unsure (0 points). Despite the changes made to the Debriefing Sheet for the final version of this study,[26] most participants (83.75%) still reported on the Debriefing questionnaire that they felt unsuccessful at learning to say the Hungarian phrases. For this study, 33 participants, or 41.25%, felt 'Not at all successful' and 34 participants, or 42.5%, reported that they felt 'Not very successful' at learning to say the Hungarian phrases. Figure 4.18 shows the response frequencies in the four conditions for this question and Table 4.29 shows the descriptive statistics in the different groups. The listen-and-repeat Singing condition had the lowest scores on this question, $M = 1.55$, and the listen-and-repeat Speaking condition had the highest mean scores, $M = 1.95$. Scores in the different groups were compared using both parametric (ANOVA) and non-parametric (Kruskal-Wallis) tests.

[26]Based on the pilot study results, the description on the Debriefing Sheet made it clear that this auditory memory task was designed to be very difficult, with an average score of only 20%.

Figure 4.18: Frequencies for Success at Learning the Hungarian Phrases – Study 2

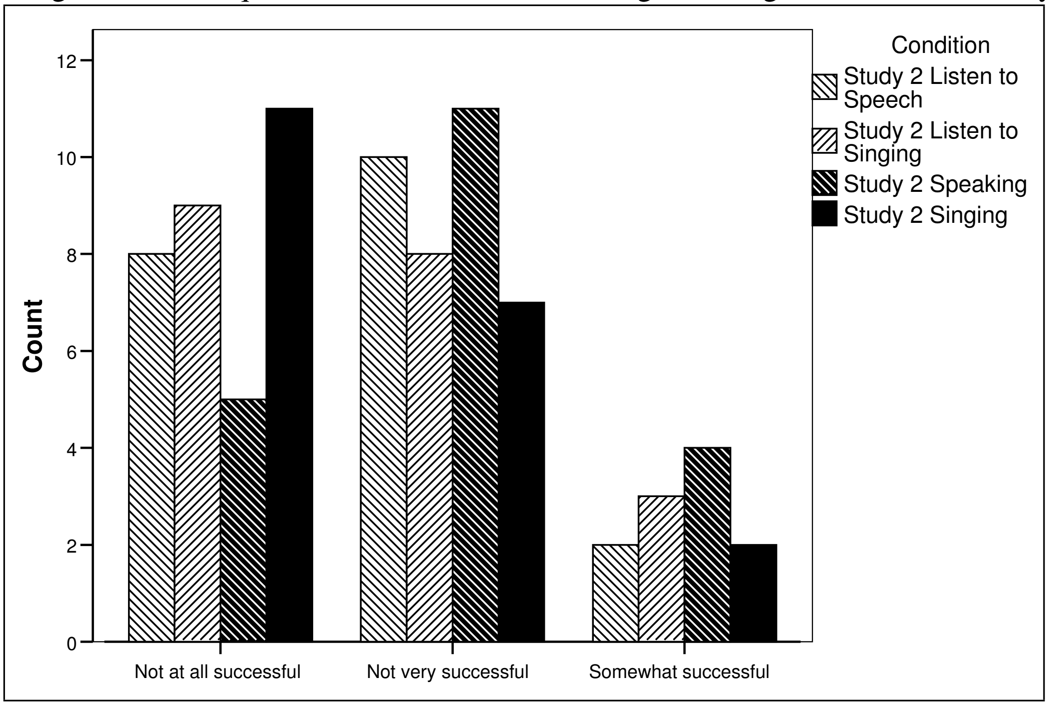


Table 4.29: Success at Learning the Hungarian Phrases – Study 2

Group	*N*	*Mean*	*Median*	*SD*	*Range*
Listen to Speech	20	1.7	2.0	0.7	1 - 3
Listen to Singing	20	1.7	2.0	0.7	1 - 3
Speaking	20	2.0	2.0	0.7	1 - 3
Singing	20	1.6	1.0	0.7	1 - 3
Active (listen-and-repeat)	40	1.8	2.0	0.7	1 - 3
Passive (listening only)	40	1.7	2.0	0.7	1 - 3
Spoken stimuli	40	1.8	2.0	0.7	1 - 3
Sung stimuli	40	1.6	1.5	0.7	1 - 3
Overall (4 groups)	80	1.7	2.0	0.7	1 - 3
Overall: Male	40	1.8	2.0	0.7	1 - 3
Overall: Female	40	1.7	2.0	0.7	1 - 3

Two-way ANOVA comparing the effects of type of stimulus and type of learning condition showed no significant differences or interactions for self-perceived success at learning to say the Hungarian phrases, nor was the interaction significant (p = .20). The Kruskal-Wallis test showed similar results for type of stimulus (p = .16 compared to p = .20 for ANOVA) and for type of learning condition (p = .76 compared to p = .75 for ANOVA).

No main effect for gender was observed on this question, p = .34. Two by three factorial ANOVA showed no significant group differences or interactions between type of learning condition, type of stimulus, and gender.

At the end of this experiment, participants' self-perceived success at learning to say the Hungarian phrases did not differ significantly in the different groups. This result is helpful because it suggests that participants' beliefs that they were doing well or poorly during the experiment were unlikely to have had a differential effect on Hungarian test performance in the different groups.

6. How successful do you think you were at learning the English meanings of the phrases? (Please circle one and/or write a few words describing your overall opinion.)

Participants again chose from six options on this Likert-style question and responses were coded: Very successful (5 points), Quite successful (4 points), Somewhat successful (3 points), Not very successful (2 points), Not at all successful (1 point), and Unsure (0 points). Scores were somewhat higher than for the previous question, which is in line with actual performance on the spoken, verbatim Hungarian Production Test compared to the easier, meaning-based English Recall Test. However, many participants still reported that they felt unsuccessful at learning the English meanings of the Hungarian phrases. Seven participants, or 8.5%, felt 'Not at all successful' and 40 participants, or 50%, reported that they felt 'Not very successful' at learning the English meanings of the Hungarian phrases. Figure 4.19 shows the response frequencies in the four conditions and Table 4.30 shows the descriptive statistics for this question in the different groups.

Results showed that scores in the different groups were approximately equal. The Singing condition had the highest scores on this question, M = 2.5, while the Listen to Singing condition had the lowest scores, M = 2.2, with no participants in

Figure 4.19: Frequencies for Success at Learning the English Meanings – Study 2

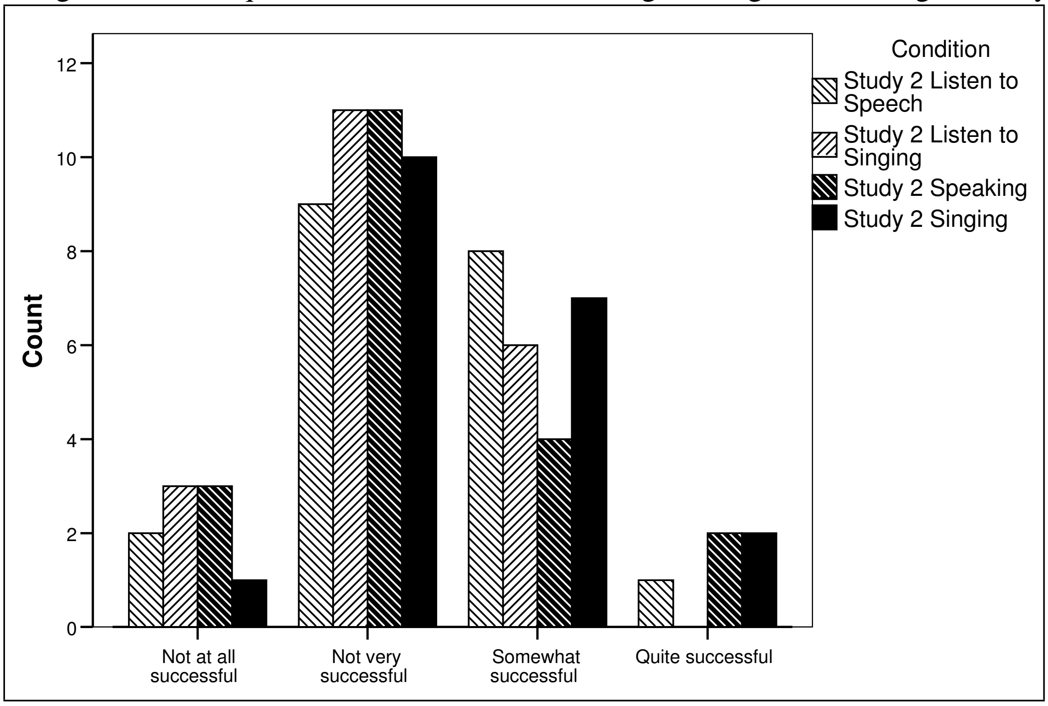


Table 4.30: Success at Learning the English Meanings of Phrases – Study 2

Group	*N*	*Mean*	*Median*	*SD*	*Range*
Listen to Speech	20	2.4	2.0	0.8	1 - 4
Listen to Singing	20	2.2	2.0	0.7	1 - 3
Speaking	20	2.3	2.0	0.9	1 - 4
Singing	20	2.5	2.0	0.8	1 - 4
Active (listen-and-repeat)	40	2.4	2.0	0.8	1 - 4
Passive (listening only)	40	2.3	2.0	0.7	1 - 4
Spoken stimuli	40	2.3	2.0	0.8	1 - 4
Sung stimuli	40	2.3	2.0	0.7	1 - 4
Overall (4 groups)	80	2.3	2.0	0.8	1 - 4
Overall: Male	40	2.5	2.0	0.8	1 - 4
Overall: Female	40	2.2	2.0	0.7	1 - 3

this condition reporting that they felt ‘Quite successful’ at learning the English meanings; no participants in any group reported that they felt ‘Very successful’ for this item.

Two-way ANOVA showed no significant differences for type of stimulus or for type of learning condition, nor was the interaction significant (p = .15). The non-parametric Kruskal-Wallis test results were somewhat similar for type of stimulus (p = .96 versus ANOVA p = .99) and for type of learning condition (p = .82 compared to ANOVA p = .56).

On this question, a significant gender difference was found, p = .038, with men reporting more success at learning the English meanings than women. Two by three factorial ANOVA did not show any other significant differences or interactions between gender, type of learning condition, and type of stimulus. Men had higher scores than women on this question, apart from in the Listen to Speech condition, where men and women had the same mean score, M = 2.4. This pattern of results may be linked to participants' change in motivation, for which women also showed a larger decrease overall in their motivation to learn the Hungarian phrases.

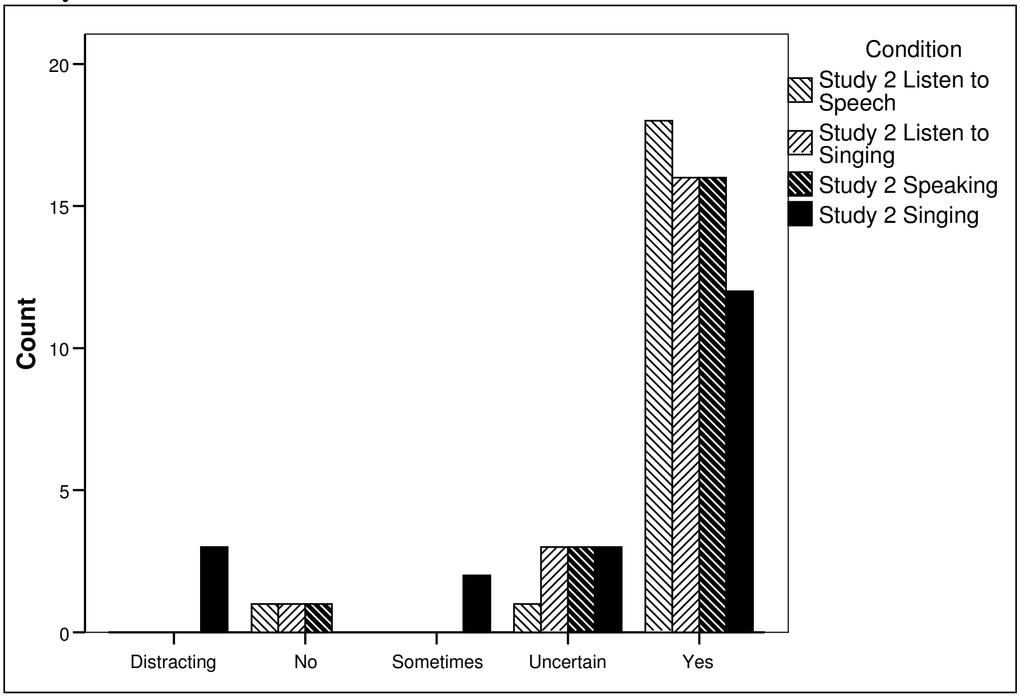
Overall, participants' self-perceived success at learning the English meanings of the Hungarian phrases were similar in the different groups, indicating that during the experiment, participants' beliefs about their learning were unlikely to have had a differential effect on Hungarian test performance at the group level. However, the overall gender difference for this question may have influenced the Hungarian test results; this is explored in section 4.5.

7. Do you believe that music and songs can support learning and/or memory?

This open-ended question was included to establish participants' beliefs about the utility of using music to support learning and also, particularly in the two sung stimuli groups, to explore whether (and to what extent) those beliefs might have influenced their Hungarian learning and test performance. If some participants who heard the sung stimuli believed that music can support learning, they might have felt more confident in their ability to learn the English-Hungarian phrases during the experiment. By contrast, if certain participants in the two sung conditions believed that music was not supportive of learning, this belief might have had a negative influence on their Hungarian test performance.

Individual responses to this question were varied, but could be coded into five categories: 'Yes' (4 points), 'Sometimes' (3 points), 'Uncertain' (2 points), 'Music during the learning process can be distracting' (1 point), and 'No' (0 points). Results were similar in the different groups, with most participants reporting that they believed music can support learning and memory; the most common response was 'Yes' in all of the groups, representing 77.5% of total responses. Figure 4.20 shows frequencies for each coded response in the four conditions and Table 4.31 shows the descriptive statistics for this question in the different groups.

Figure 4.20: Frequencies for Music and Songs Supporting Learning and/or Memory – Study 2



Although three participants in the Singing group wrote that they felt learning a melody at the same time as the phrases could be distracting, no individuals in that group wrote that they believed music and songs did not support learning and memory. One person in each of the other three groups thought music was not supportive.

Two-way ANOVA showed no main effect for type of learning condition, $p = .26$,

Table 4.31: Music and Songs Supporting Learning and/or Memory – Study 2

Group	*N*	*Mean*	*Median*	*SD*	*Range*
Listen to Speech	20	3.7	4.0	1.0	0 - 4
Listen to Singing	20	3.5	4.0	1.1	0 - 4
Speaking	20	3.5	4.0	1.1	0 - 4
Singing	20	3.2	4.0	1.2	1 - 4
Active (listen-and-repeat)	40	3.3	4.0	1.1	0 - 4
Passive (listening only)	40	3.6	4.0	1.0	0 - 4
Spoken stimuli	40	3.6	4.0	1.0	0 - 4
Sung stimuli	40	3.3	4.0	1.1	0 - 4
Overall (4 groups)	80	3.5	4.0	1.1	0 - 4
Overall: Male	40	3.4	4.0	1.2	0 - 4
Overall: Female	40	3.6	4.0	1.0	0 - 4

or for type of stimulus, p = .26. The interaction between type of learning condition and type of stimulus was not significant, p = .76. The Kruskal-Wallis tests showed similar results to the ANOVA, with no difference for type of learning condition (active or passive learning), p = .14, and for type of stimulus, the Kruskal-Wallis test was also not significant, p = .14. No main effect for gender was found, p = .36, nor were any interactions with gender observed in participants' responses to this question in any of the different groups.

Any effects of participants' beliefs regarding music's ability to support learning and memory on their Hungarian test performance are explored in section 4.5.

8. Would you suggest any changes or any ways of improving the learning experience for people participating in future studies?

Many participants (57.5%) chose not to answer this question, but some offered useful responses. Among these were that the inclusion of more visual aids would be helpful (e.g., 'It was easier when there was writing as well and even with a picture to get the context of the situation,' 'Pictures of the culture and people might be a motivating factor' 'Well for me, visual aids always help. Seeing a word burns it into my memory').

Another suggestion was to provide more variety in the type of practice during the learning sessions (e.g., ‘More matching of phrases – multiple choice questions,’ ‘Possibly break the phrases down into related modules based on context/situation,’ ‘Maybe more pictures or games in the middle to prevent the participant becoming bored’).

A third request was to reduce the number of phrases and the length of the learning sessions since the study was so long that some people felt tired (e.g., ‘Variation in the way it is taught – sample sentence exercises etc. The learning phase got too monotonous to focus,’ ‘Learning fewer phrases at a time,’ ‘A little bit shorter’).

9. Do you have any other comments or suggestions?

Most participants (87.5%) did not choose to answer this question, but some offered helpful responses. Among these were: ‘It was much easier when the text was also visible,’ ‘It would be way easier if we could write things down,’ ‘When having the small ‘conversation’, I didn’t have enough time to talk/respond back,’ and ‘Very interesting but incredibly hard.’

4.4.9 Age

Participants provided a self-report of their age at the start of the experiment session. To minimise the variability of experiment results due to age, only participants between the ages of 18 and 29 years were included in this study (two participants were removed from the data set due to age). Participants’ ages were approximately equal in the different groups, as shown in Table 4.32. The Singing condition was the youngest group, M = 21.0 years, and the oldest group was the Listen to Speech condition, M = 22.2 years.

Table 4.32: Age of participants in the four learning conditions – Study 2

Group	*N*	*M*	*SD*	Range
Listen to Speech	20	22.2 years	2.24	18 - 28
Listen to Singing	20	21.4 years	2.64	18 - 28
Speaking	20	22.0 years	2.59	18 - 28
Singing	20	21.0 years	1.76	18 - 25
Active	40	21.5 years	2.34	18 - 28
Passive	40	21.8 years	2.46	18 - 28
Spoken stimuli	40	22.1 years	2.48	18 - 28
Sung stimuli	40	21.2 years	2.23	18 - 28
Overall (4 groups)	80	21.6 years	2.34	18 - 28
Overall: Male	40	22.1 years	2.72	18 - 28
Overall: Female	40	21.1 years	1.91	18 - 26

Two-way ANOVA comparing type of learning condition and type of stimulus showed no age differences between the groups in terms of type of learning condition (active or passive), p = .58 and there was no interaction between type of learning condition and type of stimulus, p = .85. However, there was a marginal difference for type of stimulus, p = .078, with the older participants found in the two spoken stimuli groups and more younger participants in the two sung stimuli groups.

In addition, there was a trend for an overall gender difference, p = .061, with men on average older than the women. Two by three factorial ANOVA showed similar p-values when comparing age with gender, type of stimulus, and type of learning condition, with marginal differences again found for type of stimulus and for gender, but no significant interactions.

4.4.10 Gender

In this study, the four learning conditions were balanced for gender, with 10 male and 10 female participants in each condition. MANOVA showed no main effect for gender on the Hungarian tests for type of learning condition or for type of stimulus, and no significant interactions between gender and the different groups were found for the Hungarian tests.

However, significant and marginal gender differences were found for several of the individual differences measures used in this study, which were administered to establish whether the different groups were well matched for these potentially confounding factors. The Rhythm Discrimination sub-test of the Musical Ability Test also showed a main effect for gender, $p < .05$, with men showing higher performance on this measure.[27] Men reported that they felt more successful at learning the English meanings of the Hungarian phrases, $p < .05$,[28] and women reported a larger drop in overall motivation to learn the phrases as the experiment progressed, $p < .05$. In addition, a trend for an overall gender difference was found for scores on the Receptive Musical Ability Test, $p = .07$, with men outperforming women on this measure.[29] Marginal gender differences (with no interactions) were also found for Positive Affect on the PANAS mood questionnaire ($p = .079$ at pre-test and $p = .074$ at post-test), with men reporting more positive mood than women. A marginal gender difference was also found for age, $p = .061$, with the average age for men older than the average age for women.

Two by three factorial ANOVA also revealed significant interactions between gender, type of stimulus, and type of learning condition on several of the ID measures. Scores on the ten Likert-style questions of the Language Experience Questionnaire showed a significant interaction between gender and type of stimulus, $p < .05$, with men in the two sung stimuli groups showing higher scores than women. Similarly, on the ten Likert-style questions of the Musical Experience Questionnaire, an interaction was found between gender and type of stimulus, $p < .05$, with women

[27]The three-way interaction between gender, type of stimulus, and type of learning condition was also significant for the Rhythm Discrimination sub-test, $p < .001$, with women performing at a lower level than men in the Listen to Singing and the Speaking conditions.

[28]This pattern held except in the Listen to Speech condition, where scores were similar for both genders.

[29]The three-way interaction between gender, type of stimulus, and type of learning condition was also significant, $p < .001$, with women showing lower Receptive MAT scores than men in the Listen to Singing and in the Speaking conditions.

in the two spoken stimuli groups scoring higher than men. For the Melody Discrimination sub-test of the MAT, a three-way interaction was found between gender, type of stimulus, and type of learning condition, with men performing at a much higher level than women in the Listen to Singing condition. A significant three-way interaction was also found for overall MAT score, as well as an interaction between gender and condition, with men showing higher performance than women in the Speaking and Listen to Singing conditions (both interactions significant at the $p < .001$ level). A marginal interaction between gender and type of stimulus was found for overall Motivation to learn the phrases, $p = .086$, with men in the two sung stimuli groups reporting less motivation than women.

To summarise, no gender differences or interactions were found for the Hungarian tests in this study. However, for the individual differences measures, men tended to have higher scores than women, especially for musical abilities and for self-perceived success at learning the English meanings of the new phrases. By contrast, women were marginally more motivated to learn the Hungarian phrases than men in the sung stimuli groups, although across all groups women's self-reported motivation also decreased significantly over the course of the experiment session in comparison to men's change in motivation. Men also had marginally more positive mood scores than women at the start and at the end of the experiment sessions, and men were marginally older than women in this study.

In conclusion, the results of the measures of individual differences in this study showed that the groups were not well matched at the beginning of this experiment. In particular, group differences were found for the two Productive musical ability 'Happy Birthday' sub-tests, with the Singing condition showing the lowest performance on both of these tests, and for Phonological Working Memory with the two active groups (listen-and-repeat Speaking and Singing) showing higher performance. There were also a number of gender differences and interactions between women's and men's performance on the measures of individual differences. These findings are problematic because they cast doubt on the generalisability of participants' performance on the Hungarian tests in the different groups, with results potentially influenced by the group and gender differences. The next section describes statistical results that included the ID factors when comparing Hungarian test scores in the different groups, with the aim of providing a better understanding of the Hungarian test results and exploring the influence of these pre-existing differences on Hungarian test performance.

4.5 Influence of individual differences on Hungarian test performance

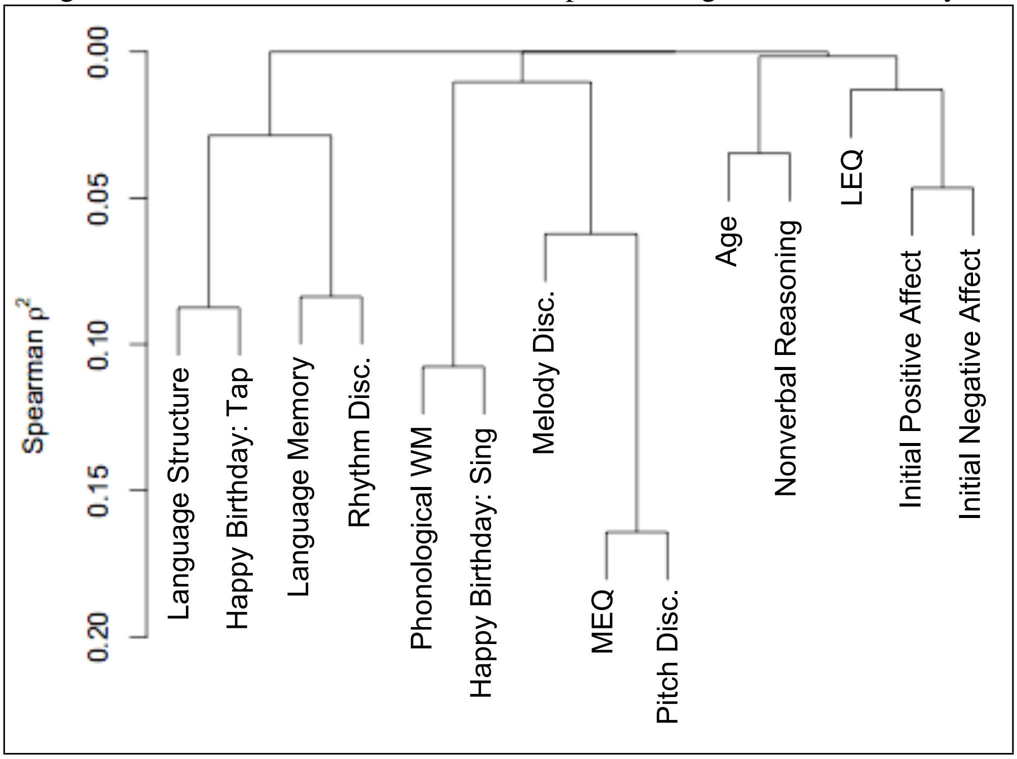
The statistical procedures detailed in this section explore the influence of the measures of individual differences on participants' Hungarian test performance. Because of significant group and gender differences and interactions found for several of the ID measures, it is important to investigate the extent to which these differences may have influenced participants' paired-associate foreign language learning. Further, this section explores whether the benefit observed in the previous study for singing to support verbatim, spoken foreign language skills can again be found, once the influence of the IDs are controlled for; or instead, whether the significantly higher Hungarian test scores in the Listen to Speech condition cannot be attributed to the effects of the ID factors.

4.5.1 MANCOVA controlling for pre-existing ID factors

Multivariate analysis of co-variance (MANCOVA) was conducted to explore the effects of the individual differences on Hungarian test performance. Apart from scores on the two LEQ and two MEQ sub-sections,[30] none of the 14 IDs pre-test measures correlated strongly with one another, as shown visually in Figure 4.21.[31] The IDs correlations showed that both of the MLAT-style language ability sub-tests clustered with the musical rhythm tests, while Phonological Working Memory clustered with the four non-rhythm music tests. The Nonverbal Reasoning factor clustered with Age.

With the dyslexia factor included for a total of 15 ID factors, the correlations showed a similar pattern except that the Language Memory Pre-test clustered with the musical rhythm tests and with the dyslexia factor, while the Language Structure Pre-test clustered with the Nonverbal Reasoning test (see Figure 4.22).

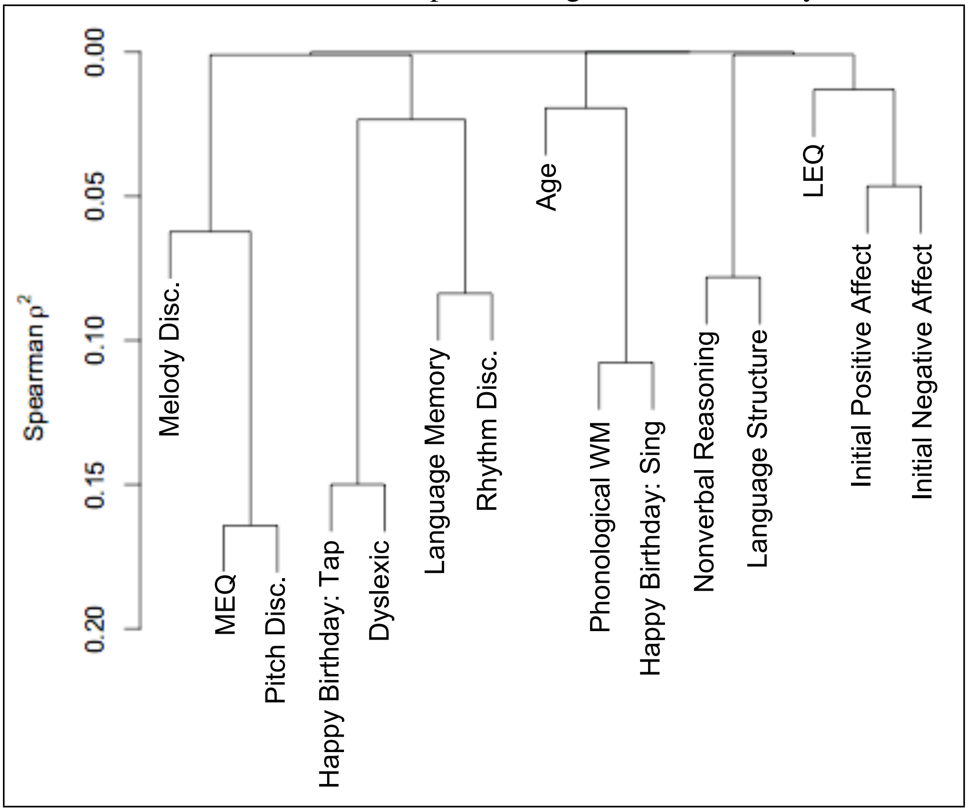
Figure 4.21: Correlations between the 14 pre-existing ID factors – Study 2



[30]Since both the LEQ and the MEQ sub-sections were significantly correlated at the $p < .001$ level, the total LEQ and MEQ scores were used.

[31]In this study, the Happy Birthday Singing and Tapping sub-tests did not correlate highly with one another (Pearson's $r = 0.23$) compared to correlations between the other ID measures, so both Productive MAT sub-test scores were used in the analysis.

Figure 4.22: Correlations between 14 pre-existing ID factors and Dyslexia – Study 2



MANCOVA for all four Hungarian tests

MANCOVA results for type of learning condition, type of stimulus, gender, handedness and the dyslexia factor with the other 14 ID measures as covariates (for a total of 17 ID factors) showed that the overall LEQ score was the best predictor of performance on the four Hungarian tests; although the significance level was marginal (p = .051) and the power was below .8, the effect size for LEQ score was large (partial η^2 = .17).

Other marginal ID factors for the four Hungarian tests in this study were the Language Structure Pre-test and Language Memory Pre-test scores (on which the listen-and-repeat Singing condition had the highest mean scores, n.s.), with p = .084 and p = .066, respectively, with observed power to detect an effect below .8 but large effect sizes. Marginal effects were also found for Initial Negative Affect (p = .095 and partial η^2 = .14) and for Handedness (p = .079 and partial η^2 = .15), again with low observed power.

Unlike for the MANOVA results for the data collected in this experiment, when controlling for all of the ID factors using MANCOVA, there was no overall main effect for Type of Learning Condition or Type of Stimulus on the four Hungarian tests (p = .43 and p = .73, respectively), nor for the interaction (p = .23), and the effect sizes were not large, although it should also be noted that the observed power to detect an effect was considerably lower than .8. In fact, for this second Hungarian study, none of the ID factors, learning conditions, or interactions had high enough observed power (above .8) to detect an effect. No significant interactions were found between the type of learning condition, type of stimulus, handedness, gender, and the dyslexia factor.[32] Table 4.33 shows the full MANCOVA results for the four Hungarian tests used in this study.

[32] An interaction between Type of Stimulus and Handedness was marginal (p = .053, partial η^2 = .17, power = .67). ANCOVA results for the Hungarian tests showed a similar pattern for the covariate ID factors, with variations for each test.

Table 4.33: MANCOVA for all four Hungarian tests with ID measures as covariates – Study 2

Effect	**df**	**Wilks'**	***F*-stat.**	***p***	**Part. η^2**	**Power**
Intercept	4, 50	.865	1.943	.12	.135	.545
Language Exp. Questionnaire	4, 50	.831	2.537	**.05**	.169	.676
Musical Exp. Questionnaire	4, 50	.972	0.360	.84	.028	.125
Phonological WM	4, 50	.990	0.123	.97	.010	.073
Rhythm Disc.	4, 50	.963	0.475	.75	.037	.153
Pitch Disc.	4, 50	.970	0.390	.82	.030	.132
Melody Disc.	4, 50	.874	1.800	.14	.126	.509
Happy Birthday: Sing	4, 50	.949	0.676	.61	.051	.205
Happy Birthday: Tap	4, 50	.886	1.610	.19	.114	.460
Language Structure Pre-test	4, 50	.851	2.185	.08	.149	.602
Language Memory Pre-test	4, 50	.841	2.361	.07	.159	.641
Nonverbal Reasoning	4, 50	.981	0.238	.92	.019	.098
Initial Positive Affect	4, 50	.891	1.530	.21	.109	.439
Initial Negative Affect	4, 50	.856	2.096	.10	.144	.582
Dyslexic	4, 50	.955	0.586	.67	.045	.181
Age	4, 50	.979	0.272	.90	.021	.105
Gender	4, 50	.955	0.594	.67	.045	.183
Hand	4, 50	.849	2.227	.08	.151	.611
Type of Learning Condition	4, 50	.928	0.976	.43	.072	.286
Type of Stimulus	4, 50	.961	0.503	.73	.039	.160
Type of Stim. x Type of Learn. Cond.	4, 50	.895	1.465	.23	.105	.421
Type of Learning Cond. x Gender	4, 50	.945	0.722	.58	.055	.217
Type of Learning Cond. x Hand	4, 50	.991	0.115	.98	.009	.072
Type of Learning Cond. x Dyslexic	4, 50	.869	1.880	.13	.131	.530
Type of Stimulus x Gender	4, 50	.954	0.603	.66	.046	.186
Type of Stimulus x Hand	4, 50	.833	2.513	**.05**	.167	.672
Type of Stimulus x Dyslexic	4, 50	.922	1.062	.39	.078	.310

MANCOVA for the spoken, verbatim Hungarian tests

MANCOVA results for the two spoken, verbatim Hungarian tests including type of stimulus, type of learning condition, gender, handedness, and dyslexia as factors (with the other 14 IDs as covariates) showed that the total percentage LEQ score had the highest significance level ($p = .04$), with a medium effect size (partial $\eta^2 = .12$), but insufficient power (below .8). Initial Negative Affect was a marginally significant predictor ($p = .099$, partial $\eta^2 = .09$, power = .47). Table 4.34 shows the full

MANCOVA results for scores on the verbatim, spoken Hungarian tests.

Table 4.34: MANCOVA for the spoken, verbatim Hungarian tests with ID measures as covariates (signif. code: * 0.05) – Study 2

Effect	**df**	**Wilks'**	***F*-stat.**	***p***	**Part. η^2**	Power
Intercept	2, 52	.970	0.815	.45	.030	.182
Lang. Exp. Quest.	2, 52	.884	3.412	**.04***	.116	.617
Music Exp. Quest.	2, 52	.987	0.353	.70	.013	.104
Phon. WM	2, 52	.994	0.151	.86	.006	.072
Rhythm Disc.	2, 52	.967	0.897	.41	.033	.196
Pitch Disc.	2, 52	.998	0.051	.95	.002	.057
Melody Disc.	2, 52	.931	1.934	.16	.069	.383
Happy Birthday: Sing	2, 52	.975	0.677	.51	.025	.158
Happy Birthday: Tap	2, 52	.962	1.018	.37	.038	.218
Language Structure Pre-test	2, 52	.959	1.122	.33	.041	.237
Language Memory Pre-test	2, 52	.961	1.051	.36	.039	.224
Nonverbal Reasoning	2, 52	.988	0.304	.74	.012	.096
Initial Positive Affect	2, 52	.959	1.109	.34	.041	.235
Initial Negative Affect	2, 52	.915	2.416	.10	.085	.466
Age	2, 52	.982	0.471	.63	.018	.123
Gender	2, 52	1.00	0.010	.99	.000	.051
Hand	2, 52	.933	1.881	.16	.067	.374
Type of Learning Condition	2, 52	.999	0.031	.97	.001	.054
Type of Stimulus	2, 52	.984	0.414	.66	.016	.114
Type of Stim. x Type of Learn. Cond.	2, 52	.902	2.823	.07	.098	.531
Type of Learning Cond. x Gender	2, 52	.954	1.257	.29	.046	.261
Type of Learning Cond. x Hand	2, 52	.992	0.215	.81	.008	.082
Type of Learning Cond. x Dyslexic	2, 52	.994	0.166	.85	.006	.074
Type of Stimulus x Gender	2, 52	1.00	0.002	.99	.000	.050
Type of Stimulus x Hand	2, 52	.950	1.364	.27	.050	.281
Type of Stimulus x Dyslexic	2, 52	.999	0.015	.99	.001	.052

After controlling for the effects of the ID factors on the two spoken Hungarian tests, in the second experimental study no main effects were found for type of stimulus or type of learning condition, nor were significant interactions found between gender, handedness, or dyslexia for the spoken Hungarian tests. The interaction between type of stimulus and type of learning condition was marginal, p = .07, but the interaction was not significant for either of the spoken Hungarian tests individually.

Because a marginal interaction was found between type of learning condition and type of stimulus, MANCOVA was conducted for the two spoken Hungarian tests while controlling for the effects of the ID factors in the four learning conditions. Results showed that handedness (p = .051), overall LEQ score (p = .061), and the Melody Discrimination test score (p = .091) were the best predictors of Hungarian test performance, although effect sizes were not large and the observed power for these factors was below .8. After controlling for the effects of the IDs, no overall group difference was observed on the two verbatim, spoken Hungarian tests, p = .85.

MANCOVA for the meaning-based Hungarian tests

MANCOVA was also conducted for the English Recall and the Hungarian Recognition tests to compare the contributions of type of learning condition, type of stimulus, gender, handedness and the dyslexia factor when controlling for the 14 ID measures as covariates. Results for these two Hungarian tests showed that the observed power was low (under .8) for all factors, and the Language Memory Pre-test score and Handedness were the only significant predictors of performance, both at the $p < .05$ level. There were also two marginal factors: LEQ score, p = .084, and Initial Negative Affect, p = .084, with medium effect sizes but low power. Table 4.35 shows the full MANCOVA results for the two meaning-based Hungarian tests in this study.

When controlling for the ID factors, no main effects or interactions were found between type of stimulus and type of learning condition for the two meaning-based Hungarian tests. However, there were significant interactions between type of stimulus and handedness at the $p < .05$ level, and a marginal interaction between type of learning condition and dyslexia, p = .083. ANCOVA showed that there was a main effect for handedness on the English Recall Test, p = .02, and a significant interaction between type of stimulus and handedness on the Hungarian Recognition Test, p = .01. Right- handed participants performed approximately the same regardless of type of stimulus, but left-handers in the spoken stimuli groups had higher scores than

Table 4.35: MANCOVA for the meaning-based Hungarian tests with ID measures as covariates (signif. code: * 0.05) – Study 2

Effect	**df**	**Wilks'**	***F*-stat.**	***p***	**Part.** η^2	Power
Intercept	2, 52	.884	3.424	.04*	.116	.618
Lang. Exp. Quest.	2, 52	.909	2.596	.08	.091	.495
Music Exp. Quest.	2, 52	.990	0.259	.77	.010	.089
Phon. WM	2, 52	.996	0.113	.89	.004	.066
Rhythm Disc.	2, 52	1.000	0.002	.99	.000	.050
Pitch Disc.	2, 52	.981	0.501	.61	.019	.128
Melody Disc.	2, 52	.969	0.835	.44	.031	.186
Happy Birthday: Sing	2, 52	.995	0.125	.88	.005	.068
Happy Birthday: Tap	2, 52	.936	1.778	.18	.064	.356
Language Structure Pre-test	2, 52	.926	2.092	.13	.074	.411
Language Memory Pre-test	2, 52	.857	4.350	**.02***	.143	.729
Nonverbal Reasoning	2, 52	.994	0.157	.86	.006	.073
Initial Positive Affect	2, 52	.931	1.927	.16	.069	.382
Initial Negative Affect	2, 52	.909	2.604	.08	.091	.496
Dyslexic	2, 52	.959	1.105	.34	.041	.234
Age	2, 52	.991	0.241	.79	.009	.086
Gender	2, 52	.963	1.008	.37	.037	.216
Hand	2, 52	.886	3.359	**.04***	.114	.609
Type of Learning Condition	2, 52	.952	1.325	.28	.048	.274
Type of Stimulus	2, 52	.991	0.238	.79	.009	.085
Type of Stim. x Type of Learn. Cond.	2, 52	.972	0.753	.48	.028	.171
Type of Learning Cond. x Gender	2, 52	.985	0.396	.68	.015	.111
Type of Learning Cond. x Hand	2, 52	.996	0.094	.91	.004	.064
Type of Learning Cond. x Dyslexic	2, 52	.909	2.618	.08	.091	.499
Type of Stimulus x Gender	2, 52	.955	1.232	.30	.045	.257
Type of Stimulus x Hand	2, 52	.882	3.475	**.04***	.118	.625
Type of Stimulus x Dyslexic	2, 52	.928	2.003	.15	.072	.395

right-handers; by contrast, right-handed participants in the sung stimuli groups had higher scores than left-handed participants.[33]

[33]However, due to the low observed power and the smaller number of left-handers (three in the spoken stimuli groups and five in the sung stimuli groups), this result should be interpreted with caution.

4.5.2 MANCOVA including debriefing questionnaire responses

An additional set of MANCOVA analyses were conducted, which in addition to the 17 ID factors already mentioned, included four items from the Debriefing questionnaire as covariates: overall Motivation, Change in Motivation, Success learning the Hungarian phrases, and Success learning the English phrases.

For all four Hungarian tests, the overall MANCOVA results including the four items from the Debriefing questionnaire was also not significant (p = .19, partial η^2 = .12, power = .45).[34] Only the Language Memory Pre-test score and Success learning the English phrases were significant predictors of performance on the four Hungarian tests (both ps < .05); the effect sizes for both factors were large, but only the Success learning the English phrases had adequate power above .8 (power = .84). A marginal effect for handedness was also found (p = .055), with observed power below the recommended .8 (power = .67), but a large effect size (partial η^2 = .18).

For the two spoken, verbatim Hungarian tests, MANCOVA results did not show any significant main effects or interactions between type of stimulus, type of learning condition and the other ID factors (overall p = .65). Success learning the English phrases and Handedness were significant factors, at the p < .05 level, both with observed power lower than .8 but large effect sizes.

MANCOVA conducted for the two meaning-based Hungarian tests showed an overall trend (p = .06, with a medium effect size but low power = .55), but no main effects or interactions between type of stimulus and type of learning condition were observed, and none of the four debriefing questions contributed significantly. Only the Language Memory Pre-test, Initial Negative Affect score and handedness appeared as significant predictors of performance, all at the p < .05 level, all with observed power

[34] ANCOVA calculations for the individual Hungarian tests showed that type of learning condition, type of stimulus, and the interaction were not significant for any of the tests.

below .8 but large effect sizes. A significant interaction between type of stimulus and handedness was found for the two meaning-based Hungarian tests (p = .046, partial η^2 = .12, power = .60).

When including the type of stimulus, type of learning condition, gender, handedness, and dyslexia without the other 14 ID factors, MANCOVA results for all four Hungarian tests showed that responses to two of the debriefing questions were good predictors of overall Hungarian test performance. Success learning the English phrases was highly significant (p = .001, partial η^2 = .26, power = .96) and Change in Motivation was marginal (p = .068, partial η^2 = .13, power = .64). Participants' overall Motivation and Success learning the Hungarian phrases were not significant, nor were any other main effects or interactions observed for the four Hungarian tests (apart from handedness at the p < .05 level, partial η^2 = .15, power = .71). These MANCOVA results indicated that participants' self-assessments of their overall success at learning the paired-associate phrases were at least moderately accurate, and that while Motivation to learn the phrases was not a good indicator of Hungarian test performance in this study, a self-reported Change in Motivation (generally a decrease in motivation) did predict performance.

4.6 Discussion

The results of this controlled experimental study showed no facilitation for paired-associate foreign language phrase learning in the active learning conditions or in the sung stimuli conditions. Contrary to predictions, in this study the passive Listen to Speech condition performed best on all four of the Hungarian language tests. A significant interaction between type of learning condition and type of stimulus was found for the Hungarian Production Test (one of two tests that showed a significant learning benefit for the Singing group in the previous study), for the English Recall

Test, and for the overall Hungarian test score. In addition, MANOVA showed a marginal difference for type of stimulus on the Hungarian Recognition Test, with the two spoken stimuli groups performing at a higher level than the two groups that heard sung phrases.

However, in this study the Hungarian test data for several groups was not normally distributed, very significantly so for some of the tests (with skewness or kurtosis values greater than 2), which casts some doubt on the validity of these statistical results. Also in contrast with the previous study (in which the groups were fairly well matched on the measures of individual differences), in this study significant group differences and interactions were found for some of the ID factors. In particular, the Singing group had much lower performance than the other three groups on the test of productive musical skills. Also, in the present study there were no gender differences found for performance on the four Hungarian tests, but significant group differences and interactions were observed for gender on several ID factors. Thus, it is somewhat problematic to directly compare the Hungarian test performance by participants in the different groups using MANOVA.

When the ID factors were included as covariates using MANCOVA, group differences in Hungarian test scores did not show the same level of significance as was found in the MANOVA results described in section 4.3. MANCOVA also showed that the level of observed power was generally insufficient to detect an effect for the data collected in this study (below .8), despite medium to large effect sizes for particular factors. When controlling for the 14 IDs as covariates and including handedness, gender and dyslexia, MANCOVA revealed no significant main effects for type of stimulus or type of learning condition, nor were any significant interactions observed between these factors on any of the four Hungarian tests in the current study. These MANCOVA results indicate that the pre-existing group differences and significant interactions found on the measures of individual differences in section 4.4

strongly influenced Hungarian test performance in the present study.

Nevertheless, the MANCOVA analyses did highlight a few individual differences factors that significantly influenced Hungarian test scores in this study, again bearing in mind that the observed power was often insufficient. Of particular importance in this study were participants' handedness,[35] previous language learning experience (LEQ scores), Initial Negative Affect, and the Language Structure and Language Memory Pre-test scores.[36] From the Debriefing questionnaire, the Success learning the English phrases was a highly significant factor predicting Hungarian test performance (with observed power greater than .8 and a large effect size). Change in Motivation was a marginal predictor of Hungarian test performance when all four Hungarian tests were included in the MANCOVA calculation but without including the other 14 ID factors as covariates. The importance of participants' perceived success learning the phrases and change in motivation in the MANCOVA results suggests that researchers should consider these factors when conducting studies on foreign language learning, even over the course of a short experiment – especially since in this study, participants' feeling of success at learning the phrases contributed more to Hungarian test performance than their prior L2 learning experience did.

In sum, the results of this study did not confirm findings from the previous experiment, which did show a facilitation for paired-associate English-Hungarian phrase learning in the listen-and-repeat Singing condition. Unexpectedly, in this study the listen-and-repeat Speaking and Singing conditions performed at a lower level on the Hungarian tests compared to performance in the corresponding Speaking and Singing conditions in the previous experiment; the decrease in performance was significant for the listen-and-repeat Singing condition. There are several possible explanations for these unexpected differences in Hungarian test performance. The

[35]In this study, there were far fewer left-handed participants than right-handers (8 vs. 72), and the different groups were not particularly well matched for handedness (see page 151 for more details).

[36]In this study, the groups were well matched for the MLAT-style Language Structure and Language Memory Pre-test scores (which was not the case in the last study). The groups were also well matched for overall LEQ scores (see page 125 for more details) and for Initial Negative Affect (see page 147).

participants may have been more tired at the end of this study (with four learning sessions rather than three learning sessions in the previous version) since the experiment lasted for approximately 75 minutes rather than 60 minutes. However, this would not explain why the Listen to Speech condition performed at a higher level than the other three groups on all four of the Hungarian tests.

Another possible explanation for the difference in performance between the two experiments is that participants in the two active learning conditions in the current study were less motivated to learn or felt less successful at learning the Hungarian phrases. This could be a reasonable explanation since participants' feeling of success learning the English meanings was a significant contributor, and change in motivation was a marginal predictor of Hungarian test performance, but there is no way to compare motivation or feelings of success in the two studies because there were no questions about this on the Debriefing questionnaire in the previous study.

The two sung stimuli learning conditions were also younger overall than the two spoken learning conditions, and in the last study there was an interaction between age and Hungarian test performance because the older participants tended to report higher LEQ scores. However, mean LEQ scores were significantly higher in the two sung groups compared to the two spoken stimuli groups in the current study (particularly in the Listen to Singing group), so this is not a very plausible explanation.

A fourth explanation for the unexpected results could be due to the significant group differences and interactions on the individual differences factors in the current study, whereas performance on the ID measures were quite similar in the three groups for the first experiment. In particular, since participants in the listen-and-repeat Singing group in this study had significantly lower productive musical abilities than those in the other three groups, and musical training and skills have shown correlations with speaking and pronunciation skills in a foreign language (Slevc & Miyake, 2006; Pastuszek-Lipinska, 2008), participants in the Singing group may have been at a disadvantage for L2 phrase learning through an aural/oral learning procedure. Also, a listen-and-repeat singing method for L2 learning might not have been a good way for individuals who felt less confident singing due to an aptitude-treatment interaction. If participants in the Singing group did not feel comfortable and confident singing, this in turn could have decreased their motivation to learn the phrases. Indeed, on the Debriefing questionnaire, the Singing group reported the lowest overall motivation, while in the Listen to Speech group – a more common way someone might try to learn phrases in a new language – participants reported the greatest overall motivation to learn the Hungarian phrases.

Another potential explanation for the unexpected results is that, due to technical difficulties, most participants in the Listen to Speech condition did not take part in the experiment at the same time that they were revising for exams. This could explain their overall more positive mood at the start of the experiment, their higher motivation to the learn the phrases, and/or their smaller decrease in motivation to learn over the course of the experiment session. A comparison of results from the first 20 participants in the Listen to Speech group whose results were excluded from analysis due to incomplete data[37] showed that participants' performance on the three remaining Hungarian tests in the four 'original' groups were similar. And for the 80 participants who took part in the experiment during the exam period, the listen-and-repeat Singing condition had the highest performance on the Delayed-Recall Hungarian Conversation (n.s.). For the measures of individual differences, the 80 participants who did the experiment during the exam period were similarly poorly matched for productive musical skills at the start of the experiment, again with the Singing condition achieving the lowest scores on this measure.[38]

[37]Namely, no Hungarian Production Test, which along with the Delayed-Recall Hungarian Conversation, was one of the two tests which showed a significant benefit for the Singing condition in the previous study.

[38]Please see Table B.2 on page 383 in the appendix to compare ANOVA results on the ID measures with the complete data set described in this chapter, to Table B.1 on page 382 which shows the ANOVA results for the four groups that instead included scores from the 20 participants in the Listen to Speech group who did not complete the Hungarian Production Test.

The fact that the Hungarian test results from the ‘original’ group of 20 participants in the Listen to Speech group were more in line with those of the other groups may suggest that the 80 participants who did the experiment during the exam period felt more stress (and/or had insufficient quantities of sleep and so on) because they were still revising for exams. By contrast, the 20 participants in the Listen to Speech group who completed the experiment when the exam period had ended may have had less overall anxiety and felt more confident to learn the phrases. This may be reflected in the Listen to Speech participants’ more positive mood and higher motivation to learn the phrases compared to the other groups presented in this chapter, although it must be stated that the group differences for those ID factors were not significant.

Thus, the higher Hungarian test scores in the Listen to Speech group whose results were described in this chapter may be due to timing (and/or anxiety) because they were not revising for exams. This interpretation of the results is also supported by the fact that performance by the 16 participants in the pilot study for this experiment tended to have much higher scores than participants in the main study, and they did not complete the experiment during the exam period either (although another important difference is that the pilot study participants also had only three learning sessions, rather than four).

It is unfortunate that we were unable to answer the question of whether there would be a difference between active and passive second language learning in this study. While there were no significant differences on Hungarian test performance between the two sung stimuli groups (both of which did the experiment during the exam period), the evidence does not allow us to determine whether singing is important, because the participants in the Singing group had lower productive musical skills than participants in the Listen to Singing group. It may be the case that when productive musical skills are equal, there will be a greater benefit for listening and singing along – or that singing may be more engaging and fun than only listening for some people, but listening to sung phrases may still provide some learning benefits for people who do not wish to sing. Future research could explore these questions.

In summary, the MANOVA results for the current study showed a significant interaction between type of stimulus in an unexpected direction for the Hungarian tests, with the Listen to Speech condition showing the highest Hungarian test performance. However, in this study there were significant group differences and interactions for a few of the ID measures, and when controlling for the 14 ID factors plus gender, handedness, and dyslexia, MANCOVA showed no significant group differences or interactions between type of stimulus and type of learning condition for Hungarian test performance. Indeed, the variation in Hungarian test scores in the different groups for this study was so large that the observed power to detect an effect for the learning condition or for the IDs rarely reached the recommended level of .8. Thus, the results of the MANCOVA analyses suggest that the pre-existing differences between the groups may, at least in part, explain why the results of this study were not in the expected direction and were not in agreement with the results of the previous experimental study.

Conclusion

The results of the research studies presented so far have shown that singing can support short-term foreign language phrase learning in a controlled, randomised, experimental design when the groups were well matched for the individual differences factors (Chapter 3). By contrast, in the current study there were large pre-existing group differences on some of the ID measures, particularly for productive musical skills, with very low performance in the listen-and-repeat Singing condition, and the IDs differences had a large influence on participants' Hungarian test performance. In both experiments, a participant's previous language learning experience, mood, and musical skills were good predictors of performance on particular Hungarian tests. This chapter also showed that motivation (or more specifically, change in motivation) and self-perceived success at learning the L2 phrases are also very important factors in predicting success in L2 learning in the experimental laboratory. Results of this second experiment also raise the possibility of an aptitude-treatment interaction, with the participants who had lower scores on a test of productive musical skills performing at a lower level on the Hungarian tests when they were randomly assigned to the listen-and-repeat singing condition for this paired-associate L2 learning task.

In the next chapter we shift away from the experimental laboratory to the development and implementation of a classroom-based research study. The chapter explores whether songs can support learning in the modern foreign language classroom, by providing cognitive or memory benefits and/or providing fun or motivating L2 learning activities. The study also examines to what extent particular individual differences may influence foreign language learning in the classroom environment, and explores the children's self-reported experiences and perceptions of learning a new language through musical and dramatic activities.

Chapter 5

Singing in the Modern Language Classroom: A Four-Week, Classroom-Based Arts Intervention Study

This chapter considers a practical question of whether listening to songs and singing can support beginning-level foreign language learning in the classroom. This educational environment and distributed practice learning procedure contrast with the controlled context and massed practice used in the two experimental studies. Anecdotal evidence about learning benefits and pedagogical recommendations from teachers to use music in the L2 classroom are frequently put forward (Murphey, 1992; Spicher & Sweeney, 2007), but it was important to explore the utility of a musical pedagogical approach for L2 learning and also to explore the influence of individual differences between learners. This study also employs opinion questionnaires to evaluate whether songs may support modern foreign language learning in other ways, such as by improving children's enjoyment or motivation to learn the new language, in addition to potentially providing memory benefits for L2 material.

For this quasi-experimental research study, a listen-and-repeat learning procedure was adapted to the modern language classroom using a variation of the

Contemporary Music Approach (Anton, 1990)[1] and recommendations for teaching choral music in a foreign language (Welch, 2007), with mini-lessons incorporating songs and dramatic dialogues that were developed to supplement the French curriculum. The present chapter explores whether the benefits for singing compared to speech that were observed in the first experimental study might occur in a longer-term, more ecologically valid educational setting. It also investigates the influence of several measures of individual differences on the French test results and reports the children's feedback regarding their experiences and perceptions of L2 learning through the additional musical and dramatic activities used in class.

5.1 Background

Songs have been shown to support literacy (Jalongo & Ribblett, 1997), science knowledge (Foster et al., 1999; Kimmel, 1998), spelling (M. Martin, 1983), and verbatim verbal memory (Thaut et al., 2008; Rainey & Larsen, 2002; Calvert & Tart, 1993) in the native language. To briefly summarise previous findings of research investigating L2 learning with music, limited support for claims of improvements in specific language areas has been found. Fomina has conducted research with adults learning English as a foreign language (EFL) in the Ukraine, showing benefits for the use of songs in learning pronunciation, prosody, and culture (Fomina, 2000, 2002, 2006). Medina (1993) showed that young Spanish-speaking children learned more English vocabulary words over a two-day period when the words were presented through songs with illustrations, compared to a presentation through speech with illustrations. Murphey (1989) found that adolescent children in rural Switzerland were very interested in learning English songs and wanted to talk about music. He reported that incorporating music into the L2 curriculum was an effective way to increase pupils' motivation to learn a new language. I previously conducted a six-

[1] Please see Chapter 2 for more details about Anton's method.

week arts intervention study with two pre-existing classes of 13-year-old children in Scotland and found that the children's overall L2 French skills in the music condition improved at a significantly higher rate than the children in the visual art and drama group, $p < .05$. Overall, in both classes the children also improved on the art form they had been practising in French class (singing or drawing), while no improvement in scores was observed for the art form the child was not learning. Although the artistic skills improvements did not reach a statistically significant level, these artistic gains did appear to transfer to improvements in French learning (Ludke, 2006).

Smith Salcedo (2002) conducted a research study with four classes of English-speaking university students who were learning Spanish. She investigated whether the introduction of Spanish songs into the classroom would improve learners' vocabulary during one 10-week semester. Learning was measured using a Cloze (fill-in-the-blank) pre-test and post-test for each song because some researchers believe that Cloze tests can assess learners' overall competence in the foreign language (Heilenman, 1983; Fischer, 1981; Hanzeli, 1977), although it is unclear to what extent scores on these tests can accurately reflect learners' L2 speaking and pronunciation skills. Smith Salcedo found that the two groups which had listened to songs did perform at a higher level at post-test than the speaking and control groups did, although the difference reached statistical significance for only two of the three songs which were used in her main study. She also found that listening to the music (without the words) during the final test resulted in no significant memory benefit. Further, the study showed that significantly more learners reported the 'din' phenomenon if they had heard songs in class than if they heard a spoken version of the text, $p < .05$ using a chi-squared test, with a mean percentage of 66.7% in the music class and 78% in the melody class (which heard the song melody during testing) compared to 33.3% for the class with the spoken text. Text recall after a delay of two weeks showed no significant difference between the classes.

In Sposet's (2008) review of literature investigating the use of music and songs to support second language education, she concludes that there is insufficient empirical evidence to support this claim. She also calls for more research – both experimental and classroom-based studies – that include control groups, particularly requesting studies with children under the age of 18 and with participants learning languages other than English. The study described in this chapter attempts to meet these criteria.

5.1.1 Research question

This chapter explores the effect of using songs in the modern foreign language classroom by investigating similar questions to Smith Salcedo (2002), but with younger, secondary school children learning French rather than adults enrolled in an introductory Spanish course at the university level. The present study investigates whether a learning procedure that incorporates French songs and singing into the curriculum can support French vocabulary and grammar learning for secondary school pupils. The primary research question is: What is the effect on pupils' vocabulary and grammar learning when new French material is presented in the foreign language classroom through songs compared to (spoken) dramatic dialogues for a total of 75 minutes over a two-week, distributed learning period?

The secondary research question asks to what extent individual differences might influence vocabulary and grammar learning through these two presentation methods. Factors considered were age (S1 level children compared to S2 level children), gender, previous language learning and artistic experience, preferences for drama, music, and/or visual art.

The study also explores the children's opinions about the new activities using feedback gathered from two questionnaires. In particular, it asks whether the pupils themselves report any benefits or challenges when learning French with these listening materials and activities. In addition, is there a higher reported incidence of 'din' for the songs than for the dramatic dialogues? Did the children's preferences for listening to the song or to the dramatic dialogue have an influence on their French vocabulary and grammar learning through these different materials?

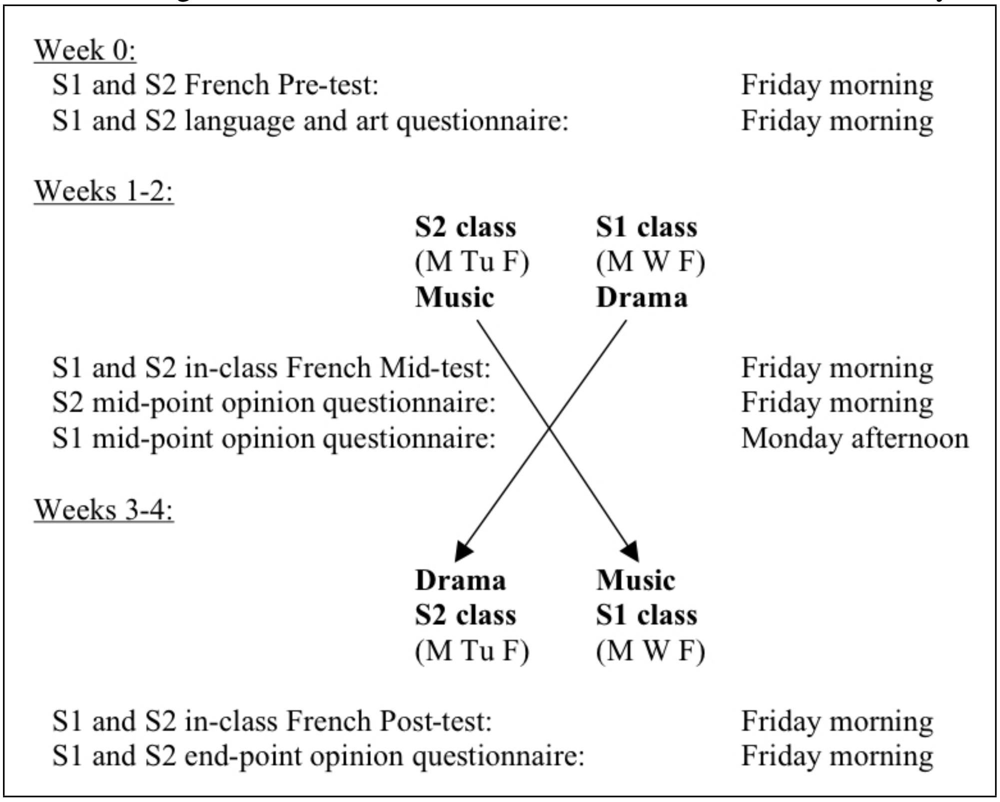
5.2 Method

This study was set up as a crossed, quasi-experimental design over a four-week period. Quasi-experimental designs are common in educational research studies where it is not feasible to break up pre-existing groups (Freeman & Tijerina, 2000), and where the question of interest is the causal influence of a treatment condition on, for example, learning outcomes. In the current study, a quasi-experimental design was used to investigate any differences in French vocabulary and grammar learning in two classroom groups after the inclusion of two arts-based teaching techniques. In each French class, two weeks were spent learning a dramatic dialogue or a song, followed by two weeks spent learning French with the other auditory art form. The method differs from an interrupted time-series design because there is only one pre-intervention data collection point and because three of the six measures were only administered once. Figure 5.1 shows an illustration of the course of this four-week study.

5.2.1 Participants

Children from two pre-existing classroom groups in a religious charter school in Edinburgh, Scotland participated in this study. This secondary school was chosen based on the recommendation of the City of Edinburgh's Children & Families Department and access was granted through the Researchers in Residence programme, which aims to foster secondary school pupils' interest in attending

Figure 5.1: Course of French four-week arts intervention study



university and conducting research. The school had five teachers in the modern languages department, who collectively taught French, Spanish, German and Italian. The teachers were very supportive of this research project because they often used drama, visual art and music to teach modern languages in this school and they wished to explore the impact of these activities on the children's learning. They also wanted information about how artistic activities could support the aims of Scotland's Curriculum for Excellence by providing learning opportunities that enabled children to become successful learners, confident individuals, responsible citizens, and effective contributors.[2]

There were 29 children in the S1 (beginning French) class and 30 children in the S2 (beginning-intermediate French) class involved in this study. Full Pre-test, Mid-

[2]More information about the Curriculum for Excellence is available at: http://www.ltscotland.org.uk/curriculumforexcellence/ (Last checked: 12 December 2009) and more details related to how the guidelines apply to modern languages are available in Appendix C.

point, and Post-test data was collected for 19 of the S1 pupils and 24 of the S2 pupils. The overall age range of pupils in the two classes was between 11 and 14 years old. The mean age of children in the S1 class was one year younger than the children in the S2 class (M = 12 years and M = 13 years, respectively), with only small age variations in each class.[3] Since the children in each French class were of approximately the same age, it is unlikely that these small differences would be a strong between-participants factor that influenced the children's French learning.

One important difference between the two classroom groups is that six children in the S1 class had special learning needs, of whom three were included in the full data analysis. All three of the included children were male, and two were also bilingual; the third boy had difficulties organising himself and there was suspicion of dyslexia. No children with special learning needs were found in the other class because in this school, the S2 French classes were divided into sets and the S2 class participating in this study was the top set of three groups of S2 French learners.

A language learning experience questionnaire (LLEQ) and an artistic experience questionnaire (AEQ) were also administered to pupils before the arts intervention began, to establish whether any pre-existing individual differences were present between learners in the two classes. Table 5.1 reports the univariate ANOVA results of the individual differences measures for children in the two classroom groups.[4]

A significant group difference was observed for Age, which was expected since the children in the S1 class were on average one year younger than the children in the S2 class. However, a group difference was also found on the first, Likert-style sub-section of the LLEQ ($p < .01$) and an overall trend was observed for the total LLEQ percentage score (this score weights both LLEQ sub-sections equally), $p = .08$, with

[3]Table 5.9 later in this chapter shows more details about the children's ages, including the ages of children in each classroom group separated by gender.

[4]These p-values are equivalent to those that would be found using Student's t-test because there are only two groups in this study, but because MANOVA and MANCOVA are used later in this chapter to investigate the French test results, ANOVA values are reported here.

Table 5.1: Univariate ANOVA for ID measures in the two French classes (signif. codes: *** 0.001, ** 0.01)

ID Measure	*N*	df	Sum Sq.	*F*-stat.	*p*-value
Lang. Learning Exp. Quest. (%)	42	1, 41	0.061	3.158	.08
- LLEQ sub-section 1 raw score	42	1, 41	402.2	7.449	**.009** **
- LLEQ sub-section 2 raw score	42	1, 41	2.279	0.260	.61
Artistic Exp. Questionnaire (%)	42	1, 41	0.002	0.209	.65
- Musical Exp. raw sub-score	42	1, 41	1.986	0.026	.87
- Drama Exp. raw sub-score	42	1, 41	13.42	0,381	.54
- Visual Art Exp. raw sub-score	42	1, 41	1.111	0.094	.76
Bilingualism	42	1, 41	0.037	0.229	.64
Age	42	1, 41	9.520	54.74	**.000** ***
Gender	42	1, 41	0.002	0.008	.93

higher scores in the S2 class. More details about the measures of individual differences are presented in section 5.4, while the effects of these IDs on the children's French learning are explored in section 5.5.

5.2.2 Materials

Three types of materials were developed for this study, consisting of: (1) overall curriculum, learning outcomes and daily lesson plans, (2) French language tests, and (3) mid-point and post-intervention opinion questionnaires. All of the lessons, French tests, and questionnaires were prepared in advance and administered by the researcher during the children's French class sessions. The materials designed for this study are described in this section.

Curricular aims, lesson plans, and additional materials

The activities related to the songs and dramatic dialogues were designed to be both fun and challenging, reinforcing French material that the children were learning in their textbooks while supporting the aims of the Curriculum for Excellence. Six 15-minute mini-lessons were created for each class session during each two-week period, to be taught by the researcher during regular class time. The lesson plans were approved by the French teacher before the intervention study began.

The majority of the grammatical structures and vocabulary in the songs and dramatic dialogues were familiar to the children or were being covered in the current unit of the textbook, but the texts also included some French material that was new and more complex than the children had previously seen. Unfamiliar vocabulary items were included because most studies exploring the benefits of including songs in the L2 classroom have assessed vocabulary learning (Sposet, 2008; Schunk, 1999; Iwata, 2005; Medina, 1993). New, challenging grammatical structures were also included to follow up the results of a previous study which showed that using songs (compared to visual art and drama activities) over a six-week period in the L2 classroom had the largest benefit on a test which assessed the children's grammatical skills (Ludke, 2006). The two French songs were chosen using the following criteria, adapted from Ludke (2006):

- Have a strong melody that is easy to sing, particularly in the refrain (Roehm, 2001; Wallace, 1994).
- Include rhymes to help facilitate memorisation and association between the written text and the sounds (Bower & Bolton, 1969; Rubin & Wallace, 1989; Smith & Phillips Jr, 2001).
- Use famous songs by native speakers to provide an accurate model for pronunciation (Anton, 1990).
- Have a strong tune-text association; that is, the words, structures, and syllables that are most important (the ones that are being taught) should last longer than the other notes, be higher pitched, or stand out in some way (Bottari & Evans, 1982; Van Voorhis, 2002; Yalch, 1991).

In addition, the mini-lessons included a variety of activities that gave children the opportunity to practise different skills that supported the aims of the Curriculum for Excellence. Activities requiring individual work in addition to pair-work, larger group and full-class activities allowed the children to learn in different ways. The activities also enabled the children to practise becoming more confident individuals and responsible citizens, and to effectively contribute to the group while successfully learning the French material.

Six mini-lessons were created for each two-week period of the quasi-experimental arts intervention study. A summary of the lessons for each class session and the corresponding learning outcomes are described below.

Day 1: Introduction to the song or dialogue. Pupils draw pictures related to words they recognise or how they feel while listening to the recording, then listen again and complete a Cloze exercise.
Learning outcome: Practise French listening skills by filling in words missing from the text.

Day 2: Pupils listen to the song and finish filling in the missing French words on their worksheets. Then the children answer a few comprehension questions in French.
Learning outcome: Practise French listening skills and identify themes and facts about the text through listening and reading.

Day 3: Teacher provides correct words for pupils to fill in and pupils listen to the teacher's English translation of text. Pupils listen again to the recording and follow along by reading the words on their worksheets.
Learning outcome: Understand overall themes and specific details about the French text.

Day 4: Pupils create their own comprehension questions for a partner to answer, then practise repeating the song or dialogue aloud as a group.

Learning outcome: Obtain and provide information by reading the French text and practise French speaking skills.

Day 5: Practise saying the dramatic dialogue or singing the song twice as a group in preparation for the audio recording.

Learning outcome: Practise French speaking skills.

Day 6: Pupils complete translation mid/post-tests, Cloze tests and opinion questionnaires. Create audio recording of the pupils saying the dialogue or singing the song. Teacher plays back the recording for the children to hear.

Learning outcome: Demonstrate recall of the text practised in class and French pronunciation skills.

Additional materials used for the French lessons included:

- CD recordings of two professionally recorded French songs: 'Les Bonbons' by Jacques Brel (S2 class) and 'Le Tourbillon' by Jeanne Moreau (S1 class).
- CD recordings of two French dialogues recorded to a professional standard by the researcher, produced by native French speakers: 'Une conversation au parc' (S1 class) and 'Un dialogue au lycée' (S2 class).
- A brief biography, written in French, about the singer of each of the two French songs that was distributed to pupils as part of the introduction during the first song lesson.
- Cloze (fill-in-the-blank) exercises for the French texts (the children in each class completed a Cloze exercise for one song and for one dialogue). For the Cloze exercise, the missing words consisted of vocabulary and grammar the

children already knew, were currently learning in class, or which were not too difficult to guess while listening to the audio recordings.

French language tests

A vocabulary and grammar translation pre/mid/post-test and two Cloze post-tests were specially developed to assess the children's French learning during the arts intervention study. Each measure is described in more detail below. The scoring procedures are outlined in section 5.2.5 and the full list of questions for these French tests is available in Appendix C.

Translation of French phrases into English. A pre/mid/post-test was developed to measure changes in the children's translation of phrases from the new language (French) into the native language (English). This test consisted of 10 items, five of which were short French phrases taken from the song while the five remaining phrases came from the dramatic dialogue used in each class. Both grammar and vocabulary learning were assessed using this test.[5]

Cloze Post-tests: Fill in missing French words. For this measure, every 7th French word was missing from the text of the dramatic dialogue or song that the children had been learning during the five previous class sessions. The children were asked to fill in as many blanks as possible and their written responses were collected for comparison to the original French text. Cloze tests were used because researchers have claimed that this type of test can provide an assessment of overall linguistic competence in a new language (Heilenman, 1983; Fischer, 1981; Hanzeli, 1977). However, since the two Cloze tests were only administered as post-tests after pupils had been practising the text for several days, in this study the Cloze tests are more

[5]More details about the scoring for vocabulary and grammar learning are provided in section 5.2.5.

likely providing a measure of vocabulary learning and memory.

Opinion questionnaires about the in-class French activities

Two opinion questionnaires were completed by pupils, once at the mid-point and once at the end of the arts intervention (e.g., at the end of each two-week period). These questionnaires provided feedback from pupils about the artistic activities we had been using together in class and whether, and how, they felt the activities had supported their French learning. The mid-point opinion questionnaire consisted of seven questions related to the study's research questions and to questions relating to Scotland's recently implemented Curriculum for Excellence. The post-test opinion questionnaire was similar to the mid-point opinion questionnaire, but with the addition of two questions asking children if the words from the dramatic dialogue or song had repeated in their heads outside of class (providing a measure of reported 'din' for the spoken text and for the sung text) and whether they had preferred listening to the song or to the dramatic dialogue together in class. For reference, the items on these questionnaires are available in Appendix C. These two questionnaires were important because the pupils' responses could provide valuable, practical information about their personal experiences of French learning through these musical and drama activities.

5.2.3 Teaching procedure

Before the arts intervention began, all pupils and the French teacher were given basic information about the aims of the study, including a description of the activities that we planned to do together in class, but without specific details about the expected outcomes. The children were told that their class had been selected to take part in a research project exploring the use of the arts on foreign language learning. The participating children were given the opportunity to ask the researcher questions about the project.

During the arts intervention, the researcher taught the first 10-15 minutes of each class period. This time was spent listening to and learning a dramatic dialogue or a song in French, with related activities to help pupils understand and remember the material. The activities for the arts intervention lasted approximately 90 minutes during each two-week period with each French class, for a total of three hours spent using the new materials during the four-week intervention period. Of this time, approximately 75 minutes was spent learning each dialogue or song, with the 15-minute period on the final day spent on testing and audio recording the children saying the dialogue or singing the song they had learned.

The majority of the in-class instruction time during the arts intervention (between 2 and 2.5 hours per week) was taught by the children's regular French teacher, who was a native speaker of French. The decision to use classes with only one teacher was made to decrease the amount of variance in learning stemming from environmental factors, differences in teaching methodology, or student-teacher rapport.

During the first 10-15 minutes of each class session, pupils were able to view the words of the French song or dialogue and helped to understand the meaning of the words and phrases. Pupils were never allowed to take a written copy of the dialogue or lyrics out of the classroom and they did not see a written translation in English at any point during the arts intervention. After three sessions of listening activities (approximately 45 minutes), pupils were encouraged to speak or sing aloud in French to practise the dramatic dialogue or song. Games and friendly competitions between one side of the class and the other side of the class or boys and girls were incorporated into the lessons, to encourage the children to say or to sing the French words aloud. The regular French teacher and the researcher were both in the classroom at all times that the two French classes met during the four-week intervention period. The researcher helped the teacher and pupils, and the researcher's presence also ensured that the children were not practising or translating the song or dramatic dialogue material except during the specified times.[6]

A DVD showcasing the children's final performances of the dramatic dialogue and the song used in each class, with subtitles in both English and French, was created after the study ended using the audio recordings made at the end of each two-week period. Copies of the DVD were given to the school at two final class sessions where the results from the research study were presented to the teacher and pupils.

The researcher's observations of each S1 and S2 class session and the materials, homework and activities presented by the regular French teacher were also noted. These field notes were intended to establish whether there were any systematic differences in the classroom during either of the two-week intervention periods that could have influenced the children's French learning. Unfortunately during the second two-week period, the S1 class completed French tests on the material they had been learning in their textbook (including speaking, writing, listening, and reading tests). As a result, the classroom atmosphere was more tense than during the previous two-week period. In addition, the song was more difficult than the dialogue that the children in the S1 class had heard (toward the end of the song, the lyrics came so fast that they almost qualified as a tongue twister). In addition, the S1 class was not able to listen to the song as many times as they had previously heard the dramatic dialogue, partly because of testing and also because of technical problems (no laptop computer and the SmartBoard would not function) on the day scheduled for going through the French words and teaching the English translation of the song. This difficulty with the words is reflected in the S1 children's rendition of the song, which was audio recorded during the final session and can be found on the DVD in Appendix E.

[6]In the S1 class, at times there was a classroom assistant to help the six children who had learning difficulties; however, she was not present during the French tests.

5.2.4 Testing procedure

For this short quasi-experimental study, all questionnaires and French tests were administered during the children's regular French classes, which were held in the same classroom, with natural and artificial lighting and adequate heating. In both classes, the pre/mid/post-test and questionnaire data was collected in the morning because the school had half-days on Fridays. The classroom was set up with desks arranged in three rows, five deep, with two desks side by side in each row. The researcher and French teacher were both present during the tests to discourage anyone attempting to copy a classmate's answers.

On the Friday before the arts intervention began, the pupils completed the French-to-English translation pre-test, the language learning experience questionnaire (LLEQ) and the artistic experience questionnaire (AEQ). The researcher gave pupils approximately 20 minutes to complete the translation pre-test and the questionnaires (the data was collected at the end of the S2 class session and at the beginning of the S1 class session).

At the mid-point and at the end of the arts intervention study, testing took place at the start of class and consisted of: (1) a Cloze test with every 7th word missing from the French dialogue/song the children had been learning in class, (2) a mid/post-test of phrases in French which were to be translated into English, and (3) an Opinion Questionnaire about the in-class French activities. The French Cloze and translation tests were administered together, without giving children the opportunity to listen to the dialogue or song prior to completing these measures. A ten-minute time limit was imposed for completion of the two French tests. After we made an audio recording of the children speaking or singing the material they had learned for two weeks in class, the children completed the Opinion Questionnaire.

5.2.5 Data analysis

Scoring of French tests

Translation of French phrases into English. This pre/mid/post-test had 10 French phrases (five from the song and five from the dramatic dialogue) which the children were asked to translate into English as best they could. No points were deducted for misspellings of the English words. Scores on this repeated measure were expected to show a gradual improvement as the children learned the song and dramatic dialogue during the arts intervention.

Initially, two grammar scores were calculated for this translation test: (1) one point (out of 10 possible) was awarded for each perfect translation, where both the grammar and vocabulary accurately expressed the English meaning of the French phrase ('Perfect'); and (2) one point (out of 10 possible) was awarded for each English translation of a French phrase with approximately the correct meaning, where the grammar and vocabulary were close to the correct English meaning, but not precisely, either due to the use of one incorrect verb tense or form, or due to one vocabulary item that was inaccurately or not translated ('Approx'). No points were awarded for grammar if the translated phrase was not at least approximately correct. Because each of the two grammar scores were quite low individually, the perfect translation score (Perfect) and the approximately correct grammar score (Approx) were added together to create an 'Acceptable' grammar score (out of 10 possible points). The Acceptable grammar scores are used in the majority of the statistical analyses.

Vocabulary scores for the French translation pre/mid/post-test were calculated by adding together the number of correctly translated English words. To receive a point for the vocabulary score, using the correct form or tense of the noun or verb was irrelevant – a point would be awarded for correctly translating the general meaning (e.g., translations such as ‘brought’ or ‘bring’ would be awarded 1 point for the vocabulary score, even if the correct translation was ‘had brought’).

Cloze Post-tests: Fill in missing French words. The children’s attempts to fill in the blanks from the dramatic dialogue or from the song were checked for accuracy against the original text. Several scores were calculated for each Cloze test: (1 – ‘Perfect’) the number of perfect answers, including correct French spelling and accent marks; (2 – ‘Misspelled’) the number of answers that were misspelled, but otherwise correct; (3 – ‘Acceptable’) the total number of perfect plus misspelled answers; (4 – ‘Incorrect’) the total number of incorrect answers; (5 – ‘Attempted’) the total number of blanks with an answer (including correct, misspelled, and incorrect answers). Since the total number of points possible on each Cloze test varied, ratio scores were calculated for each Cloze test for the ‘Acceptable’ scores out of the total possible (6 – ‘Acceptable/Possible’) and the number of ‘Acceptable’ answers out of the total number attempted (7 – ‘Acceptable/Attempted’). The ‘Acceptable/Possible’ ratio score was important because of its insensitivity to the varying numbers of blanks to fill in on the two Cloze post-tests administered to each group. Thus, only the results of the total ‘Acceptable’ Cloze post-test scores and the ‘Acceptable/Attempted’ ratio scores are reported, since these scores were most directly relevant to the research questions regarding the children’s French learning.

Scoring of opinion questionnaires

The mid-point and post-test opinion questionnaires contained open-ended questions related to the study’s research questions, exploring the children’s own reflections on their experience learning these French songs and dramatic dialogues, and whether they felt these artistic activities had supported their progress in achieving the aims of Scotland’s Curriculum for Excellence. The two opinion questionnaires were analysed by sorting the children’s responses to each question into different categories and counting the number of similar responses to each question. For a few questions that were less open-ended, it was also possible to calculate some descriptive statistics. The pupils' responses to the two opinion questionnaires are discussed in more depth in section 5.3.2.

5.3 Results

As mentioned previously (see the Participants section on page 195), due to pupil absences during the four-week arts intervention period, complete pre/mid/post-intervention data was collected for 19 (of 29) of the beginning-level French pupils (S1 class) and 23 (of 30) of the beginning-intermediate French pupils (S2 class). The eight bilingual children's scores in the two classroom groups are included in the data analysis, but will also be listed separately as descriptive statistics. First the French test results are described, followed by discussion of the children's responses to the opinion questionnaires.

5.3.1 French tests

The French tests developed for this study were a French-to-English translation administered as a pre/mid/post-test to assess vocabulary and grammar learning. Responses were also collected for two Cloze post-tests with every 7th French word removed from the dialogue or song the children had been learning in class. The results of each measure are discussed in turn.

Translation of French phrases into English

As described in section 5.2.5, the Acceptable grammar score was used in these statistical analyses because the overall mean grammar scores were very low in both classes. Table 5.2 shows the descriptive statistics for the pre-test, mid-point, and post-test grammar scores on the French-to-English translation test. A separate score was calculated for vocabulary (the number of translated words with the correct meaning, but not necessarily in the exact form), as shown in Table 5.3. Performance on the French translation test was generally higher in the S2 class, particularly for vocabulary. Despite the fact that the tests included phrases taken from a previously unknown song or dialogue, this result is not surprising since the S2 pupils had already taken one year of French when the arts intervention began.[7]

Table 5.2: French Translation Test: Grammar scores ('Acceptable')

Pre-test	*N*	*M*	*SD*	Range
S1 Class	19	0.32 points	0.58	0 - 2
S2 Class	23	0.87 points	0.97	0 - 3
Mid-Point test				
S1 Class	19	1.05 points	1.13	0 - 3
S2 Class	23	2.00 points	1.35	0 - 4
Post-test				
S1 Class	19	0.79 points	1.13	0 - 3
S2 Class	23	2.65 points	1.61	0 - 6

Table 5.3: French Translation Test: Vocabulary scores

Pre-test	*N*	*M*	*SD*	Range
S1 Class	19	3.26 points	4.24	0 - 12
S2 Class	23	6.96 points	6.49	0 - 20
Mid-Point test				
S1 Class	19	7.53 points	7.20	0 - 22
S2 Class	23	14.2 points	8.00	0 - 32
Post-test				
S1 Class	19	6.89 points	6.72	0 - 22
S2 Class	23	16.0 points	7.37	4 - 32

Repeated-measures ANOVA showed that overall French grammar and vocabulary translation scores improved significantly from the Pre-test to the Mid-Point test, both at the $p < .01$ level in the two classroom groups. No difference was found between the two groups for change in performance at the mid-point of the arts

[7]MANOVA showed significant group differences for all of the French measures used in this study apart from the second Cloze post- test ($p = .10$), with the S2 class consistently outperforming the S1 class (at the $p < .05$ and $p < .01$ levels).

intervention. This trend of improvement on the French translation test continued for the S2 class from the Mid-Point test to the Post-test with another significant increase in mean grammar and vocabulary scores, $p < .01$ (see Figures 5.2 and 5.3 on the next page). By contrast in the S1 class, no improvement for the French translation grammar or vocabulary scores was observed from the Mid-Point test to the Post-test; in fact, overall performance decreased slightly for the Post-test. The marked decrease in scores in the S1 class may have resulted from having had less time to learn the words of the song because of French testing during regular class time, and/or due to the song's greater level of difficulty compared to that of the dramatic dialogue.

For the French translation grammar and vocabulary scores, the bilingual children tended to perform at a slightly higher level than their native English speaking peers. This pattern was not true for the vocabulary translation Mid-Point test scores in the S1 class,[8] but the mean differences for bilingualism on the French tests were small. Overall, no significant main effect was observed for bilingualism, nor were the interactions between bilingualism, group, or gender significant. However, gender differences were observed for the French translation Pre-test and Mid-Point vocabulary scores, both at the $p < .05$ level, with girls outperforming boys on these measures. The gender difference for the French vocabulary translation score at Post-test was no longer significant, $p = .11$.

[8] In part, this may be because two of the bilingual boys in the S1 class also had special learning needs.

Figure 5.2: French Translation Grammar Test scores in the two groups (10 points possible) (signif. code: ** 0.01)

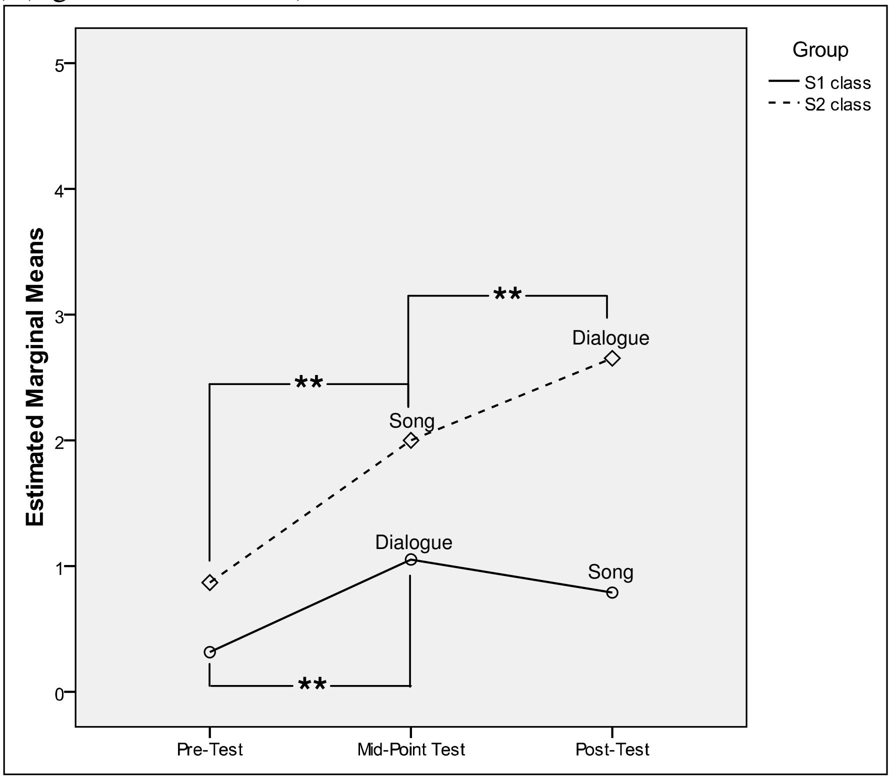
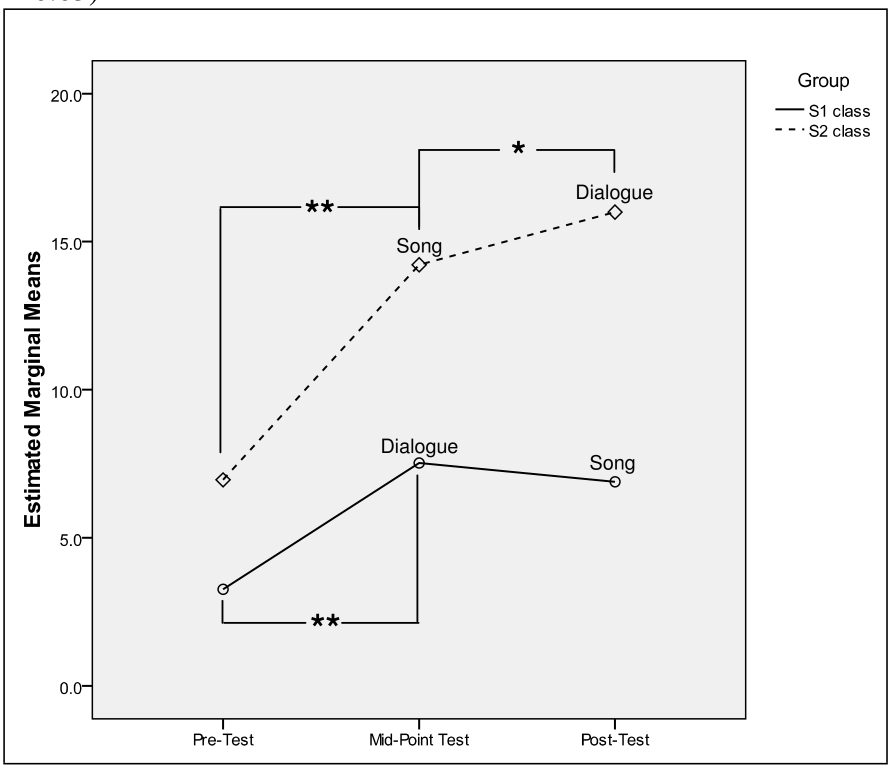


Figure 5.3: French Translation Vocabulary Test scores in the two groups (signif. codes: ** 0.01, * 0.05)



Cloze Post-tests: Fill in missing French words

After the first two week-period of the arts intervention, the Cloze post-tests consisted of 23 items in the S1 class (dramatic dialogue) and 32 items in the S2 class (song). For the second two-week period, the Cloze post-tests consisted of 40 items in the S1 class (song) and 37 items in the S2 class (dramatic dialogue). Table 5.4 shows the mean 'Acceptable' (Correct + Misspelled) scores in each class on the Cloze post-tests.[9]

Table 5.4: French Cloze post-tests: 'Acceptable' scores

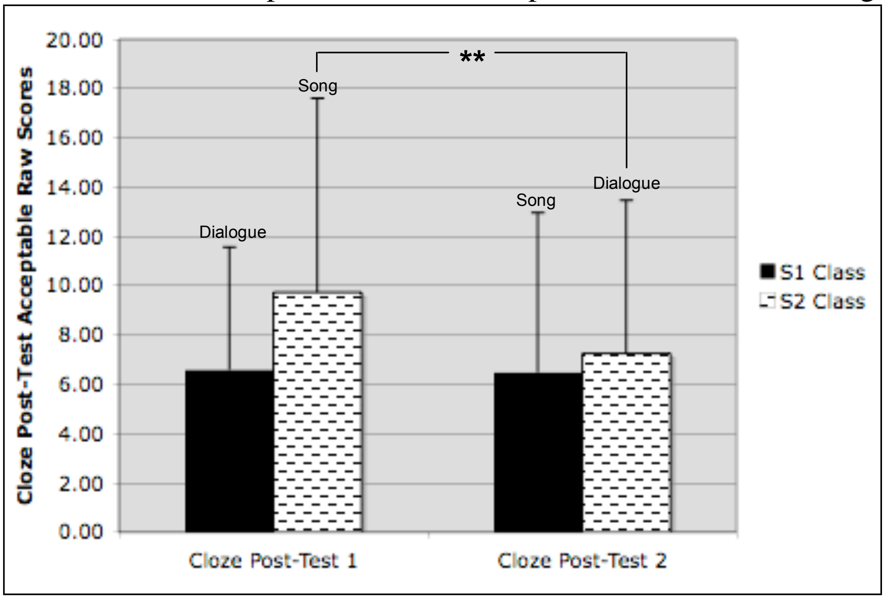
Cloze Test 1	*N*	*M*	*SD*	Range	Possible
S1 Class (dialogue)	19	7.95 points	5.23	0 - 20	0 - 23
S2 Class (song)	23	17.1 points	5.79	8 - 31	0 - 32
Cloze Test 2					
S1 Class (song)	19	7.58 points	7.92	0 - 27	0 - 40
S2 Class (dialogue)	23	13.7 points	7.37	2 - 29	0 - 37

Again on this measure, performance was generally much higher for the S2 class compared to the S1 class. However, the large decrease in the S2 class for the mean Acceptable words from the first Cloze test to the second Cloze test reached statistical significance ($p < .01$), as shown in Figure 5.4. This may have reflected an observed general decrease in interest or motivation for the activities, perhaps because so many French tests were being administered, or perhaps the children were less motivated to learn the dramatic dialogue than they had been to learn the song. By contrast, the rate of Acceptable words on the second Cloze post-test did not decrease substantially in the S1 class, although the mean was still consistently lower than overall scores in the S2 class, with some children in the S1 class still earning zero points on this test.

For the first Cloze post-test, the native English speaking children had slightly higher scores than their bilingual peers. The opposite pattern was found on the second

[9] As previously mentioned, the French teacher thought that the material and rate of presentation of the words in this song was too challenging for beginning-level French learners. The level of difficulty may have had a negative impact on results.

Figure 5.4: French Cloze post-tests: Raw acceptable answers in the two groups



Cloze post-test in the S1 class; however, the mean differences were small and no significant differences were observed between the groups for bilingualism on these measures. Significant gender differences were found for the first Cloze post-test ($p < .05$) and for the second Cloze post-test ($p < .01$), with girls outperforming boys on these measures. In addition, there was a marginal interaction between gender and bilingualism on the second Cloze post-test, $p = .052$, with the bilingual girls outperforming boys; a similar pattern was observed for the first Cloze post-test, but the difference was smaller, $p = .20$. No three-way interactions between group, gender, and bilingualism were observed on either of the Cloze post-tests in this study.

For the ratio of ‘Acceptable’ answers to the ‘Attempted’ responses on the two Cloze post-tests, results in the S1 class showed greater consistency, whereas the mean ratio and standard deviation for scores in the S2 class decreased slightly from the first Cloze post-test to the second Cloze post-test. Table 5.5 shows the descriptive statistics for the ratio of Acceptable (Correct + Misspelled) to Attempted responses in each class on the two Cloze post-tests.[10]

[10]Because one child made zero attempts to fill in the blanks on the first Cloze post-test and six children made no attempts on the second Cloze post-test, unfortunately this ratio could not be calculated for those children, so the total number of participants is correspondingly lower in the S1 class on these measures.

Table 5.5: French Cloze post-tests: Ratio of Acceptable:Attempted answers

Cloze Test 1	*N*	*M*	*SD*	Range	Possible
S1 Class (dialogue)	18	.86	.14	.50 - 1	0 - 1
S2 Class (song)	23	.86	.16	.48 - 1	0 - 1
Cloze Test 2					
S1 Class (song)	13	.86	.18	.40 - 1	0 - 1
S2 Class (dialogue)	23	.74	.20	.38 - 1	0 - 1

The native English speakers again had very slightly higher ratio scores than their bilingual peers on the first Cloze vocabulary test, with the opposite pattern on the second Cloze vocabulary test, but again no significant group differences were observed for bilingualism on either of the Cloze post-tests. No gender differences were observed, nor were any significant interactions found between gender, group, and bilingualism for the ratios of Acceptable:Attempted answers on the two Cloze post-tests.

5.3.2 Opinion questionnaires

On the mid-point and post-intervention opinion questionnaires, most children in the two classes (92.4%) reported that they enjoyed trying the artistic activities in class. Many of the pupils reported that they felt that the new activities had helped improve their French skills, specifically mentioning vocabulary, speaking, pronunciation, and listening skills. Several pupils offered helpful observations and suggestions for how to improve the activities for future projects, including the idea of including videos to accompany the songs and dialogues, sometimes watching with English subtitles, and incorporating more variety into the activities by sometimes dividing the class into even smaller groups. In answer to the research questions for this study, more than half of the children in the two classes (52%) reported that the lines from the song had repeated in their heads after French class had ended; the incidence of 'din' was much less common for the dramatic dialogue, at 9.4%. Table 5.6 shows the children's responses to question 8, regarding their preferences for the song vs. the dramatic dialogue at the end of the arts intervention. The percentage of children preferring the song to the dramatic dialogue was greater than 50% in both classes.

Table 5.6: Listening preferences for French song vs. dramatic dialogue

	S1 Class	**S2 Class**	**Both Classes**
Song	56.5% (13 pupils)	68.0 % (17 pupils)	62.5% (30 pupils)
Dramatic dialogue	8.7% (2 pupils)	24.0% (6 pupils)	16.7% (8 pupils)
Both	13.0% (3 pupils)	0.0% (0 pupils)	6.25% (3 pupils)
Blank or unclear	21.7% (5 pupils)	8.0% (2 pupils)	14.6% (7 pupils)

Comments from the children regarding the artistic activities:

'I enjoyed trying to guess the words in the empty gaps. I enjoyed this because it was quite hard to hear all the words, this was challenging!'

'My favourite activity was in a day reading the dialogue we have been learning. It was fun because we got to learn a lot of new words in French.'

'Singing all together. It was a really fun song and even more fun to sing it.'

'When we were recorded because it's minted listening to it.'

The children also reported that the additional artistic activities had helped them to achieve all four of the aims of the Curriculum for Excellence. The activity that was most often cited as an example of Successful Learner was learning new vocabulary words. For Confident Individual, many children reported that speaking or singing in French had helped them achieve this goal. The activity most often cited for Responsible Citizen was active participation in the activities. For Effective Contributor, the most common example was that they had answered comprehension questions and speaking or singing out loud. Many children also wrote that they felt the activities had helped improve their French skills, particularly in terms of listening and speaking skills. There were no noticeable differences in the children's responses to the opinion questionnaires between the S1 and S2 classes.

Successful Learner (35 children in the two classes)

'I think I was a successful learner because I widened my vocabulary using the song in just a few days.'

'I felt I was a successful learner during the singing because I understood everything even things we had just learnt.'

Confident Individual (32 children in the two classes)

'Singing because I felt confident to sing loud and I liked the feeling of confidence.'

'My favourite activity was when we said the part of Rachelle in the story [dramatic dialogue] because it made me more confident in reading in French out loud.'

Responsible Citizen (6 children in the two classes)

'I felt I listened to everything and did the work I was given. I also understood some parts in the song.'

Effective Contributor (19 children in the two classes)

'An effective contributor in the recordings because I was loud and I tried to pronounce everything properly.'

'I thought that I was an effective contributor when we were singing "Le Tourbillon" [The Whirlwind] as I tried hard to sing it well.'

Several Aims of the Curriculum for Excellence

'I think that I completed all these in the talking exercise because I was able to speak without hesitation, I learnt new words, I contributed to group talking and stuck to the dialogue.'

'Reading out you had to be an effective contributor and a confident individual.'

Overall, the results of this arts intervention support educators' claims that presenting new foreign language material through both musical and dramatic activities can provide an enjoyable and challenging addition to French class for beginning-level learners, even though a greater learning benefit was not observed for the songs compared to the dramatic dialogues. Artistic activities also enabled pupils to achieve the aims of Scotland's Curriculum for Excellence in the modern language classroom. The next section describes several measures of individual differences and the following section explores their influence on the children's French learning in each class.

5.4 Measures of individual differences

It could be argued that measures of individual differences might be less crucial in a crossed research design where all participants take part in both learning conditions. However, collecting background information was important to compare whether the children were well matched in terms of these factors and the extent to which differences in the children's learning in the two groups or within each group could be explained by their prior language learning experience, by their initial motivation to learn French, or by their training and preferences for particular art forms. Thus, prior to start of the arts intervention programme, data from two background questionnaires and a few other individual differences were collected in both classes. The complete list of questions for all of the ID measures is available in Appendix C.

The Language Learning Experience Questionnaire (LLEQ) contained questions about the children's prior language learning experience, including their motivation and enjoyment of learning French. The Artistic Experience Questionnaire (AEQ) gathered information about the children's artistic experience, including musical, dramatic, and visual art training. Both the LLEQ and AEQ were both administered on the Friday prior to the start of the arts intervention. Results of the ID measures are described in this section, with scores reported separately for age, gender, and bilingualism wherever these factors had a significant influence on results.

5.4.1 Language learning experience questionnaire

The LLEQ was composed of 10 Likert-scale questions, plus four items about the amount of time the children had spent in non-English speaking foreign countries and any language learning prior to travel. The questions were similar to those contained in the questionnaires from the two experimental studies (Chapters 3 and 4).

For the ten Likert-scale items, 'Very untrue' was awarded one point; 'Somewhat untrue' received two points; 'Equally true and untrue; unsure' earned three points; four points for 'Somewhat true'; and five points were awarded for 'Very true'. A total of between 10 and 50 points were possible on the first section. For the second part of the Language Learning Experience Questionnaire, responses earned between 0 and 2, 4, and 5 points for each question, with a possible total between 0 and 16 points.[11] Table 5.7 shows the descriptive statistics for each LLEQ sub-section and the overall LLEQ percentage score.[12]

Univariate ANOVA results showed a main effect for group on the first section of the LLEQ, with the S1 class showing significantly lower mean scores than the S2 class, $p < .01$. There was also a marginal group difference for the LLEQ item 'I find it easy to learn French,' with children in the S1 class reporting lower scores on this

[11] The questions and scoring details for this questionnaire are available in the appendix on page 421.
[12] An overall percentage score was used in order to weight each LLEQ sub-section equally.

Table 5.7: Language Learning Experience Questionnaire scores for children in the S1 and S2 classes

LLEQ Total Percent	*N*	*M*	*SD*	Range	Possible
S1 Class	19	47.2%	15.9%	24.4 - 83.4%	0 - 100%
S1 Class: Male	8	52.1%	15.8%	31.1 - 83.4%	0 - 100%
S1 Class: Female	11	43.7%	15.7%	24.4 - 80.8%	0 - 100%
S2 Class	23	54.9%	12.1%	37.5 - 81.5%	0 - 100%
S2 Class: Male	10	50.8%	8.4%	37.5 - 65.1%	0 - 100%
S2 Class: Female	13	58.0%	13.8%	37.4 - 81.5%	0 - 100%
LLEQ sub-section 1	*N*	*M*	*SD*	Range	Possible
S1 Class	19	31.0	8.7	23 - 50	10 - 50
S1 Class: Male	8	33.1	9.8	23 - 47	10 - 50
S1 Class: Female	11	29.5	7.9	25 - 50	10 - 50
S2 Class	23	37.2	6.1	24 - 49	10 - 50
S2 Class: Male	10	35.5	4.8	24 - 34	10 - 50
S2 Class: Female	13	38.5	6.8	34 - 49	10 - 50
LLEQ sub-section 2	*N*	*M*	*SD*	Range	Possible
S1 Class	19	5.18	3.54	0 - 14	0 - 16
S1 Class: Male	8	6.06	3.90	0 - 11	0 - 16
S1 Class: Female	11	4.55	3.30	2 - 14	0 - 16
S2 Class	23	5.65	2.39	3 - 12	0 - 16
S2 Class: Male	10	4.90	1.79	3 - 9	0 - 16
S2 Class: Female	13	6.23	2.68	3 - 12	0 - 16

question.[13] No group differences were found for the second section of the LLEQ (p = .61), but a trend was observed for the overall LLEQ percentage score (p = .083). In part, these differences may be due to the age difference between the children in the two classroom groups, or it may also be due to the S2 class being the top set for French, whereas the S1 children had not yet been divided into sets.

The LLEQ results also revealed that 12 children in the two French classes were not monolingual native speakers of English. Unsurprisingly, the eight bilingual children for whom complete data was collected had significantly higher scores on the LLEQ than their native English speaking peers. Three-factor MANOVA including

[13]It should be noted that Levene's test of equality of error variances for this question was violated, p = .001, indicating that the variance of scores in the two groups were not very similar. The variance on the final Likert-style LLEQ question about beginning to learn a foreign language during primary school was also very different in the two classes, p = .003.

group, gender, and bilingualism showed the differences between the bilingual children and their native English speaking classmates were significant at the $p < .01$ level for the second section and overall on the LLEQ, while the difference on the first section of the LLEQ was marginal, $p = .065$. Further details about the bilingual children are provided in subsection 5.4.4.

For gender, boys in the S1 class had slightly higher mean LLEQ scores than girls, whereas girls generally had higher scores than boys in the S2 class. No main effects for gender were observed on the LLEQ measure.[14] However, three-factor MANOVA showed a significant interaction between bilingualism and gender in this study, $p = .014$; the three-way interaction between group, gender, and bilingualism was highly significant, $p = .001$. However, these results should be treated with caution because there was only one bilingual girl in the S1 class and she had a much higher score than the two bilingual boys in her class, both on the second section of the LLEQ and overall (80.8% vs. $M = 47.4\%$, which was lower than the native English speaking boys' mean score in the S1 class, $M = 53.6\%$). Total percentage LLEQ scores for the four bilingual girls ($M = 71.8\%$) and the boy (65.1%) in the S2 class tended to be more similar, and scores by their native English speaking classmates were also fairly balanced for gender ($M = 51.9\%$ for girls and $M = 49.2\%$ for boys). No gender differences were found on any particular Likert-style LLEQ item.

5.4.2 Artistic experience questionnaire

The first section of the AEQ consisted of 10 Likert-scale questions that were adapted from the musical experience questionnaire developed for the two experimental studies. Information about each child's background and training in three different art forms were collected for this study: music, drama, and visual art. For the first AEQ

[14]One-way ANOVA showed a marginal effect for gender on the second section of the LLEQ, $p = .09$, with girls showing higher scores than boys.

section, pupils responded to four Likert-scale items about their musical experience, three items about their experiences with drama, and three questions about their experiences with visual art. The second section of the AEQ contained twenty-five additional questions designed to provide a fuller picture of each child's musical, dramatic, and visual art training, as well as the child's artistic interests and preferences.

The first AEQ section was scored in the same manner as the first part of the LLEQ, and again a total of between 10 and 50 points were possible for these Likert-type items. The twenty-five questions on the second AEQ section was scored by categorising the children's responses and assigning a value between 0 and 5 for each item.[15] AEQ scores were calculated for each of the two sub-sections and an overall percentage for the two parts of the questionnaire. Table 5.8 shows descriptive statistics for the three AEQ scores. Separate sub-scores were also calculated for the children's musical, dramatic, and visual art experience.[16]

MANOVA results (including group, gender, and bilingualism as factors with the three AEQ scores as dependent variables) showed no overall group differences between the two classes for reported overall artistic experience (p = .94). However, there was a significant difference for gender on this measure (p = .005), with girls scoring higher than boys on the first AEQ section (p = .005) and an overall trend (p = .078), while the girls scored lower than boys on the second section of the AEQ (n.s.). In addition, a marginal overall effect for bilingualism was found (p = .075), with bilingual children tending to score lower than their native English speaking classmates, but separate one-way ANOVAs showed that the differences were not significant for bilingualism for any of the AEQ sub-scores.

Musical experience and training. For the three AEQ sub-scores calculated to

[15]The full list of questions and scoring details for this questionnaire are available in the appendix on page 423.

[16]Details of the scoring procedure as well as descriptive statistics for these sub-scores are shown in Tables C.5 to C.7 starting on page 423 in the appendix for this chapter.

Table 5.8: Artistic Experience Questionnaire scores for children in the S1 and S2 classes

AEQ Total Percent	*N*	*M*	*SD*	Range	Possible
S1 Class	19	47.8%	11.4%	27.7 - 75.1%	0 - 100%
S1 Class: Male	8	47.1%	13.2%	31.9 - 75.1%	0 - 100%
S1 Class: Female	11	48.3%	10.5%	27.7 - 70.8%	0 - 100%
S2 Class	23	46.3%	10.5%	28.6 - 71.9%	0 - 100%
S2 Class: Male	10	39.3%	6.4%	28.6 - 47.1%	0 - 100%
S2 Class: Female	13	51.6%	10.1%	35.9 - 71.9%	0 - 100%
AEQ sub-section 1	*N*	*M*	*SD*	Range	Possible
S1 Class	19	37.0	6.6	23 - 50	10 - 50
S1 Class: Male	8	35.1	7.2	23 - 47	10 - 50
S1 Class: Female	11	38.4	6.2	25 - 50	10 - 50
S2 Class	23	36.8	7.7	24 - 49	10 - 50
S2 Class: Male	10	29.8	3.5	24 - 34	10 - 50
S2 Class: Female	13	42.2	5.1	34 - 49	10 - 50
AEQ sub-section 2	*N*	*M*	*SD*	Range	Possible
S1 Class	19	28.2	16.7	7 - 73	0 - 130
S1 Class: Male	8	31.3	21.5	10 - 73	0 - 130
S1 Class: Female	11	25.9	12.9	7 - 54	0 - 130
S2 Class	23	24.5	14.6	5 - 62	0 - 130
S2 Class: Male	10	24.8	11.8	8 - 38	0 - 130
S2 Class: Female	13	24.3	16.9	5 - 62	0 - 130

provide information about the children's musical experience, MANOVA showed no differences between the two groups (p = .95) or for bilingualism (p = .51).[17] However, there was a significant gender difference, p = .007, with the girls showing higher scores than boys, particularly for the first music sub-section, p = .002. No significant interactions were observed between group, gender, and bilingualism for the musical experience and training sub-scores.

Drama experience and training. MANOVA results comparing the three drama experience sub-scores showed no significant differences between the two groups (p = .72) or for gender (p = .29).[18] The bilingual children tended to have higher scores than the native English speakers for drama training, but the differences were not

[17] Descriptive statistics for these scores are available in Table C.5 on page 428 in the appendix.
[18] Descriptive statistics for the drama sub-scores are shown in Table C.6 on page 429 in the appendix.

significant (p = .83). A significant interaction was found between group and gender, p = .016, particularly for the first drama sub-section, p = .005. The interaction between group and gender was also significant for the overall percentage drama experience sub-score, p = .03, with girls in the S1 class scoring lower than boys whereas in the S2 class, girls had higher scores than boys.

Visual art experience and training. For the children's visual art training and experience, three-factor MANOVA with the three visual art AEQ sub-scores showed no differences between the two classroom groups (p = .98) or for gender (p = .42).[19] However, a significant difference was observed for bilingualism, p = .01, with the bilingual children scoring higher than the native English speakers, particularly for the first sub-section of the visual art AEQ questions. No significant interactions between group, gender, and bilingualism were found.

To summarise, for the Artistic Experience Questionnaire and for the sub-scores calculated to measure the children's musical, drama, and visual art experience and training, no group differences were observed between the two classes. Significant gender differences were found, particularly on the Likert-style questions of the AEQ, with girls tending to score higher than boys. A marginal overall difference in the Likert-style AEQ scores was also observed for bilingualism, reaching statistical significance for the visual art sub-score, with bilingual children tending to report more artistic training and experience than their classmates on the first AEQ section. However, since the group sizes were not equal for the bilingual children (8 pupils) and the native English speakers (34 pupils), these results should be interpreted with caution and not be generalised to other populations.

[19]Table C.7 on page 430 in the appendix shows descriptive statistics for the visual art sub-scores on the AEQ.

5.4.3 Artistic preferences

The Artistic Experience Questionnaire scores were also used to establish whether any children had a stronger background or preference for a particular art form (music, drama, or visual art). This was important because a marked difference in previous training or preferences for drama or for music could result in an aptitude-treatment interaction, as was possibly found in the results of Chapter 4.

Thus, scores on the three AEQ sub-sections were compared to identify whether the children had a much higher sub-score for music, drama, or visual art compared to their scores on the other sub-scores. For the Likert-scale items, if a child had a sub-score for one art form that was 1 point higher than their score for either of the other two art forms, this was categorised as a preference for that art form. For the second sub-section of the AEQ and for the overall AEQ scores, percentage scores were compared because the number of points possible varied for music, drama, and visual art background. If a child's percentage score was higher for one art form compared to the others by 20% or more, this was classified as a preference for the art form.[20]

On the first AEQ section, 19 children in the two classes had similar scores for each of the three art forms. Four children overall had a strong preference for Music compared to both Drama and Visual Art (three children in S1 class, one child in S2 class); one strongly preferred Drama to Music and Visual Art (in S1 class); and two children in the S2 class strongly preferred Visual Art to both Music and Drama. A few children had higher scores for Music (three children; one in S1 class) or for Drama (one child in S1 class) compared to Visual Art, and one child in each class had a higher score for Visual Art than for Drama on the first AEQ section. In addition, four children (three in S1 class) had a much lower score for Drama than for Visual Art and for Music; five children (two in S1 class) had a much lower score for Visual Art

[20]This percentage was chosen because it corresponded to the same cut-off point used for the artistic preference scores in the first AEQ section.

than for Drama or Music; and one child in the S1 class had a much lower score for Music than for Drama or Visual Art. Three-factor MANOVA showed no group or gender differences for the Artistic Preference scores in the two classes, nor was there a difference for bilingualism or any significant interactions.

Compared to the first AEQ section, percentage scores for the second AEQ section tended to be lower, but on this part, most children (23 out of 42) had roughly similar percentage scores for the three art forms. Four children had a strong preference for Music (one in S1 class) and one child in the S1 class had a strong preference for Drama; no children reported a strong preference for Visual Art, but this may in part be due to some children not completing all questions of the second part of the AEQ. Scores of seven children (four in S1 class) indicated a preference for Music compared to Visual Art; one child in the S1 class had a higher score for Drama than for Music; three children (one in S1 class) had higher Drama scores than scores for Visual Art; and three children (one in S1 class) had 20% higher scores for Music and Drama compared to Visual Art. Again, no children preferred Visual Art and no differences or interactions were found for group, gender, or bilingualism on the second AEQ section using three-factor MANOVA.

Overall AEQ percentage scores showed even fewer children with strong artistic preferences: 28 of 42 children reported similar overall percentage scores for Music, Drama, and Visual Art. One child in the S2 class had a much higher percentage score for Music than for Drama and Visual Art; five children (one in S1 class) had higher scores for Music than for Visual Art; and two children in the S1 class had higher scores for Music than for Drama. One child had a higher score for Visual Art than for Drama. Four children (two in S1 class) had higher scores for Drama and Music than for Visual Art, and one child in the S1 class had higher scores for Visual Art and Drama than for Music. However, no overall differences or interactions were observed between group, gender, or bilingualism on the Artistic Preference scores.

In summary, while some children who took part in this study showed preferences for one art form or another, these artistic preferences were fairly well balanced, with no significant differences observed for group, gender, or bilingualism in the two classroom groups. Since the music, drama, and visual art AEQ sub-scores contain greater detail than the Artistic Preference scores, the AEQ sub-scores are used in later MANCOVA analyses.

5.4.4 Bilingualism

The children's responses to the LLEQ revealed that there were six bilingual children in the S1 class, although complete data was collected for only three: Polish/English (one boy), Arabic/English (one boy), and Greek/English (one girl).[21] In the S2 class, there were also six bilingual children, with complete data collected for only five: Croatian/English (one girl), Greek/English (one boy), Polish/English (one girl), Italian/English (one girl), and Slovene/English (one girl). As previously mentioned, the eight bilingual children in the two classes have been included in the full statistical analyses, but bilingualism has also been included as a separate factor.

For the French translation pre/mid/post-tests on grammar and vocabulary, no significant differences in performance were found between the bilingual children and the native English speakers involved in this study. There were no differences observed for the French Cloze post-tests, although a marginal interaction between gender and bilingualism was found on the second Cloze post-test, with bilingual girls outperforming boys ($p = .052$).

In line with expectations, a significant difference on the LLEQ was found between the bilingual children and their native English speaking classmates, particularly for the second section and the overall percentage LLEQ score (both at the $p < .01$ level). There was also a significant interaction between bilingualism and

[21] Both of the bilingual male pupils also had special learning needs.

gender (p = .014), and a significant three-way interaction between group, gender, and bilingualism (p = .001), although these results should be treated with considerable caution due to the unequal group sizes for bilingual and native English speaking children in the two groups.[22]

Previous sections also showed that for the first section of the Artistic Experience Questionnaire, there was a marginal difference for bilingualism, with bilingual pupils reporting lower AEQ scores than their classmates, although the difference was not significant for any of the individual AEQ scores. On the total visual art sub-score of the AEQ, there was a significant main effect for bilingualism, p = .01, with bilingual children showing higher scores than their native English speaking peers.

5.4.5 Age

The mean age of participating children in the S1 class was one year younger than the children in the S2 class, with only small variations in age within each class. Table 5.9 shows more details about the children's ages, including the mean age in each group separated by gender.[23]

Table 5.9: Age of participants in the S1 and S2 classes

Group	*N*	*M*	*SD*	Range
S1 Class	19	12.0 years	0.33	11 - 13
- S1 Class: Male	8	12.0 years	0.0	12
- S1 Class: Female	11	12.0 years	0.45	11 - 13
S2 Class	23	13.0 years	0.48	12 - 14
- S2 Class: Male	10	13.0 years	0.63	12 - 14
- S2 Class: Female	13	12.9 years	0.29	12 - 13

No significant differences were found for age, although the effects of age on French learning and test performance were obscured to some extent by the fact that

[22]For more details, please refer to page 221 in section 5.4.1.

[23]The children's ages that were provided to the researcher were only years of age – not years and months – so a fine-grained analysis of the effects of age in this study was not possible.

the children in the S1 class were on average one year younger than children in the S2 class, so any differences in age would tend to correspond with differences between the two classroom groups involved in this study. For the measures of individual differences, there were no significant differences observed for age, although there was a significant three-way interaction between age, gender, and bilingualism on the second LLEQ section (p = .004), very similar to the three-way interaction found between group, gender, and bilingualism (p = .001).

5.4.6 Gender

The two pre-existing classroom groups involved in this study were fairly well balanced for gender. In the S1 class, complete data was obtained for a total of 19 children: 11 girls and 8 boys. In the S2 class, data was collected for 10 boys and 13 girls, for a total of 23 pupils.

In this study, results of the French tests showed consistent gender differences, with girls tending to outperform boys on the French translation vocabulary tests (p < .05 for the Pre-test and Mid-point test; but at post-test, p = .11) and for the Cloze post-tests (p < .05 for the first Cloze post-test and p < .01 for the second Cloze post-test). No gender differences were observed for the French translation pre/mid/post-test grammar scores.

For the measures of individual differences, no gender differences were observed for the Language Learning Experience Questionnaire, but an overall gender difference was found for the Artistic Experience Questionnaire, with girls scoring significantly higher than boys on the first AEQ section (p < .01) and a trend for the overall percentage AEQ score (p = .078). For the AEQ sub-scores, a significant gender difference was found for the Music sub-section of the AEQ at the p < .01 level, with girls again scoring higher than boys. There was also a significant interaction found between group and gender on the Drama sub-section of the AEQ, both for the overall percentage score and for the first sub-section, with girls scoring higher than boys in the S2 class while the opposite pattern was found in the S1 class. The effects of gender and the other ID measures on French performance are explored in the next section.

5.5 Influence of individual differences on French test performance

This section explores the influence of the individual differences described in the last section, both on the children's French pre-test scores and their French test performance after learning French songs and dramatic dialogues. This section has the potential to clarify which of the IDs for which data was collected in this study have the greatest influence on classroom-based foreign language learning. The individual differences under consideration are: previous language learning experience, previous artistic experience and training, preferences for music, drama, or visual art, age, gender, and bilingualism.[24]

5.5.1 MANCOVA controlling for IDs at pre-test

MANCOVA (multivariate analysis of co-variance) was conducted to investigate the relationship between the pre-existing individual differences factors and the children's French pre-test translation grammar and vocabulary scores. A MANCOVA analysis including group, gender, and bilingualism as factors, along with the two LLEQ sub-sections, two AEQ sub-sections, and age as covariates, showed that no factors were significant predictors of performance on the French translation pre-test. The best

[24]Because of the small number of bilingual participants compared to monolingual native English speakers, a separate study would be needed to more adequately investigate the role of early bilingualism on subsequent foreign language learning.

predictors showed only marginal effects: the first LLEQ section sub-score ($F(2,27) = 1.394$, $p = .051$, partial $\eta^2 = .20$, power = .58), the first AEQ section sub-score ($p = .085$, partial $\eta^2 = .17$, power = .49), and gender ($p = .094$, partial $\eta^2 = .16$, power = .47). These three factors were especially predictive of performance on the French translation grammar pre-test scores, with less of an influence for the vocabulary translation scores.

When a similar MANCOVA was run substituting the two sub-scores on the artistic experience questionnaire with two scores each for music, drama, and visual art, again none of the factors were significant predictors of French pre-test performance. The first LLEQ sub-section score was the best predictor ($F(2,23) = 3.315$, $p = .054$, partial $\eta^2 = .22$, power = .57), followed by the first visual art AEQ sub-score ($p = .069$, partial $\eta^2 = .21$, power = .53). Again, these ID factors were good predictors of French translation grammar pre-test scores, but less effective for vocabulary translation scores.

5.5.2 MANCOVA controlling for IDs at mid-point

MANCOVA was conducted for the French translation mid-point grammar and vocabulary test scores and for the first Cloze post-test, including group, gender, and bilingualism factors plus the two LLEQ and AEQ scores and age as covariates. Results showed that the score on the first LLEQ section was the best predictor of French mid-point test performance ($F(3,26) = 10.153$, $p < .001$, partial $\eta^2 = .20$, power = .99). Univariate ANCOVAs showed that the first LLEQ score was significant at the $p < .001$ level on all three French mid-point tests, with large effect sizes and power greater than .9. There was also a marginal effect found for group ($p = .09$) and for the interaction between group and gender ($p = .085$), but with low observed power. Univariate ANCOVAs showed a significant group difference for the Cloze post-tests, with the S2 class scoring higher ($p = .02$, partial $\eta^2 = .19$, power = .69).

A second MANCOVA analysis was conducted which included the two LLEQ sections, the two music AEQ sub-scores, the two drama AEQ sub-scores, the two visual art AEQ sub-scores, and age as covariates. The results were similar to the results described in the previous paragraph, again showing that LLEQ score was a significant predictor of French mid-test performance (p = .001, partial η^2 = .53, power = .98 for the first LLEQ sub-section and p = .06, partial η^2 = .28, power = .59 for the second LLEQ sub-section). No other ID factors were significant predictors, although marginal effects were observed for group (p = .058, partial η^2 = .28, power = .61), gender (p = .072, partial η^2 = .27, power = .57), as well as a significant three-way interaction between group, gender, and bilingualism (p = .04, partial η^2 = .30, power = .65), again bearing in mind the small, and unequal, sample size for bilingualism in this study.

5.5.3 MANCOVA controlling for IDs at post-test

For the French grammar and vocabulary translation post-tests and the Cloze post-tests, MANCOVA was conducted including group, gender, and bilingualism as factors and the two LLEQ and AEQ scores and age as covariates. Scores on the first LLEQ sub-section were the best predictor of French post-test performance (p = .001, partial η^2 = .48, power = .98), followed by a marginal effect for the first AEQ sub-section (p = .077, partial η^2 = .23, power = .56). The interaction between group and gender was also significant (p = .04, partial η^2 = .26, power = .65). Boys outperformed girls on the French translation grammar post-test in the S2 class while girls outperformed boys in the S1 class, whereas girls outperformed boys in both classes on the French vocabulary translation and Cloze post-tests.

The second MANCOVA analysis, including age, the two LLEQ sections, and the AEQ sub-scores for music, drama, and visual art as covariates, showed that LLEQ score was the most significant predictor of French post-test performance ($p < .001$, partial η^2 = .56, power = .99 on the first LLEQ section and marginal, p = .07, partial η^2 = .27, power = .57, on the second LLEQ section). The Likert-style AEQ subsection for drama was also a significant predictor (p = .031, partial η^2 = .33, power = .71) and a marginal effect for gender was also found (p = .09, partial η^2 = .25, power = .52).[25]

5.6 Discussion

Overall, this arts intervention study produced positive outcomes for pupils, both in terms of French learning and support of the aims of Scotland's Curriculum for Excellence in the modern language classroom. The results of the French tests in this classroom-based L2 learning study provide support for the claim that learning French songs and dramatic dialogues (over a two-week period) can be beneficial for L2 learning. In this quasi-experimental arts intervention study, French test results showed that in the S2 class, both the song and the dramatic dialogue significantly increased the children's French vocabulary and grammar translation scores over each two-week period. In the S1 class, children's vocabulary and grammar translation scores improved significantly for the dramatic dialogue (from pre-test to mid-point test). By contrast, the song post-test results for the French-to-English translation in the S1 class (from mid-point to post-test) were lower, in part due to the researcher's choice of a very difficult French song and a lack of time to learn the song during the second two-week period, due to in-class testing conducted by their French teacher.

Children's responses to the two opinion questionnaires also provided valuable information about the benefits of incorporating arts activities into the modern foreign language classroom, with many children in both classes reporting that their confidence in speaking French was improved. Based on in-class observations, in general there

[25] Separate ANCOVAs showed a significant gender effect for the second Cloze post-test, p = .01, but no effect for the French translation post-test scores.

were few opportunities for the children to speak to one another in French, and it was unusual for a child to say more than two sentences in a row. The children's responses to the opinion questionnaires showed that increasing the amount of listening and singing/speaking practice increased these beginning French learners' confidence to speak in the target language. This finding fits with reports of increased confidence in a study of ESL learners at different proficiency levels conducted by de Guerrero (1987). The effects of learning songs and dramatic dialogues on L2 speaking skills would be an interesting question to pursue in future classroom-based research.

Consistent with previous findings by Smith Salcedo (2002), the questionnaire responses also showed a higher incidence of self-reported 'din' (where the French words repeated in the child's head after class) for the song than for the dramatic dialogue (52.0% vs. 9.4%). The children in the two classes also reported an overall preference for the song (62.5%) compared to the dramatic dialogue (16.7%), although a few children enjoyed the song and the dramatic dialogue equally (6.25%).

For the measures of individual differences before the arts intervention began, ANOVA showed the two pre-existing classroom groups were not well matched for previous language learning experience, with the S2 class scoring higher on this questionnaire (particularly for the Likert-style questions). However, this may be linked to the significant age difference between the groups, which was expected because children in the S2 class were on average one year older than children in the S1 class. There were no group differences for overall artistic experience or preferences at the start of the study, but there was a significant gender difference, with girls scoring higher than boys (particularly for the Likert-style questions); this trend of girls having higher scores than boys was also found on the first music sub-score of the AEQ (both at the $p < .01$ level). Scores on the Likert-style sub-sections of the LLEQ and AEQ correlated positively and significantly with one another.[26] Before the

[26]One-tailed Pearson's $r(40) = .42$, $p = .002$. The correlation was also significant for both the S1 class and for the S2 class separately ($r(17) = .53$, $p = .01$ and $r(21) = .43$, $p = .02$, respectively). Significant positive correlations were also found for the children's scores in both classes between all of

arts intervention study began, the children's previous artistic experience, language learning experience, and gender were the only predictors of French translation grammar and vocabulary test scores, all at marginal levels (with power below the recommended .8 but large effect sizes). The Likert-style AEQ sub-score for visual art was also a marginal predictor of pre-test scores; all of these ID factors were most significant for the French grammar translation pre-test scores, rather than for vocabulary.

Similar to findings of the previous experimental chapters with adults learning Hungarian phrases, MANCOVA showed that the children's prior language learning experience had the greatest influence on their French language learning (significant at the $p < .01$ level for both the French mid-point and post-test scores, with large effect sizes and high levels of observed power). In this study gender was also a frequent predictor of French performance, with girls tending to score higher than boys (although generally only at a marginal level, albeit with large effect sizes). At post-test, the first AEQ sub-section for drama predicted French test performance, with a large effect size, at the $p < .05$ level. Significant and marginal main effects for group and interactions between group and gender were also found, with low observed power but large effect sizes. MANCOVA results showed that age, bilingualism, musical and visual art training tended to be less predictive of children's French learning.

Variations in French test performance were not shown to be due to the use of songs vs. dramatic dialogues to learn the L2 material in this study. Higher scores on the grammar and vocabulary translation test were observed for the dramatic dialogue than for the song in the S1 class, which may have arisen due to the choice of a very challenging song, insufficient time to learn the lyrics, and/or the tense classroom atmosphere because the children were taking French tests. The decrease in Cloze post-test 2 performance in the S2 class may have been due to internal factors such as change in motivation, whereas children in the S1 class may have been more motivated

the sub-sections of the Artistic Experience Questionnaire (all ps $< .01$).

to fill in the missing words for the song because they had enjoyed listening to it. More uncontrolled variables exist in the foreign language classroom than are found in an experimental study, so it is difficult to interpret the reason for these results with much confidence. However, using multiple methods to explore the educational research questions posed in this thesis can help create a more complete picture of the potential of songs and singing to support foreign language learning.

While the French test scores did not show the hypothesised pattern of greater improvements for the songs than for the dramatic dialogues in this study, clear benefits of including musical and dramatic activities in the L2 curriculum were found in the children's responses to the two opinion questionnaires. Based on in-class observations, the songs also seemed to be more fun or memorable, even though children in the S1 class also appeared to greatly enjoy the dramatic dialogue, which was set in a park, partly because there were dog noises that someone needed to act out. Also, a few boys in the S1 class started dancing to the song when it started to play, although their French teacher asked them to stop. Children in the S2 class tended to be more reserved, but in both classes children asked to hear the song again several times at the end of class (which was not allowed, due to the research design), whereas only the children in the S1 class asked to hear the dramatic dialogue again, and only once. In addition, when the completed DVDs and the results of the study were presented to both classes, the children asked to hear the songs again. Thus, this study showed practical benefits for learning a new language through songs and dramatic dialogues in the classroom.

Although the improvements in L2 grammar and vocabulary test scores were not significantly greater for the songs compared to the dramatic dialogues in this study, the finding that both songs and dramatic dialogues can greatly improve L2 grammar and vocabulary learning (in the S2 class) can give educators confidence that taking the time to teach pupils the words of an L2 song with an appropriate difficulty level can support pedagogical aims to a similar degree as dramatic dialogues. The fact that most children preferred the songs to the dialogues is a practical reason to incorporate more music into the modern languages curriculum, since songs can provide fun and motivating L2 material for pupils to learn.

Conclusion

In two modern foreign language classrooms over a longer-term, distributed practice learning procedure over a four-week period, learning new L2 French material through songs was not shown to be more effective than learning new material through dramatic dialogues, based on learners' performance on two Cloze (fill in the blank) post-tests and a French-to-English translation grammar and vocabulary pre/mid/post-test. In this classroom learning context, children's French test performance was most often mediated by previous language learning experience, group, and gender, with girls tending to score higher than boys (all with large effect sizes).

The results of two self-report opinion questionnaires indicated that the songs and dramatic dialogues that were incorporated into the French curriculum had increased many of the pupils' French listening and speaking skills, and their confidence to speak in French. The children also preferred listening to the song than to the dramatic dialogue and they reported a greater incidence of 'din' for the French words of the song than for the dialogue. Overall, these findings support the claim that songs can support L2 learning in the modern language classroom over longer periods of time, although these benefits cannot be ascribed to the possible cognitive benefits of a musical learning method.

The next chapter describes the development of a workbook and afternoon workshop designed to increase communication between researchers and teachers on the effective use of songs and singing to support foreign language learning in the classroom. This final component of the research project was aimed at bringing the research findings to key stakeholders – modern language teachers without formal musical training – while at the same time informing future research by taking into account teachers' needs and feedback about the materials that were prepared to support foreign language teaching through music-related activities and singing.

Chapter 6

Teaching Foreign Languages with Music and Songs: A Knowledge Exchange Project

This chapter discusses the design of a workbook and workshop that were designed to facilitate the exchange of ideas between modern foreign language teachers who wished to learn more about the use of songs in L2 education and provide the research community with knowledge of the teachers' practical concerns and experiences when using songs and music in the L2 curriculum. The chapter describes how the knowledge exchange workshop was implemented, the modern language teachers' reports of what would encourage them to use songs in the L2 classroom, and the challenges that teachers cited around the use of L2 songs for their pedagogical goals. It also outlines whether the modern language teachers felt that the workshop and workbook materials were valuable, based on their responses to two brief questionnaires about the material covered during the workshop. The chapter also explores whether the workshop format was suited to sharing knowledge between researchers and modern foreign language practitioners.

6.1 Background

An important part of the research process is communication between the research community and the teachers and other stakeholders who can use those research findings in their practice. The flow of practitioners' observations, ideas, and needs back to the research community can also create fruitful new research directions and approaches. Since 2005, the UK's Arts and Humanities Research Council (AHRC) has promoted projects that encourage researchers 'to disseminate and transfer their knowledge to other areas where it can make a difference' and in particular, the AHRC backs innovative, creative projects that 'support two-way KT [Knowledge Transfer] processes rather than one-way dissemination' (AHRC, 2008).

For knowledge exchange to be successful, an open atmosphere of mutual trust and respect between the different parties is essential (Levin & Cross, 2004). One concern for developers of knowledge exchange projects is that the conversation may become unbalanced if one party is perceived as the 'expert' and other participants as the 'learners,' with the result that there is less sharing of knowledge in both directions (Thomas-Hunt, Ogden, & Neale, 2003). Thus, it is important to design knowledge exchange projects which have a quality of openness, and to provide frequent opportunities for informal discussion and feedback between all participants.

With these considerations in mind, a knowledge exchange workshop was created to present current research about the potential of music and songs to support verbal skills in the native language and in new languages, and to share some of the results of the preceding chapters with a small group of modern language teachers who wanted to include more L2 songs in their educational practice. The pedagogical and motivational value of L2 songs and singing to engage learners and to support broader curricular aims, such as Scotland's Curriculum for Excellence, were also discussed. Information was sought from the teachers about their own experiences using songs in the L2 classroom through the consideration of a few discussion questions. The aims of the project were: (1) To conduct targeted workshops for an appropriate group of teachers to share new teaching materials; and (2) To produce professional-quality printed teaching materials for the workshops. The chapter also addresses whether a knowledge exchange workshop can provide a good means of sharing information between the research community and modern foreign language teachers.

6.2 Method

This section outlines the design of the workbook and workshop, as well as information about the pilot version and background information about the modern language teachers who participated. It also describes the development of questionnaires designed to obtain teachers' feedback about the workshop and workbook.

As part of a larger knowledge transfer grant awarded to Dr. Katie Overy by the College of Humanities and Social Sciences at the University of Edinburgh, funding was provided to support the creation and printing of a workbook and an afternoon workshop that would help Edinburgh-based teachers use music and songs in the modern language classroom. The workbook was designed to be useful for practitioners who had not previously used songs in the curriculum by providing activities, tips and song suggestions that could be used by modern language teachers without any formal musical training. The three-hour afternoon workshop was designed to foster the dissemination and exchange of ideas and experiences between researchers and practitioners. Both the workbook and workshop were piloted before the final versions were completed. Two post-workshop questionnaires were distributed to all participants to help determine whether the workshop and workbook had met the aims of this knowledge exchange project.

6.2.1 Design of the workbook

The workbook was developed to provide practical ideas and activities for modern language teachers who did not have formal musical training and who had not used songs or music extensively in their previous teaching. The preface and introduction sections provided background information and research findings related to the use of songs in language education. These sections also provided some practical information about how the workbook activities could be implemented in the modern language classroom.

The main body of the workbook was divided into ten activity cards, providing a sequence of lesson activities with each mini-lesson focusing on a different foreign language skill. The ten activity cards described how to prepare for the lesson and then ideas about how to use a song to teach particular L2 skills in the classroom. The mini-lesson topics included: how to introduce a song in the target language; ideas for teaching L2 vocabulary, grammar, listening comprehension, speaking, pronunciation, reading, writing, and cultural understanding through songs; and finally, some tips about how to teach learners to sing a song in the target language. Some activity cards provided variations on the activity and/or sample materials that could help teachers prepare a lesson for different groups (e.g., beginners vs. intermediate learners, primary school children vs. adolescents or adults). Some activity cards also provided additional ideas that could be used as small-group extension activities or assigned for homework.

The final section of the workbook contained a list of references and additional song-related resources for teaching particular languages. It also contained a questionnaire that teachers could distribute to learners at the beginning of the year to determine their musical background and preferences (either by photocopying the English questions or by creating a new version in the target language for more advanced learners). The final section contained a grid of song suggestions for teaching a variety of different languages, to which the teacher and/or pupils could add their own songs.

It was important that the ‘look and feel’ of the workbook be professional, yet easy to read and use. The budget for the grant included a small fee for a graphic designer to create the cover and page design and to handle the typesetting. Since the workbook activities were intended to be used and/or adapted by modern language teachers, the printed workbooks were spiral- bound to make it easy to photocopy a particular lesson for use in the classroom. Also, the paper for the final print run was sturdy card to create a durable product, and the workbooks were printed in colour to create an appealing look. The printing and binding was done by professional printers in Edinburgh. A copy of the workbook is available in the appendix.

6.2.2 Design of the workshop for modern language teachers

A knowledge exchange workshop was developed to facilitate communication between researchers and teachers with an interest in how to use songs in the foreign language classroom to facilitate learning. The workshop was designed to have a warm, friendly atmosphere in which all participants would feel comfortable to share ideas and opinions. The workshop was advertised to modern language teachers in the local Edinburgh area as an ‘opportunity for everyone to hear about current research in this area, to exchange tips and to learn from one another.’ Participants were asked if they had taught any lessons using music or songs in the past, and if so, to think about an idea or two that they could share with the group.

6.2.3 Pilot workshop with Masters degree students

Three weeks before running the workshop session for practising modern language teachers, a pilot version of the workshop was conducted with Masters degree students at the University of Edinburgh. There were 10 participants in total who attended the pilot workshop session: three students from the Music in the Community MSc programme, six from the MEd TESOL (Teaching English to Speakers of Other Languages) programme, and one from the Language Teaching MSc programme. The MSc students reported that they had between 2-14 years of teaching experience, although several attendees had never taught foreign languages because they were musicians and had therefore taught music. All of the pilot workshop participants reported that they had previously used music and songs in their teaching.

During the pilot workshop, a short discussion about three different questions was audio recorded. The discussion questions and a summary of responses to each are detailed below.[1]

1. What motivates or encourages you as a teacher to use songs in the foreign language classroom? What are some benefits of using songs to teach foreign languages?

Discussion of this question resulted in a variety of responses. Several pilot workshop attendees mentioned that they and their students thought that listening to music and songs in the L2 classroom was a lot of fun. Other responses were:

- Music is engaging
- Music builds confidence
- Music 'wakes up' lots of different areas of the brain
- Music is a good way to improve students' passion in learning the new language
- Music lowers the 'affective filter' so it is easier to learn a new language

2. What do you think the challenges or difficulties might be with trying to use songs in

[1]A full transcript is available in Appendix D.

a foreign language classroom? What are some possible ways of resolving them?

After taking a short break, this was the first question discussed. Several themes arose during the pilot workshop session:

- Lack of resources (time and money) to find suitable audio recordings. This challenge could be resolved by asking the students to find songs – subject to review by the teacher before they are used in the classroom. However, this solution might not work well with very young learners.
- Being shy and lack of confidence in singing (teachers and/or pupils). This issue could be resolved by introducing the new musical method slowly to build everyone's confidence.
- Amusicality or lack of enjoyment of music in general, or certain types of music in particular (teachers and/or learners). A proposed solution to the second part of this issue was to include a variety of music styles, or to ask students to fill in a questionnaire about their musical preferences and then to choose musical styles that the learners would be more likely to enjoy.
- Unwillingness to take part in singing or musical activities (sometimes, but not always, due to not enjoying the music). Possible solutions to this challenge included making clear to learners what the teacher's reasons for including music in the lesson are – for particular pedagogical reasons, such as to practise speaking or pronunciation skills. Another idea was to encourage learners to sub-vocalise (move their mouth in time with the words) if they were unwilling to sing aloud.

3. How you would adapt or use songs with your own students? What are some ideas of song activities?

Unfortunately there was insufficient time to discuss this question during the pilot workshop session. However, since the workshop session was fairly interactive, a few ideas were mentioned at other times during the discussion. Some of the themes were:

- Teaching vocabulary and chunks of material in the new language
- Getting students to suggest songs or artists to use in class
- Gap-fill (Cloze) activities or photocopying the lyrics sheet
- Singing an L2 song as a group
- Especially with younger learners, use a variety of music, nursery rhymes, poetry, rhythmic, whole-body activities (such as the Orff method)

At the end of the pilot workshop, a brief post-workshop questionnaire was distributed to the Masters degree students. The questionnaire was designed to evaluate the extent to which the workshop had met the aims set out at the beginning. On the whole, responses to the post-workshop questionnaire for the pilot workshop were helpful and moderately positive, with all 8 respondents[2] reporting that they felt satisfied with the workshop content (5 participants were 'Satisfied' and 3 participants were 'Very satisfied') and with the workshop materials (7 participants were 'Satisfied' and one participant was 'Very satisfied'). Responses were more mixed for the question of whether the method of combining the presentation with participatory discussion and activities was satisfactory, with five participants reporting that they were satisfied (2 were 'Very satisfied' and 3 'Satisfied'), while two participants reported they felt 'Indifferent or Unsure'.[3]

The pilot workshop provided a useful opportunity to fine-tune the timings before

[2]Three pilot workshop participants had to leave early and two of those individuals did not respond via email to the post-workshop questionnaire.

[3]One additional respondent wrote that she was 'Unsure' because she had to leave halfway through the workshop, so she was not present during the discussion/participation section.

the workshop was presented again to practising modern language teachers in Edinburgh. A few of the Masters degree students suggested that while the amount of information about the research findings was not quite enough for them, they felt it might be appropriate to further reduce the level of detail about previous research studies for the final workshop with modern language teachers. The Masters degree students also provided some additional song-related resources to include and suggested a few changes to the workbook content before the final version was printed. Thus, conducting a pilot workshop with Masters degree students provided very useful feedback and information which was used to create the final workbook and knowledge exchange workshop for modern foreign language teachers. The slides used for the final workshop are available in Appendix D.

6.2.4 Participants

Modern language teachers were recruited through an email invitation sent by a contact at the City of Edinburgh Council who worked with the Department of Schools and Families. Six modern language teachers from local Edinburgh schools took part in the afternoon workshop. The participants' language teaching experience ranged from two in-service teachers at the end of the training period to 30 years of experience, with all but two of the teachers having at least 5 years of language teaching experience. All of the modern language teachers who attended the session were female. Four teachers worked in secondary schools while two participants taught at the primary school level; one of the primary school teachers also taught evening classes for adult learners.

6.2.5 Measures

Information about the participating teachers' opinions of the workshop was gathered using two post-workshop questionnaires. The two questionnaires were created to help determine whether the workshop and workbook had met the aims of this knowledge exchange project. One questionnaire was distributed at the end of the afternoon workshop and had 14 questions, including five open- ended questions. The second questionnaire had a total of 7 questions (with two open-ended questions) and was sent by email to the participating teachers two weeks after the workshop took place.[4]

6.2.6 Procedure

The modern language teachers attended this knowledge exchange workshop during their own time, since Edinburgh schools finish at 12.30 pm on Friday afternoons. A catered lunch was served before the workshop began, from 1-1.30 pm, as a way to thank the teachers for their participation. The workshop began with introductions, followed by a warm-up activity and then a 40-minute presentation by the researcher. A small-group discussion, with time to write responses to four discussion questions that were posted on several whiteboards in the room, was followed by a full-group discussion.

After a 30-minute catered tea and coffee break, the researcher led a short demonstration of a few lesson activities and ideas that were outlined in the workbook by teaching a French song. This presentation was followed by the distribution of the workbooks and a brief discussion of the activities and lesson ideas. For the final 10 minutes of the afternoon session, the participating teachers completed a post-workshop feedback form.

Two weeks after the knowledge exchange workshop, an email was sent to all

[4]Copies of the two post-workshop questionnaires are available in Appendix D starting on page 460).

participants asking them to fill in and return the second post-workshop questionnaire, which was attached as a Word document. A PDF version of the Powerpoint slides that were presented by the researcher during the workshop session was also sent to all of the teachers.

6.2.7 Data analysis

The teachers' feedback was collected using two post-workshop questionnaires. All responses were compiled into a database. For questions 1 to 4 (and for questions 6-9 of the questionnaire distributed at the end of the workshop), each response was coded with a value between 5 ('Very Satisfied' or 'Fully') and 1 ('Very Dissatisfied' or 'Not at all'). For the six items in question 5, regarding which components of the workshop and workbook had been most useful, the total number of ticks for each item was calculated. Responses to the five open-ended questions on the first questionnaire and the two open-ended questions on the second questionnaire were also compiled. The participating teachers' responses to both questionnaire sub-sections are discussed in the next section.

6.3 Outcomes

This section describes the outcomes of the afternoon workshop, which aimed to bring together researchers and practitioners to share their knowledge and ideas regarding the use of music and songs in modern foreign language education. First, this section will provide some personal reflections on whether the workshop met this goal, followed by outlining the teachers' responses to the four discussion questions. Finally, the results of the post-workshop questionnaires are presented, along with brief comparisons with the questionnaire responses gathered after the pilot workshop with Masters degree students.

6.3.1 Reflections on the workshop

The knowledge exchange workshop for modern language teachers went smoothly. All of the participants met each other during the time set aside for lunch, and the informal introductions helped create a friendly and open atmosphere. It was helpful to have a small group of attendees, with a few teachers who already knew each other. All of the participating teachers reported that they had prior experience with songs in the modern language classroom, either through listening to songs in a new language they were learning or else in their language teaching. It was also useful that there was so much variation in terms of the participants' prior teaching experience, because it enabled a good exchange of interesting viewpoints and ideas.[5]

A shorter version of the slides about previous research findings was appropriate for the group's interests, because none of the teachers said that they particularly wanted to attend the workshop to gain in-depth information about research in the area. Indeed, on the post-workshop questionnaire, only one individual reported that the summary of research findings was 'most useful and effective,' compared to all 8 participants in the pilot study.

The full-group discussion of the teachers' responses to the four questions on the whiteboards was not given in-depth coverage, but during the tea and coffee break afterward, many of the attendees continued to talk about the discussion questions. Toward the end of the second half of the workshop session, everyone was able to share ideas about how they had used songs in the modern language classroom in the past, which was the main reason for including that discussion activity.

The demonstration of the workbook activities was conducted in French, a language which all of the teachers understood. The Mad Libs® game was new to the teachers, and they also took part in the simple 'Bonjour' warm-up activity before

[5] An audio recording of the final workshop session was not made, but the teachers' responses to the discussion questions on the whiteboards were photographed to provide evidence of what was discussed.

starting to sing ‘Les Champs-Elysées’ by Joe Dassin. In the group discussion after the demonstration of the workbook activities, one teacher observed that ‘as the student’ being taught a song in French, she felt unclear about what exactly was going to happen next and her uncertainty made the song activities more engaging. The demonstration of a few song-related activities was scheduled to take approximately 20 minutes followed by a 10-minute group discussion. This practical portion of the workshop could have been usefully extended by reducing the amount of time allotted to discussion of the research findings, because in the post-workshop questionnaire, three teachers commented that they would have appreciated more time devoted to the practical activities.

Overall, the participating teachers were very enthusiastic about the use of songs and music to support L2 education. In terms of the potential of music to support learning, the teachers frequently mentioned that people can remember songs that they learned in the new language for years, even after they have forgotten everything else. This observation could usefully link back to direct future research in psychology, where most theories of cognition hypothesise that learning the melody of the music in addition to the words would be an additional memory load during the learning process and therefore songs would not be expected to facilitate learning.

The teachers in both afternoon workshops also requested a theory or a model that would show how singing can support modern language learning, because this would help them explain to colleagues why it can be useful to incorporate songs into the L2 curriculum. An educational framework that is based on research evidence was subsequently developed and is presented in Chapter 8.

6.3.2 Results of the discussion questions

First in pairs or as individuals, the participating teachers wrote their responses to four discussion questions, which were then briefly discussed together as a group.

Photographs of the teachers' responses to these questions are available in Figures 6.1 to 6.4 and are also written in the text since the photographs were not clear enough to read all of the handwriting. Teachers' responses to the first discussion question were:

- It's fun! ✓
- Retention of vocabulary ✓
- + also reinforcement ✓
- Fulfils Curriculum for Excellence Criteria ✓
- Pupil engagement
- genuine experience of the country/culture
- When pupils are enjoying themselves they learn.
- lots of material & variety

Figure 6.1: Teachers' Responses to Question 1: What do you think the value of using songs in the modern language classroom might be for you?

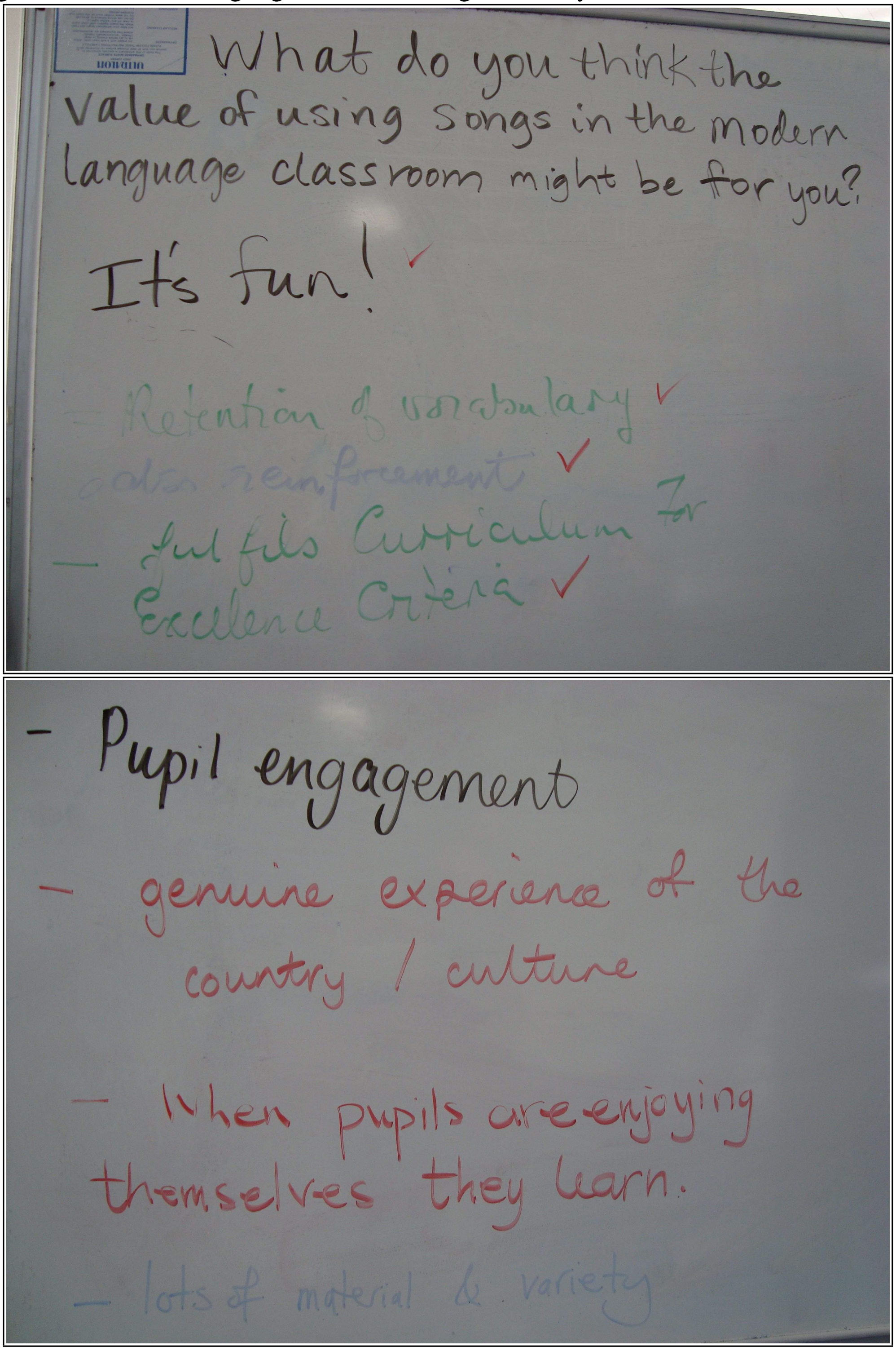
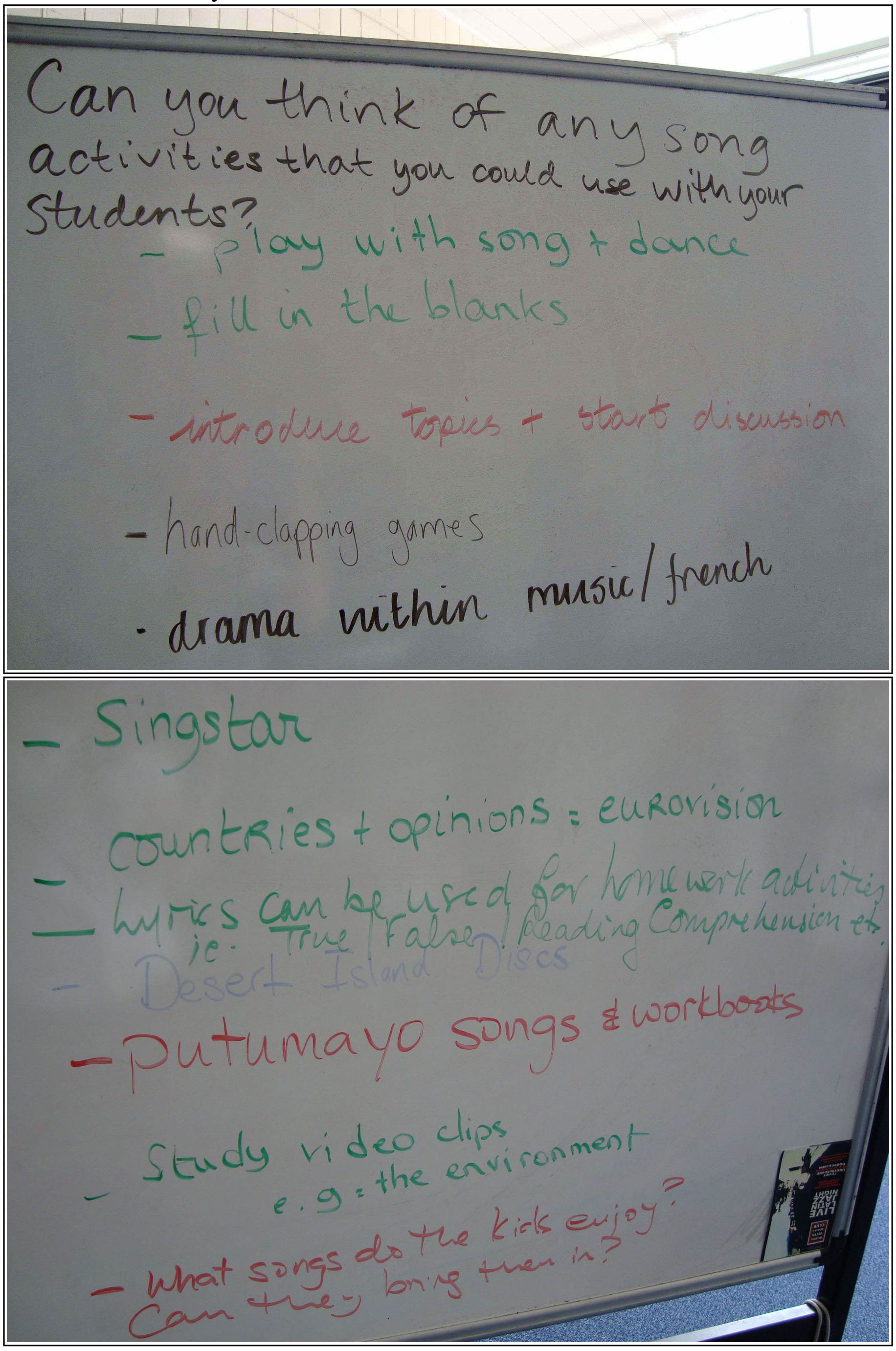


Figure 6.2: Teachers' Responses to Question 2: Can you think of any song activities that you could use with your students?

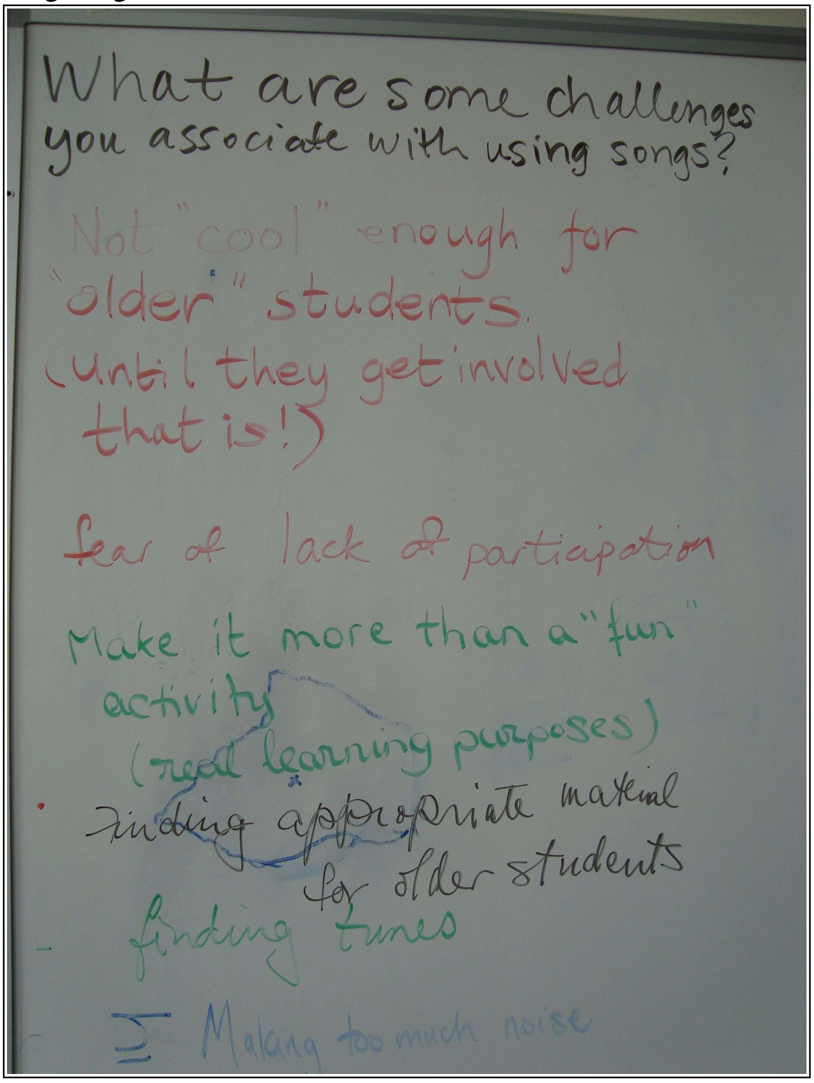


For the second discussion question, teachers' written comments were:

- play with song and dance
- fill in the blanks

- introduce topics + start discussion
- handclapping games
- drama within music/French
- Singstar
- countries + opinions = Eurovision
- Lyrics can be used for homework activities, i.e. True/False/Reading Comprehension, etc.
- Desert Island Discs.
- Putumayo songs & workbooks
- Study video clips, e.g. the environment
- What songs do the kids enjoy? Can they bring them in?

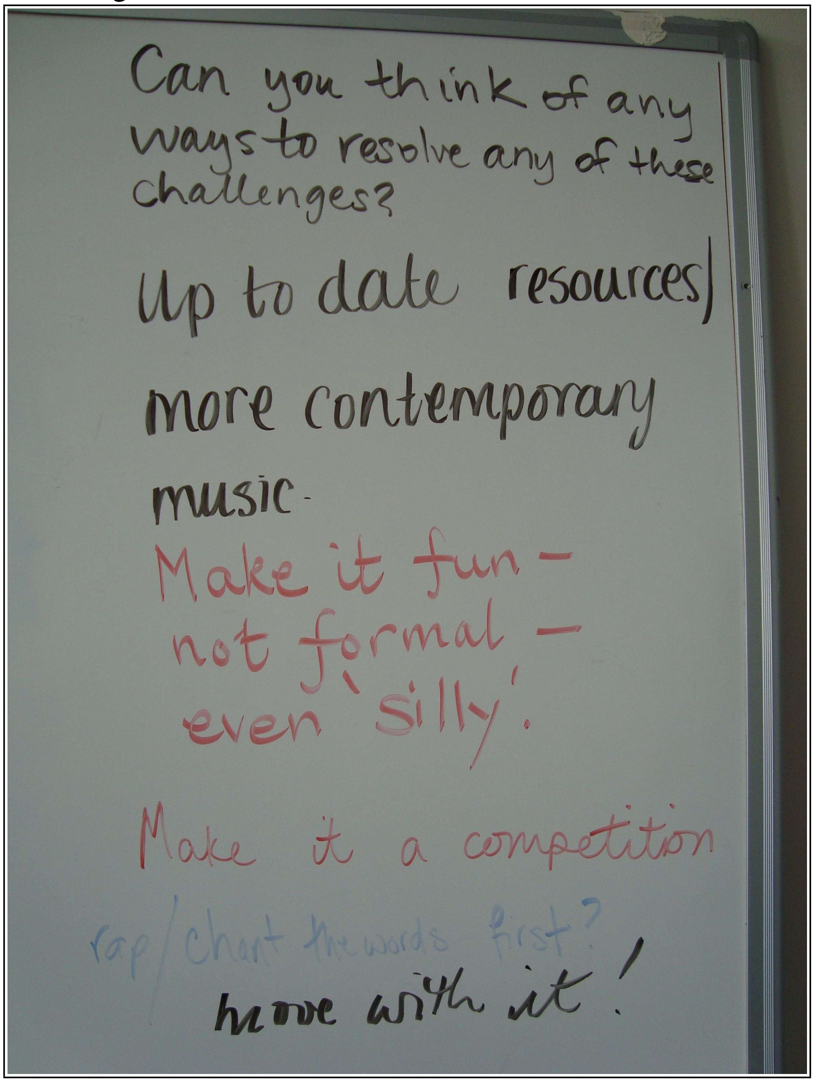
Figure 6.3: Teachers' Responses to Question 3: What are some challenges you associate with using songs?



For the third discussion question, the modern language teachers wrote:

- Not "cool" enough for "older" students (until they get involved that is!)
- fear of lack of participation
- Make it more than a "fun" activity (need learning purposes)
- Finding appropriate material for older students
- Finding tunes
- Making too much noise

Figure 6.4: Teachers' Responses to Question 4: Can you think of any ways to resolve any of these challenges?



The modern language teachers' written responses to the fourth discussion question were:

- Up to date resources
- more contemporary music
- Make it fun – not formal – even 'silly'.
- Make it a competition

- rap/chant the words first?
- move with it!

6.3.3 Teachers' feedback on the post-workshop questionnaires

At the end of the workshop, the six teachers completed a 14-item questionnaire which was described in section 6.2.5. The participants' responses to each question are outlined in turn.

1. How satisfied are you with the workshop content?

For this question, all participants in the final workshop reported that they were satisfied, with two reporting that they were 'Very satisfied' and four reporting that they were 'Satisfied,' which was similar to responses after the pilot workshop (5 were 'Satisfied' and 3 were 'Very satisfied').

2. How satisified are you with the method of combining the presentation with participatory discussion/ activities?

Responses to this question were also positive, with three teachers reporting that they felt 'Very satisfied' and three reporting that they were 'Satisfied' with the combination of presentations and participation/discussion. This was much more positive than responses after the pilot workshop, with 2 attendees 'Very satisfied,' 3 'Satisfied,' and 2 'Indifferent/Unsure.'

3. How satisfied are you with the workshop materials?

Again, responses to this question after the final workshop were even more positive than after the pilot workshop, with three teachers writing that they were 'Very satisfied' or 'Satisfied' with the workbook, compared to one individual 'Very satisfied' and 7 attendees 'Satisfied' with the materials after the pilot workshop.

4. How do you perceive the information and assistance received prior to the workshop?

Most (5 of 6) participants were 'Satisfied' or 'Very satisfied' with the amount of information and assistance about the workshop, which was similar to the pilot workshop (in both workshops, one individual wrote that they felt 'Indifferent/Unsure' for this question).

5. Which portions of the workshop session did you find most useful and effective? Please tick all that are applicable.

- Research findings: 1
- Small-group brainstorming: 5
- Group discussion of brainstorming: 3
- Sample activities: 4
- Group discussion of activities: 5
- Printed workbook materials: 1

6. How familiar are you now with:

a.) Academic research. This item had the lowest score overall, with one teacher reporting that she felt 'Fully' familiar with academic research in this area, one writing 'To a good extent,' and four writing 'Somewhat.' Nevertheless, these responses were considerably higher than the results of the pilot workshop (3 reported 'To a good extent,' 2 reported 'Somewhat,' and 3 reported 'Very little' for this item).

b.) Practical tips and lesson ideas. Four teachers responded ‘To a good extent’ and one wrote that she was ‘Somewhat’ familiar with practical tips and lesson ideas after the workshop (one individual chose not to respond to this question). These responses were more positive than those collected after the pilot workshop (5 of 8 attendees wrote ‘Somewhat’ and 3 wrote ‘To at good extent’).

c.) Challenges and how to overcome them. Again, four teachers wrote ‘To a good extent,’ while two teachers wrote ‘Somewhat’ about their familiarity with how to overcome challenges that could arise when using songs in the L2 classroom. These responses were slightly more positive than those gathered for the same item after the pilot workshop, where 4 participants wrote ‘To a good extent’ and 4 participants wrote ‘Somewhat.’

7. How much has this workshop helped increase your practical knowledge of how to teach foreign languages using music and songs?

No participants reported that they would be ‘Fully’ able to use what they had learned in the workshop for their teaching, but two wrote ‘To a good extent,’ while four wrote ‘Somewhat.’ Responses to this item were somewhat lower than those reported after the pilot workshop, where the majority (7 of 8) wrote ‘To a good extent.’

8. How much do you think you can apply what you learned from the workshop to your teaching?

No participants reported that they would be ‘Fully’ able to use what they had learned in the workshop for their teaching, but 5 wrote ‘To a good extent,’ while one wrote ‘Somewhat.’ These were similar to responses after the pilot workshop.

9. To what extent will you be able to teach your colleagues about the topic?

Again, no participants reported that they would be 'Fully' able to teach their colleagues about what they had learned in the workshop, but 5 participants wrote 'To a good extent,' while one wrote 'Somewhat.' Responses to this question were more positive than those in the pilot study, where the majority (5 of 8 participants) wrote 'Somewhat' for this item.

The six teachers' responses to the five open-ended items of the post-workshop questionnaire were varied, as outlined below.

10. How do you think you can apply what you have learned from this workshop in your classroom?

Overall, most teachers responded positively to this question. Several stated that they had gained some good ideas and that they planned to use more songs in their foreign language classes. However, one teacher wrote that it would be difficult for her to include more songs in the curriculum.

- 'Difficulty in finding time in crowded syllabus.'
- 'Will try to do more work through song.'
- 'This has given me various ideas to use.'
- 'Some good ideas, I really like the presentation on "Aux Champs-Elysees", I would definitely use that in my lessons.'
- 'I hope I'll be able to apply the ideas in my new school where the language level + motivation is greater!'
- 'This workshop reinforced my belief that music is a powerful tool in the classroom.'

11. What was one of the greatest benefits to you from this workshop?

Four of the six modern language teachers cited the exchange of practical ideas and materials as the most beneficial components of the workshop. One teacher wrote that she found it useful to meet new and old colleagues in the 'pleasant atmosphere' of the course, and another stated that she found 'learning some of the theory behind this idea' beneficial.

- 'Discussion – exchange of ideas and experiences
- 'Seeing colleagues & meeting new colleagues & music staff of University. Pleasant atmosphere. Organisation of course.'
- 'Sharing practice & materials.'
- 'Good atmosphere / good exchanges of ideas. Gave me some more motivation to use songs in the classroom.'
- 'Materials + ideas :)'
- 'Learning some of the theory behind this idea.'

12. Are there any topics that you wish had been covered in more depth?

Most of the teachers chose not to answer this open-ended question, but the two respondents had similar requests. A desire for more sample songs and for more practical presentations was expressed.

- 'More sample songs.'
- 'More practical presentations (cf "Aux Champs-Elysees"). How to adapt songs for kids who don't understand much French.'

13. Are there any additional topics that you wish the workshop had covered?

The only response to this question was from a teacher who taught in a primary school: 'More simple songs for use with Primary children.'

14. Do you have any other comments or suggestions?

Four teachers chose to respond to this question, with most writing that they had enjoyed the workshop and had gained useful resources and ideas. In addition, one attendee wrote that the slide presentation was a little too 'wordy' and thought it would be helpful if there had been more time to share ideas and resources that could be put into practice.

- 'Lots of ideas here to follow up.'
- 'Very interesting afternoon. Very good links to websites.'
- 'One of the most exciting workshop I have attended this year! Thank you.'
- 'I think that the powerpoint was a little too 'wordy'. It would have been nice to share some more practical resources.'

Only one teacher completed the questionnaire that was sent via email two weeks after the workshop was held. Her responses were very positive, as outlined below, but she reported that she had not yet used any songs in her lessons due to time pressures as the end of the academic year approached.

1. How satisfied are you with the workbook content? Very satisfied.

2. How satisfied are you with the workbook layout? Very satisfied.

3. How much did the workshop help increase your practical knowledge of how to teach foreign languages using music and songs? To a good extent.

4. Do you think the workshop format was a helpful way to learn about this topic and to share ideas with other modern language teachers? Fully.

5. Which portions of the workshop have you found to be most useful in your own *teaching practice? Please tick all that are applicable.* Small-group brainstorming, Sample activities, Group discussion of activities, Printed workbook materials.

6. How have you applied what you learned from the workbook in your classroom? If you have not tried any of the activities yet, can you tell us why? Have not taught a song yet due to pressure of finishing off course work before end of term.

7. Do you have any other comments or suggestions? Excellent atmosphere in our group. We felt fully welcomed and the catering was excellent! Thanks, again!

6.4 Discussion

The primary aim of the project described in this chapter was to bring together practitioners and researchers to discuss and share ideas about the potential of using songs to support foreign language education. To that end, an afternoon workshop and workbook were developed to enable modern language teachers without formal training in music to incorporate more songs and related musical activities into their classrooms. The project also allowed an evaluation of the extent to which the workshop format was suited to facilitating knowledge exchange between the two groups of stakeholders.

The afternoon workshops were designed to have a friendly and open atmosphere that would encourage interaction between the teachers and the researcher. Responses to the feedback questionnaires indicated that for the participants in the pilot session, and to an even greater extent for the teachers who took part in the final workshop, the participatory discussion format of the workshop was useful. On the whole, the teachers' responses were also more positive than those of the Masters degree students who took part in the pilot version of the knowledge exchange workshop. Many attendees also reported that they felt the sharing of ideas, the demonstration of activities, and the workbook materials would improve their teaching of songs in the second language classroom.

The feedback from the group of Masters degree students also differed from that of the practising teachers, with all of the students reporting that they felt that the presentation of previous research findings were among the most valuable information they had gained from the workshop compared to only one teacher reporting that the research findings were similarly useful. Nevertheless, several participants wrote that they still did not know much about the research in this area after taking part in the workshop. It should have been made clearer to workshop attendees that to date, there is not a great deal of evidence-based research about this topic.

Conclusion

Feedback on the two post-workshop questionnaires support the idea that the information and ideas gained from the knowledge exchange workshop by the participating teachers would be useful in their future teaching. Ideas also flowed from the practitioners to the researcher, providing questions and requests that can direct future research in the psychology of music and related disciplines. The interactive workshop format provided a successful means of disseminating and sharing knowledge about the benefits of using songs and singing in foreign language education.

Many of the teachers also requested a theory or model that would make clear how and why songs can provide L2 learning benefits. The next chapter outlines a series of statistical analyses that will lead to the development of a framework for future research that investigates whether and how songs and singing may support foreign language learning.

Chapter 7

Stepwise Regression with Individual Differences

In this chapter, stepwise multiple regression was conducted to explore which of the measures of individual differences collected in the three research studies had the greatest predictive effects on L2 test performance. First, the results of the two Hungarian experiments were analysed for each Hungarian language test. Stepwise regression was then conducted for the data collected in the four-week French arts intervention study. Finally, a stepwise regression analysis was conducted that had fewer ID measures, but which included participants from all three research studies.

7.1 Stepwise regression for the Hungarian studies

This first section describes the results of stepwise regression analyses using data from the two Hungarian studies. The stepwise regression using all participants in the two Hungarian experiments (140 individuals in total) provided results that were broadly similar to those found using MANCOVA to analyse the experimental results. This subsection describes the stepwise regression results for each Hungarian test, which would be similar to the results of ANCOVA calculations for each measure of

Hungarian language learning. The fifteen measures of individual differences that were used in the stepwise regression analyses were: total LEQ scores; total MEQ scores; Phonological Working Memory scores; receptive MAT Rhythm Discrimination, Pitch Discrimination, and Melody Discrimination scores; total productive MAT 'Happy Birthday' Singing and Tapping scores; Language Structure and Language Memory test scores; Positive and Negative Affect scores at the start of the experiment session; Age; Gender; and the Special Learner factor.[1]

Participants in the listen-and-repeat Speaking and Singing conditions in the two Hungarian studies were not collapsed into the same condition because there were a number of differences in the learning and testing procedures for the two Hungarian studies, and also because doing so would result in unequal group sizes.Thus, when the learning condition was included as a factor in these analyses, all of the original groups were kept separate, such that there were seven learning conditions in total. Stepwise regression analyses were also conducted without including Condition to check whether there were any differences in the models. In addition, it should be borne in mind that there was an important group difference in the second Hungarian study which was not accounted for by the ID measures – all participants in the highest-performing Listen to Speech condition completed the experiment after the exam period had ended – so including the unusual results of that group in these regression analyses may reduce the explanatory power of the resulting models; but it is nevertheless hoped that the findings may help guide future research in this area.

Stepwise regression results for the Hungarian Production Test

Including the fifteen ID measures that were shared between the two experimental studies, two predictor variables were extracted using stepwise regression analysis (both with and without including the learning condition). As shown in Table

[1]Participants in these two experimental studies who were identified as at risk of having undiagnosed dyslexia.

7.1, Model 2 shows that the first extracted factor was total LEQ score (beta = .237), followed by the Happy Birthday: Tapping test score (beta = .169). Although the observed power was decent (almost .8 for the total percentage LEQ score), the effect sizes (Cohen's f^2) were quite low[2] and together these two extracted variables only explained a cumulative 7.4% of the variance in Hungarian Production Test scores.

Table 7.1: Regression for Hungarian Production Test

Models with/without Group[a]		B	Std. Error B	Beta	t	f^2	Power
1	(Constant)	-0.187	0.972		-0.192		
	LEQ Total percent	4.798	1.628	.243**	2.947	.06	.85
2	(Constant)	-1.621	1.184		-1.368		
	LEQ Total percent	4.667	1.611	.237**	2.898	.06	.76
	Happy Birthday: Tap	0.429	0.207	.169*	2.071	.03	.43

[a] R^2 = 0.06 for Model 1 and effect size (Cohen's f^2) = .06; change in R^2 = .03 for Model 2 ($p <$ .05) and effect size = .10. Adjusted R^2 = .08 for Model 1 and adjusted R^2 = .07 for Model 2.

Stepwise regression results for the English Recall Test

The total percentage LEQ score and the MLAT-style Language Structure subtest score were the only two significant factors extracted for the English Recall Test, both with and without including Condition with the other thirteen ID factors in the stepwise regression calculation. Model 2 in Table 7.2 shows more details of the stepwise regression results. The observed power for the English Recall Test was lower (.70 for the total percentage LEQ score and .52 for the Language Structure test) than for the Hungarian Production Test, and so were the effect sizes. The two extracted variables explained a cumulative 7.5% of the variance in English Recall Test scores, which is a small amount.

[2] Cohen defines f^2 values of .02, .15, and .35 as low, medium, and large effect sizes, respectively (Cohen, 1992).

Table 7.2: Regression for English Recall Test

Models with/without Group[a]		B	Std. Error B	Beta	t	f^2	Power
1	(Constant)	3.653	1.426	*	2.561		
	LEQ Total percent	6.601	2.389	.229**	2.764	.05	.80
2	(Constant)	-0.341	2.224		-0.153		
	LEQ Total percent	6.467	2.352	.224**	2.749	.05	.70
	Language Structure	0.445	0.192	.189*	2.316	.04	.52

[a] R^2 = 0.05 for Model 1 and effect size = .05; change in R^2 = .04 for Model 2 ($p < .05$) and effect size = .10. Adjusted R^2 = .05 for Model 1 and adjusted R^2 = .08 for Model 2.

Stepwise regression results for the Hungarian Recognition Test

The total percentage LEQ score was the only significant predictor extracted for the Hungarian Recognition Test, both with and without including Condition with the other ID factors. Table 7.3 shows details for the extracted factor, total LEQ score (beta = .236). The observed power was high enough (above .8), but the effect size was again very low. The LEQ total percentage score extracted in this model explained only a cumulative 4.9% of the variance in Hungarian Recognition Test scores.

Table 7.3: Regression for Hungarian Recognition Test

Models with/without Group[a]		B	Std. Error B	Beta	t	f^2	Power
1	(Constant)	12.922	0.841	***	15.365		
	LEQ Total percent	4.012	1.408	.236**	2.849	.06	.83

[a] R^2 = 0.06 for Model 1 and effect size = .06. Adjusted R^2 = .05 for Model 1.

Stepwise regression results for the Multiple-Choice Hungarian Vocabulary Post-test

For the Hungarian vocabulary post-test, scores were only available for the first experimental study, meaning there was only a total of 60 participants rather than the 140 participants included in the other regression analyses. Results showed that the

Happy Birthday Tapping sub-test score was the first significant predictor extracted, followed by the LEQ score. Model 2 had a higher overall effect size (f^2 = .19) for the Multiple-Choice Hungarian Vocabulary Post-test compared to the other Hungarian tests, and there was also a decent level of observed power (at or above .8) for both extracted factors, as shown in Table 7.4. Model 2 explained 12.8% of the overall variation in Hungarian vocabulary post-test scores, which is higher than for the other Hungarian tests, but still low.

Table 7.4: Regression for Multiple-Choice Hungarian Vocabulary Post-test

Models with/without Group[a]		B	Std. Error B	Beta	t	f^2	Power
1	(Constant)	8.388	1.412	***	5.942		
	Happy Birthday: Tap	0.935	0.386	.303 *	2.425	.10	.97
2	(Constant)	5.129	2.071	*	2.476		
	Happy Birthday: Tap	0.892	0.375	.290*	2.379	.10	.93
	LEQ Total percent	5.839	2.780	.256*	2.100	.07	.80

[a]R^2 = 0.09 for Model 1 and effect size (Cohen's f^2) = .10; change in R^2 = .07 for Model 2 (p < .05) and effect size = .19. Adjusted R^2 = .08 for Model 1 and adjusted R^2 = .13 for Model 2.

Stepwise regression results for Delayed-Recall Hungarian Conversation

The regression results for this measure are similar to the other Hungarian tests, as shown in Table 7.5. With and without Condition included in the analysis along with the 15 ID factors, the total LEQ percentage score was the first extracted factor, followed by the productive MAT 'Happy Birthday' tapping sub-test score. The observed power for the extracted total percentage LEQ score was high enough (above .8), although lower for the other variable, and again the effect sizes were again very low for both of the extracted predictors. Model 2 explained a cumulative 11.1% of the variance in scores on the Delayed-Recall Hungarian Conversation test.

Stepwise regression results for overall raw Hungarian test score

As shown in Table 7.6, the results of the stepwise regression for the raw

Table 7.5: Regression for Delayed-Recall Hungarian Conversation

Models with/without Group[a]		B	Std. Error B	Beta	t	f^2	Power
1	(Constant)	-0.183	0.851		-0.215		
	LEQ Total percent	4.743	1.425	.273**	3.327	.08	.92
2	(Constant)	-1.924	1.038	.	-1.854		
	LEQ Total percent	4.533	1.393	.260**	3.253	.08	.86
	Happy Birthday: Tap	0.526	0.188	.224**	2.797	.05	.68

[a] R^2 = 0.07 for Model 1 and effect size (Cohen's f^2) = .08; change in R^2 = .05 for Model 2 ($p <$ .01) and effect size = .14. Adjusted R^2 = .07 for Model 1 and adjusted R^2 = .11 for Model 2.

Hungarian test performance (calculated by adding together scores on all four Hungarian tests in common between the two experiments) showed that in Model 3, three predictors were extracted (both with and without Condition included). The total LEQ score was the first factor (beta = .318), followed by the productive MAT 'Happy Birthday' Tapping test score (beta = .166), and the MLAT-style Language Memory sub-test score (beta = .166) was the third significant predictor of participants' total Hungarian test performance. The observed power for the extracted total percentage LEQ score was high (above .9) but again the power was lower for the other two variables. The effect sizes were also very low for the three extracted predictors. Model 3 explained a cumulative 14.9% of the variance in overall Hungarian test scores, which is higher than those found for the individual Hungarian tests, but still not a very high level of explanatory power.

In summary, using data from the 140 individuals who took part in the two Hungarian studies, these regression analyses were not able to explain much of the variability in scores, resulting in very low effect sizes and explanatory power. This is likely due to the high variability in performance between participants and the fact that the groups in the second Hungarian experiment were not well matched for productive musical skills or for the time they took part in the study (during or after the exam period), thus complicating the relationship between the individual differences data and Hungarian test results. The fact that the groups in the second study were not well matched casts

Table 7.6: Regression for overall raw Hungarian test score

Models with/without Group[a]		B	Std. Error B	Beta	t	f^2	Power
1	(Constant)	16.205	2.941	***	5.510		
	LEQ Total percent	20.154	4.925	.329***	4.092	.12	.98
2	(Constant)	11.133	3.616	***	3.731		
	LEQ Total percent	19.542	4.854	.319***	3.853	.11	.96
	Happy Birthday: Tap	1.532	0.655	.185*	2.423	.04	.50
3	(Constant)	9.707	3.646	**	2.663		
	LEQ Total percent	19.493	4.800	.318***	4.061	.11	.94
	Happy Birthday: Tap	1.375	0.652	.166*	2.107	.04	.43
	Language Memory	0.314	0.156	.159*	2.015	.03	.32

[a] R^2 = 0.11 for Model 1 and effect size (Cohen's f^2) = .12; change in R^2 = .03 for Model 2 ($p <$ .05) and effect size = .17; change in R^2 = .03 for Model 3 ($p < .05$) and effect size = .20. Adjusted R^2 = .10 for Model 1, R^2 = .13 for Model 2, and R^2 = .15 for Model 3.

serious doubts on the generalisability of the Hungarian test results collected in that study and, to a lesser extent, the validity of these regression models.

Nevertheless, these stepwise multiple regression analyses have indicated that previous language learning experience and productive musical skills (in particular, the ability to accurately tap along to the syllables of 'Happy Birthday') were the most consistently extracted ID factors for the Hungarian language tests.

For language learning abilities, the MLAT-style Language Structure sub-test was the second predictor extracted for the English Recall Test (after LEQ score). The Language Memory sub-test was extracted as the third predictor of overall raw Hungarian test performance (after LEQ and Happy Birthday: Tap scores). This result fits with evidence cited in Skehan (1989), that the MLAT receptive Language Structure and Language Memory sub-tests were good predictors of L2 achievement.

Since productive musical skills, and in particular the ability to tap along with the syllables of 'Happy Birthday,' appeared frequently in the results of the stepwise regression analyses using data from all 140 participants, this lends support to one explanation of the second Hungarian experiment results. The regression results again suggest that, at least in part, participants in the Singing condition of the second Hungarian study performed so poorly on the Hungarian language tests because they had significantly lower Productive MAT scores compared to the other groups. Having lower productive musical skills and being assigned to the listen-and-repeat singing condition may have also created an aptitude-treatment interaction, but these regression analyses do support the idea that individuals with higher levels of musicianship also tended to be more successful L2 learners, at least for tasks related to listening and speaking skills learned through these aural/oral methods.

Mood prior to the experiment session, phonological working memory, receptive musical skills, prior musical training and experience, age, gender, and the learning condition were not extracted as significant predictors of Hungarian test performance when using the data from all 140 participants. However, as previously mentioned, the small effect sizes and low explanatory power of the regression models and predictor variables suggest that too many independent variables or an insufficient number of participants were included in these stepwise regression analyses, and/or that other independent variables which were not measured in these studies – such as stress or fatigue related to taking exams, or IQ – may have been better overall predictors of L2 learning over this short time period. This second possibility is supported by the finding of a higher effect size for the Hungarian Multiple-Choice Vocabulary Post-test, for which data was only available for the 60 participants in the first Hungarian experiment. A considerable amount of 'noise' in the data for the second Hungarian experiment could have resulted in lower overall effect sizes and explanatory power than might have been found otherwise, if all of the groups had been better matched for IDs, as they were in the first Hungarian experiment.

Considering these results as a whole, the fact that the learning condition was not extracted as a predictor in these stepwise regression models suggests that learning L2 material through a listen-and-repeat singing method can be beneficial, particularly for speaking skills in a new language, but considerable inter-learner variability was found and therefore L2 learning through a singing method is unlikely to be the strongest overall predictor of successful L2 learning. The regression models showed that these 140 adult students' previous language learning experience and productive musical skills had the greatest influence on their L2 learning over the course of these two short experimental studies, albeit with small effect sizes and low explanatory power.

7.2 Stepwise regression for the French study

This section describes the stepwise regression analyses that were conducted using the data collected in the classroom-based French arts intervention study. The factors used in these analyses were: Group; Age; Language Learning Experience Questionnaire total percentage scores (averaging together Likert-style items in section one and open-ended questions in section two); Artistic Experience Questionnaire total percentage scores (combining together the Likert-style items and the open-ended questions about the children's music, drama, and visual art experience and preferences); Gender; and the Special Learner factor.[3]

Stepwise regression results for the French-to-English Translation Pre-test

For the Acceptable grammar translation scores on the French translation pre-test, the only significant factor extracted was Group, which was not very surprising based on the children's French pre-test scores in the two classes (with consistently lower performance in the S1 class, at least in part because children in the S2 class had already learned French for one year in secondary school and were the top set for French). When Group was removed from the stepwise regression calculation, no factors were extracted as predictors of grammar translation pre-test performance.

[3] This factor was used to classify children who: (1) were bilingual; (2) had special learning needs; or (3) were bilingual and also had special learning needs.

Table 7.7 shows the results of the stepwise regression with Group as the only predictor of performance (beta = .327); the observed power for Group was not very high (.60) and the effect size did not quite reach a medium level. In addition, the model explained only a cumulative 8.4% of the variance in overall French grammar translation pre-test scores.

Table 7.7: Regression for French Pre-test Grammar Translation ('Acceptable')

Model with Group[a]		B	Std. Error B	Beta	t	f^2	Power
1	(Constant)	-6.883	3.435	.	-2.004		
	Group	0.554	.253	.327*	2.186	.12	.60

[a]R^2 = 0.11 for Model 1 and effect size (Cohen's f^2) = .12. Adjusted R^2 = .08 for Model 1.

For the French vocabulary translation pre-test scores, the stepwise regression results extracted one factor when Group was included in the analysis, the LLEQ percentage score (beta = .374) and this model explained 11.9% of the variance. With Group included, the observed power for LLEQ score as the sole predictor did not reach a high level (.73) and the effect size was also low. Two factors were extracted when Group was not included in the calculation: in Model 2, the total LLEQ percentage score (beta = .578) was followed by a negative influence of the Special Learner factor (beta = -.383), as shown in Table 7.8. Without including Group, Model 2 explained a cumulative 20.6% of the variance in overall French vocabulary translation pre-test scores, but the effect sizes only reached a medium level and the observed power was lower than .8.

Table 7.8: Regression for French Pre-test Vocabulary Translation (words)

Model with Group[a]		**B**	**Std. Error B**	**Beta**	**t**	f^2	**Power**
1	(Constant)	-2.555	3.185		-0.802		
	LLEQ Total percent	15.253	5.975	.374*	2.553	.16	.73
Models without Group[b]		**B**	**Std. Error B**	**Beta**	**t**	f^2	**Power**
1	(Constant)	-2.555	3.185		-0.802		
	LLEQ Total percent	15.253	5.975	.374 *	2.553	.16	.73
2	(Constant)	-2.746	3.024		-0.908		
	LLEQ Total percent	23.568	6.702	.578 **	3.517	.16	.62
	Special Learner	-3.118	1.340	-.383 *	-2.327	.12	.47

[a] R^2 = .14 for Model 1 and effect size (Cohen's f^2) = .16. Adjusted R^2 = .12 for Model 1.

[b] R^2 = 0.14 for Model 1 and effect size (Cohen's f^2) = .16; change R^2 = 0.11 for Model 2 (p < .05) and effect size f^2 = .33. Adjusted R^2 = .12 for Model 1 and adjusted R^2 = .21 for Model 2.

Stepwise regression results for French Cloze Post-test 1

Regression results for the first Cloze (fill-in-the-blank) post-test, which primarily provided a measure of French vocabulary learning in this study,[4] showed that Group was the only significant factor extracted (beta = .645), which explained 40.2% of the variance and had a very large effect size and power (see Table 7.9). Since the significant result for Group likely arose because children in the S2 class consistently attempted to fill in more blanks than pupils in the S1 class did, a second regression analysis was conducted with Group removed. The second regression calculation showed that the only significant factor extracted was Age (beta = .389), with a medium effect size and power = .77. Because the children in the S2 class were on average one year older than children in the S1 class, Age was positively correlated with Group (Pearson's $r(42) = 0.76$, $p < .001$) so in effect this result indicates that the difference between the two classes explained far more of the performance differences on the first French Cloze test than the LLEQ, AEQ, Gender, or Special Learner factors did.[5]

[4]As previously mentioned, if the Cloze tests had also been administered as pre-tests, the pre-test scores would have provided a measure of the children's overall linguistic competence in French (Heilenman, 1983; Fischer, 1981; Hanzeli, 1977).

[5]This is similar to the ANCOVA results for this French measure, as discussed in Chapter 5.

Table 7.9: Regression for French Cloze Post-test 1 (‘Acceptable’)

Model with Group[a]		B	Std. Error B	Beta	t	f^2	Power
1	(Constant)	-86.261	18.215	***	-4.736		
	Group	7.174	1.344	.645***	5.339	.71	.99
Model without Group[b]		**B**	**Std. Error B**	**Beta**	**t**	f^2	**Power**
1	(Constant)	-32.072	16.145	.	-1.987		
	Age	3.434	1.288	.389 *	2.667	.18	.77

[a] $R^2 = 0.42$ for Model 1 ($p < .001$). Adjusted $R^2 = .40$ for Model 1.
[b] $R^2 = 0.15$ for Model 1 ($p < .01$). Adjusted $R^2 = .13$ for Model 1.

Stepwise regression results for the French-to-English Translation Mid-Point test

For the French grammar translation test scores at the mid-point of the arts intervention study, Group was the only significant factor extracted (see Table 7.10) with a medium effect size (Cohen’s $f^2 = .15$) although the power did not reach a high level. When Group was removed from the regression analysis for the mid-point French grammar translation score, no factors were extracted as significant predictors.

Table 7.10: Regression for French Mid-Point Grammar Translation (‘Acceptable’)

Model with Group[a]		B	Std. Error B	Beta	t	f^2	Power
1	(Constant)	-11.263	5.272	*	-2.136		
	Group	0.947	0.389	.359*	2.436	.15	.69

[a] $R^2 = 0.13$ for Model 1 ($p < .05$). Adjusted $R^2 = .11$ for Model 1.

For the French vocabulary translation mid-point test scores, the total LLEQ percentage score (beta = .672) and Special Learner (beta = -.513) were the two significant factors extracted in the final model (Group was initially added and later deleted in Model 4). Without including Group as a factor, results for the vocabulary translation mid-point test were similar (see Table 7.11 for more details). The effect sizes were at a medium level although the power did not reach .8.

Table 7.11: Regression for French Mid-Point Vocabulary Translation (words)

Models with Group[a]		B	Std. Error B	Beta	t	f^2	Power
1	(Constant)	-79.458	32.067	*	-2.478		
	Group	6.691	2.365	.408**	2.829	.20	.82
2	(Constant)	-70.009	30.988	*	-2.259		
	Group	5.312	2.351	.324*	2.260	.20	.71
	LLEQ Total percent	17.952	8.280	.311*	2.168	.10	.41
3	(Constant)	-41.261	30.655		-1.346		
	Group	3.087	2.331	.188	1.324	.20	.64
	LLEQ Total percent	33.536	9.606	.581**	3.491	.10	.34
	Special Learner	-5.049	1.868	-.437**	-2.703	.14	.46
4	(Constant)	-1.003	3.983		-0.252		
	LLEQ Total percent	38.808	8.826	.672***	4.397	.19	.69
	Special Learner	-5.923	1.765	-.513**	-3.356	.23	.78
Models without Group[b]		**B**	**Std. Error B**	**Beta**	**t**	f^2	**Power**
1	(Constant)	-0.640	4.463		-0.143		
	LLEQ Total percent	23.014	8.371	.399 *	2.749	.19	.79
2	(Constant)	-1.003	3.983		-0.252		
	LLEQ Total percent	38.808	8.826	.672 ***	4.397	.19	.69
	Special Learner	-5.923	1.765	-.513 **	-3.356	.23	.78

[a] R^2 = 0.17 for Model 1 and effect size = .20; change R^2 = 0.09 for Model 2 ($p < .05$) and effect size f^2 = .34; change R^2 = 0.12 for Model 3 ($p < .01$) and effect size = .60; change R^2 = = -.03 for Model 4 ($p > .05$) and effect size = .53. Adjusted R^2 = .15 for Model 1, .22 for Model 2, .33 for Model 3, and .31 for Model 4.

[b] R^2 = .16 for Model 1 and effect size = .19; change R^2 = .19 for Model 2 ($p < .01$) and effect size = .53. Adjusted R^2 =.14 for Model 1 and adjusted R^2 =.31 for Model 2.

Stepwise regression results for French Cloze Post-test 2

For the second Cloze post-test measuring the pupils' French vocabulary learning of the texts (song or dramatic dialogue) they had been learning in class, no factors were extracted as significant predictors of performance, either with or without including Group in the calculation.

Stepwise regression results for French-to-English Translation Post-test

Similar to the regression results for the mid-point grammar translation test, results showed that Group (beta = .557) was the only significant factor for the French grammar translation post-test, with a large effect size and power (see Table 7.12). A second regression analysis which did not include Group as a factor showed that Age (beta = .370) was the only significant predictor for the French grammar translation post-test scores (with a medium effect size and smaller level of power). Again, because of the correlation between Age and Group, this result effectively indicates that the differences between the two classroom groups was even more significant than any of the other ID factors (Gender, Special Learner, LLEQ, and AEQ scores).

Table 7.12: Regression for French Post-test Grammar Translation ('Acceptable')

Model with Group[a]		**B**	**Std. Error B**	**Beta**	**t**	**f^2**	**Power**
1	(Constant)	-23.426	5.958	***	-3.932		
	Group	1.863	0.439	.557 ***	4.239	.45	.99
Model without Group[b]		**B**	**Std. Error B**	**Beta**	**t**	**f^2**	**Power**
1	(Constant)	-10.497	4.898	*	-2.143		
	Age	0.983	0.391	.370 *	2.516	.16	.72

[a] R^2 = 0.31 for Model 1. Adjusted R^2 = .29 for Model 1.
[b] R^2 = 0.14 for Model 1 ($p < .05$) and effect size = .16. Adjusted R^2 =.12 for Model 1.

For the French vocabulary translation post-test, the second stepwise regression model extracted two factors: Group (beta = .453) and the total percentage LLEQ score (beta = .351). Without including Group as a factor, the best predictor of vocabulary translation post-test performance were the total LLEQ percentage score and a negative effect of the Special Learner factor (see Table 7.13 on the next page for details).

Stepwise regression results for average French test score

An overall French test score was calculated by averaging together the children's scores on the six French translation grammar and vocabulary pre/mid/post-tests plus the two French Cloze post-tests.

The stepwise regression results in Model 4 showed that Group was again the first

Table 7.13: Regression for French Post-test Vocabulary Translation (words)

Models with Group[a]		B	Std. Error B	Beta	t	f^2	Power
1	(Constant)	-111.474	29.776	**	-3.744		
	Group	9.105	2.196	.548 ***	4.146	.43	.99
2	(Constant)	-100.668	27.863	**	-3.613		
	Group	7.529	2.114	.453 **	3.562	.43	.97
	LLEQ Total percent	20.529	7.446	.351 **	2.757	.13	.51
Models without Group[b]		**B**	**Std. Error B**	**Beta**	**t**	f^2	**Power**
1	(Constant)	-2.360	4.344		-0.543		
	LLEQ Total percent	27.703	8.148	.474 **	3.400	.29	.93
2	(Constant)	-2.647	4.062		-0.652		
	LLEQ Total percent	40.205	9.002	.687 ***	4.466	.29	.86
	Special Learner	-4.689	1.800	-.401 *	-2.605	.13	.52

[a] R^2 = 0.30 for Model 1 and effect size (Cohen's f^2) = .43; change R^2 = 0.11 for Model 2 ($p <$ .01) and effect size f^2 = .71. Adjusted R^2 = .28 for Model 1 and adjusted R^2 = .39 for Model 2.

[b] R^2 = .22 for Model 1 and effect size = .29; change R^2 = .12 for Model 2 ($p < .05$) and effect size = .51. Adjusted R^2 =.21 for Model 1 and adjusted R^2 =.31 for Model 2.

significant predictor extracted (with a large effect size and observed power), then the total LLEQ percentage score (with a small effect size and low power), followed by a negative influence of the Special Learner factor (beta = -.459 with a medium effect size) and Age (beta = -.346 and a small effect size); full details are shown on the next page in Table 7.14.

Without including Group in the stepwise regression calculation, the best predictors of performance were LLEQ score and the Special Learner factor, with medium effect sizes and a high level of power for both predictors. Details of the stepwise regression results for average French test score without Group are also shown in Table 7.14.

Table 7.14: Regression for average French test score

Models with Group[a]		B	Std. Error B	Beta	t	f^2	Power
1	(Constant)	-49.555	13.187	**	-3.758		
	Group	4.131	0.973	.558***	4.247	.45	.99
2	(Constant)	-45.204	12.546	**	-3.603		
	Group	3.496	0.952	.472**	3.674	.45	.97
	LLEQ Total percent	8.267	3.352	.317*	2.466	.10	.42
3	(Constant)	-32.010	12.067	*	-2.653		
	Group	2.475	0.918	.334**	2.697	.45	.95
	LLEQ Total percent	15.420	3.781	.591 ***	4.078	.10	.36
	Special Learner	-2.317	0.735	-.444**	-3.151	.14	.48
4	(Constant)	-32.118	11.566	**	-2.777		
	Group	4.341	1.254	.586**	3.462	.45	.92
	LLEQ Total percent	16.284	3.648	.624***	4.464	.10	.31
	Special Learner	-2.397	0.706	-.459**	-3.395	.14	.42
	Age	-2.037	0.975	-.346*	-2.088	.05	.17
Models without Group[b]		**B**	**Std. Error B**	**Beta**	**t**	f^2	**Power**
1	(Constant)	0.451	1.971		0.229		
	LLEQ Total percent	11.599	3.697	.444 **	3.137	.25	.89
2	(Constant)	0.266	1.673		0.159		
	LLEQ Total percent	19.646	3.708	.753 ***	5.299	.25	.80
	Special Learner	-3.018	0.741	-.578 ***	-4.071	.31	.89

[a] R^2 = 0.31 for Model 1 and effect size (Cohen's f^2) = .45; change R^2 = 0.09 for Model 2 ($p <$ .05) and effect size = .68; change R^2 = .12 for Model 3 ($p < .01$) and effect size = 1.11; change R^2 = .05 for Model 4 ($p < .05$) and effect size = 1.36. Adjusted R^2 = .29 for Model 1, R^2 = .37 for Model 2, R^2 = .49 for Model 3, and R^2 = .53 for Model 4.

[b] R^2 = .20 for Model 1 and effect size = .25; change R^2 = .24 for Model 2 ($p < .001$) and effect size = .78. Adjusted R^2 =.18 for Model 1 and adjusted R^2 =.41 for Model 2.

In summary, when Group was included in the stepwise regression calculation, the Group was the most frequently extracted factor for most of the French tests. However, Group was not a good predictor of performance for the French vocabulary translation pre-test or mid-point test, or for the second Cloze post-test.

With and without Group included, the total percentage LLEQ score was the most consistently extracted factor for French vocabulary translation scores at pre-test, mid-point test, and post-test. By contrast, the LLEQ score was not a good predictor of Cloze test scores or French grammar translation scores (with or without Group).

When Group was not included, Age was extracted as a predictor of French performance on the first Cloze post-test. No factors were extracted as predictive of performance on the second Cloze post-test, with or without including Group as a factor, suggesting that the results of this test were very unusual. On this measure, children in the S1 class had low performance because they had not listened to the song as much as they had practised the dramatic dialogue, whereas the S2 class attempted to fill in fewer blanks for the dramatic dialogue (Cloze post-test 2) than they had done for the song (Cloze post-test 1). The large variation in test results led to the lack of significant predictors of performance on the second Cloze test.

When Group was excluded from the stepwise regression analysis, no predictive factors were extracted for the French grammar translation pre-test or mid-point test. Age was the only significant predictor for the French grammar translation post-test, although the highly significant positive correlation between age and group suggests that the predictive power of age on French performance was likely reflecting a large underlying group difference rather than a true effect for age. However, age did appear as a negative predictor (after Special Learner, LLEQ and Group) of average French test scores.

Without including Group, the Special Learner factor was also a consistent, but negative predictor of French vocabulary translation test scores (after the LLEQ score). By contrast, gender and the total percentage AEQ score were never extracted as significant factors, whether or not Group was included in the analysis.

While there were fewer measures of individual differences included in the French study compared to the two Hungarian studies, previous language learning experience was again a consistent, significant predictor, particularly of French vocabulary test scores. The Special Learner factor was also a significant, negative predictor of French vocabulary scores. Age (which correlated highly with Group in this study) was a significant predictor for French grammar and Cloze test 1 performance when Group was excluded from the analysis; taken with the fact that Group often appeared as a significant predictor of French test performance, this suggests that the group differences in this classroom-based study were very large.

7.3 Stepwise regression including all data from the Hungarian and French studies

To more deeply explore the relative influence of the individual differences factors on different aspects of L2 learning, an attempt was made to combine the French study results plus the data from the two Hungarian experiments into one stepwise regression analysis. Several dependent variables and measures of individual differences were combined together, creating one data set with the following measures of individual differences: (1) previous language learning experience (using LEQ or LLEQ total percentage scores); (2) prior musical experience and training (MEQ or Musical experience sub-scores, as overall percentage scores); (3) age; (4) gender; (5) special learner characteristics (bilingualism and/or special learning needs, such as dyslexic tendencies); and (6) learning condition. This resulted in a complete data set for 182 individuals on four measures of L2 learning: (1) Vocabulary production in the target language; (2) Grammar translations of phrases from the target language into English; (3) Vocabulary translations of words from the target language into English; (4) Average target language learning score. The calculation of the four L2 learning scores is described below, along with the stepwise regression results for each one.

Stepwise regression results for Vocabulary Production in the Target Language

To calculate an overall L2 vocabulary production score, the ‘number of correct words' sub-score for the Hungarian Production Test was used and for the French study, the sum of the number of 'Acceptable' words on the two French Cloze post-tests was calculated. These roughly corresponded to vocabulary scores in the target language, although the oral production required in the two Hungarian experiments was more difficult than filling in the missing words (a written task) in the French study. Since there were different possible numbers of correct words for the Hungarian Production Test (between 0 and 43 words possible) and for the French Cloze post-tests (between 0 and 63 words possible in the S1 class and between 0 and 69 words possible in the S2 class), a total percentage score was calculated.

When including the learning condition as a factor, results of the stepwise regression analysis shown in Table 7.15 revealed that the significant predictors in Model 4 were: overall LEQ (or LLEQ) score; Condition; the Special Learner factor; and Age. When the learning condition was not included, stepwise regression showed that the total percentage L(L)EQ score, Age, and the Special Learner factor were extracted. However, all of the effect sizes and observed power were low (below .05 effect sizes for the different variables and below .8 for observed power). The final regression models only explained 11% of the variance in vocabulary production scores, which is a low amount.

Table 7.15: Regression for Vocabulary Production in Target Language

Models with Group[a]		B	Std. Error B	Beta	t	f^2	Power
1	(Constant)	11.197	4.598	*	2.435		
	L(L)EQ Total percent	21.694	7.887	.201 **	2.751	.04	.79
2	(Constant)	5.079	5.146		.987		
	L(L)EQ Total percent	24.511	7.854	.227**	3.121	.04	.69
	Condition	0.493	0.196	.183*	2.510	.03	.60
3	(Constant)	2.720	5.179		.525		
	L(L)EQ Total percent	28.271	7.917	.262***	3.571	.04	.63
	Condition	0.591	0.198	.219**	2.981	.03	.53
	Special Learner	-4.254	1.803	-.174*	-2.359	.03	.46
4	(Constant)	14.659	6.970	*	2.103		
	L(L)EQ Total percent	33.637	8.088	.311***	4.159	.04	.57
	Condition	0.425	0.206	.157*	2.059	.03	.48
	Special Learner	-5.889	1.892	-.241**	-3.113	.03	.41
	Age	-0.675	0.268	-.207*	-2.514	.03	.45
Models without Group[b]		**B**	**Std. Error B**	**Beta**	**t**	f^2	**Power**
1	(Constant)	11.197	4.598	*	2.435		
	L(L)EQ Total percent	21.694	7.887	.201**	2.751	.04	.79
2	(Constant)	19.725	5.850	**	3.372		
	L(L)EQ Total percent	25.901	8.002	.240**	3.237	.04	.69
	Age	-0.558	0.241	-.171*	-2.314	.03	.53
3	(Constant)	22.721	5.820	***	3.904		
	L(L)EQ Total percent	32.319	8.136	.299***	3.972	.04	.63
	Age	-0.852	0.257	-.262**	-3.321	.03	.46
	Special Learner	-5.590	1.904	-.229**	-2.937	.04	.66

[a] R^2 = .04 for Model 1 and effect size (Cohen's f^2) = .04; change R^2 = .03 for Model 2 ($p < .05$) and effect size f^2 = .08; change R^2 = .03 for Model 3 ($p < .05$) and effect size f^2 = .11; change R^2 = .03 for Model 4 ($p < .05$) and effect size f^2 = .15. Adjusted R^2 = .04 for Model 1, R^2 = .06 for Model 2, R^2 = .09 for Model 3, R^2 = .11 for Model 4.

[b] R^2 = .04 for Model 1 and effect size = .04; change R^2 = .03 for Model 2 ($p < .05$) and effect size = .07; change R^2 = .04 for Model 3 ($p < .01$) and effect size = .12. Adjusted R^2 =.04 for Model 1, R^2 = .06 for Model 2, and R^2 = .10 for Model 3.

Stepwise regression results for Grammar Translations of Phrases from the Target Language into English

This regression analysis used the English Recall score (perfect translation) in the Hungarian experiments and for the French study, the French-to-English grammar translation (acceptable) post-test score was used. Since a total of 0-20 points were possible for the English Recall Test compared vs. 0-10 points possible on the French grammar translation post-test, the raw scores were converted to percentages.

Stepwise regression results showed that the extracted factors were: Age, total percentage L(L)EQ score, and the Special Learner factor (with and without including the learning condition). Age had a high level of observed power (.99) and a medium effect size (.16), but these values were lower for the other two variables (see Table 7.16). Overall, Model 3 explained 20.7% of the variance in the data.

Table 7.16: Regression for Grammar Translation of Phrases from L2 into English

Models with/without Group[a]		**B**	**Std. Error B**	**Beta**	**t**	f^2	**Power**
1	(Constant)	0.379	6.158		0.062		
	Age	1.670	0.307	.376***	5.438	.16	.99
2	(Constant)	-14.943	7.425	*	-2.012		
	Age	1.428	0.306	.321***	4.666	.16	.99
	L(L)EQ Total percent	35.318	10.156	.239**	3.477	.06	.83
3	(Constant)	-11.826	7.445		-1.589		
	Age	1.123	0.328	.252**	3.421	.16	.99
	L(L)EQ Total percent	41.994	10.408	.285***	4.035	.06	.77
	Special Learner	-5.815	2.435	-.174*	-2.388	.03	.41

[a] R^2 = .14 for Model 1 and effect size = .16; change R^2 = .05 for Model 2 (p < .05) and effect size = .24; change R^2 = .03 for Model 3 (p < .01) and effect size = .28. Adjusted R^2 =.14 for Model 1, R^2 = .19 for Model 2, and R^2 = .21 for Model 3.

Stepwise regression results for Vocabulary Translations of Words from the Target Language into English

For the two Hungarian experiments, this vocabulary translation score was based on one of the English Recall Test sub-scores (the number of words correctly translated into English, out of 0-70 words possible) and for the French study, the French-to-English vocabulary translation (acceptable) post-test score (out of approximately 0-53 words possible in both classes).[6]

Results of the stepwise regression analysis for the L2 vocabulary translation scores extracted two predictors, with high levels of observed power: Age and total LEQ score (both with and without including Condition). Model 2 explained 23.1% of the variance and the effect size for age was at a medium level (see Table 7.17).

Table 7.17: Regression for Vocabulary Translations of Words from L2 into English

Models with/without Group[a]		B	Std. Error B	Beta	t	f^2	Power
1	(Constant)	3.288	5.748		0.572		
	Age	1.726	0.287	.409***	6.020	.20	.99
2	(Constant)	-13.474	6.842	*	-1.969		
	Age	1.461	0.282	.347***	5.181	.20	.99
	L(L)EQ Total percent	38.640	9.359	.276***	4.129	.08	.93

[a] R^2 = .17 for Model 1 and effect size = .20; change R^2 = .07 for Model 2 ($p < .001$) and effect size = .32. Adjusted R^2 =.16 for Model 1 and adjusted R^2 = .23 for Model 2.

Stepwise Regression Results for Average Target Language Learning Score

For the final stepwise regression analysis, an average L2 learning score was calculated. The percentage scores for (1) L2 Vocabulary Production, (2) Grammar Translations of Phrases into English, and (3) Vocabulary Translations of Words into English were averaged together and this score was used in the calculation.

Results of the stepwise regression analysis showed that the significant extracted factors were: total percentage LEQ score and a negative influence of the Special Learner factor (with and without including the learning condition). The observed power was high for both factors, but the effect sizes did not reach a medium level for

[6] There was not a strict ‘correct’ number of words possible for the English translations in the French study because some variation in English word choice was allowed, as long as the translated word(s) were correct (e.g., the French word ‘surtout’ could be translated into English as ‘especially’ or ‘above all’).

either predictor separately. Model 2 explained a cumulative 15.8% of the variance in average L2 test scores for the 182 individuals whose data was included.

Table 7.18: Regression for average target language learning score

Models with/without Group[a]		B	Std. Error B	Beta	t	f^2	Power
1	(Constant)	8.980	4.898	.	1.833		
	L(L)EQ Total percent	39.145	8.401	.328***	4.659	.12	.99
2	(Constant)	7.189	4.771		1.507		
	L(L)EQ Total percent	44.179	8.258	.370***	5.349	.12	.99
	Special Learner	-6.696	1.869	-.248***	-3.583	.06	.87

[a] R^2 = .11 for Model 1 and effect size = .12; change R^2 = .06 for Model 2 (p < .001) and effect size = .20. Adjusted R^2 =.10 for Model 1 and adjusted R^2 = .16 for Model 2.

In summary, stepwise regression results using data from all three studies showed that age, previous language learning experience, and the Special Learner factor were consistently extracted predictors of these four scores of L2 learning. The learning condition was also a significant predictor, but only for vocabulary production in the target language, not for the English translations or for average L2 performance. Gender and previous musical training and experience did not appear in the regression results as significant predictors of L2 scores when using data from all 182 participants.

7.4 Discussion

Stepwise regression analyses showed a significant positive relationship between L2 test scores and the amount of previous language learning experience, which explained an average 11.3% of the variability in L2 test scores across the three studies (with small to medium effect sizes). When including data from all three research studies, the LEQ score was the first factor predicting L2 vocabulary production scores, both with and without the learning condition as a factor. The LEQ score was also the best predictor of average L2 scores. The LEQ score was the second factor extracted (after age) for the grammar and vocabulary translations from the target language into English.

The Special Learner factor was often extracted in the stepwise regression analyses after the measure of previous language learning experience, with a negative influence on L2 scores, which is perhaps unsurprising since this factor identified participants who had learning difficulties such as dyslexia. Across the three research studies, the Special Learner factor appeared as the second or third predictor of L2 production, grammar translation scores, and overall L2 learning, although it was not a significant factor for the L2-to-English vocabulary translation scores.

Age also frequently appeared as either the first or second extracted factor in the regression analyses for grammar and vocabulary translations of L2 phrases and words into English when including data from all three studies, with medium effect sizes and a high level of power (both when condition was and was not included as a factor). Age was also a predictor of L2 vocabulary production, although with a much lower effect size and observed power. It is likely that age was often extracted as a significant predictor of L2 test scores because children in the S1 class tended to have much lower mean percentage scores than participants in the other groups.

The learning condition was extracted as a significant factor (after the L(L)EQ percentage score) for the vocabulary L2 production score. This significant result for condition likely arose because children in the S1 class had low scores for the French Cloze post-test on which this percentage score was based, especially compared to performance in the listen-and-repeat Singing condition in the first Hungarian study and the Listen to Speech condition in the second Hungarian study. However, the effect size and observed power of the learning condition were again quite low.

Gender was not extracted as a significant predictor in the stepwise regression analyses that used data from all 182 individuals who took part in these three research studies, but female participants did tend to score higher than male participants on the L2 tests, especially for the vocabulary translations of L2 words into English.

The stepwise regression results with all 182 participants generally showed that self-reported musical experience (based on questionnaire scores) was not a significant factor. This is similar to the stepwise regression results that combined data from the two Hungarian studies, where MEQ score was not extracted as a significant factor. Whereas MEQ scores were not good predictors of participants' L2 test scores, the test of productive musical skills (especially syllable-tapping skills) was often extracted as a significant factor in the regression analyses in the two Hungarian studies, albeit with small effect sizes. Unfortunately it was not possible to measure musical skills in the classroom-based French study, so it was impossible to use the data from the three research studies to determine whether (productive) musical skills might also be an important predictor of adolescents' success in second language learning. However, there is some support for this possibility based on explorations of the relationship between second language pronunciation and musical skills with slightly younger children (Milovanov et al., 2008, 2009).

Table 7.19 shows the most often extracted predictors from the stepwise regression analyses, as described in this chapter. Despite generally having sufficient power and decent correspondences between the observed R^2 and adjusted R^2 values (relating to the generalisability of the sample's results to the population), the small effect sizes for the regression models suggest that there was excessive 'noise' in the data and that other factors which were not measured are important factors in L2 learning. The results of the stepwise regression analyses presented in this chapter do not indicate that singing is one of the most supportive factors that can predict overall L2 learning, but the evidence does suggest that singing can be beneficial for productive, speaking skills in a new language (although with small effect sizes and low explanatory power). The effects of L2 learning through singing vs. speaking, perhaps including IQ and motivation in place of less relevant factors, could provide useful directions for future research.

Table 7.19: Stepwise regression: Three most frequently extracted ID factors

Data set	*N*	Predictor 1	Predictor 2	Predictor 3
Hungarian studies	140	LEQ Total percent	Happy Birthday: Tapping	Language Structure and Memory
French study	42	Group (and Age)	LLEQ Total percent	Special Learner
All three studies	182	Age	L(L)EQ Total %	Special Learner

Conclusion

This chapter presented the results of stepwise regression analyses that explored the effects of individual differences between learners on the second language test results in the three empirical studies. The stepwise regression analyses showed that the most frequently extracted predictors of L2 learning were prior language learning experience, productive musical skills, age, and the special learner factor. After L(L)EQ score, the learning condition was a significant overall predictor of productive L2 skills when including data from all 182 participants in the three research studies (albeit with a small effect size and observed power below the recommended .8).

The final chapter summarises the research findings and outlines implications for educational practice arising from these research studies that explored the use of songs to support L2 learning. It also suggests future directions for research into the role of singing in modern foreign language learning. These stepwise regression results and previous research findings form the basis of an educational framework that may aid researchers in future explorations of how singing can support L2 learning.

Chapter 8

Conclusion

This chapter offers a final summary of the outcomes of the three research studies and the knowledge exchange project. It also describes the main contributions and some potential applications of these research findings, and offers a framework that can help direct future researchers investigating the effects of listening to songs and singing for modern foreign language learning and for the L2 curriculum. The chapter also highlights some important unresolved questions that are worthy of further research.

8.1 Summary of research findings

The preceding chapters have outlined a research project that explored whether listening to songs and singing can be an effective means of learning and teaching material in a new language, compared to spoken presentation methods. The research studies also investigated the extent to which a number of individual learner differences influenced the results of L2 learning through these aural/oral learning procedures and the practicalities of using songs in the modern foreign language classroom. Table 8.1 offers a brief summary of the findings of the three empirical studies conducted for this research project.

Table 8.1: Summary of results in the three empirical studies

Study	**Hungarian Study 1**	**Hungarian Study 2**	**French Study**
Greater benefit for singing?	Yes	No	No
Groups matched for pre-existing ID factors?	Yes	No (Productive MAT)	No (LLEQ, Age)
Summary of results	Benefit greatest for two verbatim, spoken Hungarian tests, $p < .05$.	Passive Listen to Speech group had highest scores, but MANCOVA n.s.	Song and drama equal L2 gains in S2 class; S1 class did worse.
Predictive ID factors	LEQ, positive mood	LEQ, hand, negative mood, motivation	Age and LLEQ (S2 > S1 class)
Gender differences?	Marginal F > M for Hungarian tests	Only for ID factors	F > M for French vocab. translation

A practical, classroom-based study (Chapter 5) showed that incorporating songs and dramatic dialogues into the French curriculum for 15 minutes per class session over a four-week period provided a number of learning benefits. For the adolescent learners in the S2 class, who were at the beginning-intermediate levels in French, both songs and dramatic dialogues provided significant gains in French vocabulary and grammar learning over the four-week period. Children in the S1 class (beginning-level French) showed a significant increase in both vocabulary and grammar translation scores for the dramatic dialogue, but due to external factors their performance on the post-tests decreased for the song. While a larger benefit on French learning for the songs compared to the dramatic dialogues was not observed, the questionnaire results showed that most of the participating children preferred listening to the songs and that the words from the songs were more likely to repeat in their heads after class than words from the dialogues did. Many pupils also reported that they thought the new activities had improved their overall French listening comprehension, speaking, and pronunciation skills, and that their confidence to speak in French had increased as a result of learning the songs and dramatic dialogues. This study also showed that the children's previous language learning, age, and gender had a strong influence on their French learning.

Chapter 6 described the development of a workbook created to help modern foreign language teachers without extensive musical training to use songs more effectively in the classroom to support L2 learning. Ten activities were created as a starting point for teaching different L2 skills in the classroom. An afternoon knowledge exchange workshop was also developed, which successfully enabled researchers and practitioners to exchange ideas that could inform their future work.

In Chapter 4, a controlled, randomised experimental study was conducted to investigate the effects on second language learning of an active ('listen-and-repeat') vs. a passive ('listen-and-learn') learning procedure using spoken or sung phrases in the new language. Results of this study revealed that there were significant pre-existing group differences on a measure of participants' productive musical skills (singing and tapping along to the syllables of 'Happy Birthday'), with the listen-and-repeat Singing condition showing the lowest scores. The passive Listen to Speech group unexpectedly had the highest performance on the Hungarian language tests, but the difference was no longer significant when the individual differences factors were included as covariates in the statistical analyses. For this experiment the variation in Hungarian test scores was so great that the observed power using MANCOVA was often inadequate to detect an effect, although large effect sizes were found for some of the ID factors. MANCOVA results indicated that the group differences on the IDs measures in this study had a substantial influence on participants' Hungarian test performance. In particular, scores on the Language Experience Questionnaire and the 'Happy Birthday' tapping sub-test were predictive of Hungarian test performance. Participants' feelings of success learning the English phrases and their change in motivation were also good predictors of Hungarian test scores. In addition, this study raised the possibility of an aptitude-treatment interaction because participants who were randomly assigned to the listen-and-repeat Singing condition had the lowest productive music scores (significantly lower than the other three groups) and had much lower Hungarian test scores than expected. The fact that the Singing group did score highly on the Delayed-Recall Hungarian Conversation (despite low productive music test scores) would suggest that L2 phrase-learning benefits for a singing method might appear more strongly for longer-term recall, rather than an immediate memory benefit.

By contrast, in Chapter 3 the participants were well matched on the measures of individual differences in the three learning conditions.[1] This aural/oral learning study compared the relative effects of an active learning procedure that involved listening to and repeating paired-associate English-Hungarian phrases that were either spoken, rhythmically spoken, or sung. Results showed that the Singing condition tended to have higher Hungarian test performance than the other two groups, particularly for the spoken, verbatim Hungarian tests (at the $p < .05$ level, with a large effect size). MANCOVA results indicated that the significant learning advantage observed for the Singing condition on the verbatim, spoken Hungarian tests was likely due to the beneficial effects of the 'listen-and-repeat' singing procedure during the learning phase, and not due to pre-existing differences between participants in the three groups. MANCOVA for this study also showed that in addition to the learning condition, several ID factors were predictive of Hungarian test performance (with large effect sizes), especially previous language learning experience (based on LEQ scores) and positive mood at the start of the experiment; for the two verbatim, spoken Hungarian tests, phonological working memory (measured with a nonword repetition test) and musical pitch discrimination test scores were also significant at the $p < .05$ level, with large effect sizes despite low observed power.

Finally, Chapter 7 presented stepwise regression analyses exploring the effects of a number of individual differences factors on second language learning outcomes in the three research studies. Not surprisingly, previous language learning experience

[1]There was a marginal difference on the MLAT-style Language Memory test, with the Singing condition showing the highest performance, but this measure was not a significant predictor of performance on the Hungarian language tests.

was the best predictor of Hungarian test performance in the two experimental studies, and similarly, children who reported greater language learning experience tended to score higher on the French language tests than their peers with lower scores on the language learning experience questionnaire. Other significant factors predicting L2 test performance in the two Hungarian studies were the participant's productive musical skills (especially syllable-tapping), MLAT-style Language Structure and Language Memory test scores.[2] Age was another significant predictor when including data from all 182 participants, with the younger children scoring lower than the adults (although this age difference may have been due to differences in previous language learning experience). Nevertheless, in general the regression models did not provide a great deal of explanatory power, suggesting that there was a lot of 'noise' in the data and/or that other ID factors that were not measured across these three empirical studies – such as intelligence or motivation – were important.

In line with previous findings, gender differences were observed for several of the measures of individual differences used in these research studies, plus marginal effects for gender on some tests of L2 learning. In general, differences were in the expected direction, with female participants tending to show higher performance on language tasks compared to male participants.

The overall picture which develops from this research project is that singing and listening to songs can provide an effective and enjoyable way of learning and teaching material in a new language, enhancing verbal memory and L2 skills through cognitive and/or affective enhancements when using a musical presentation method. This research project provides the first controlled experimental evidence that a singing 'listen-and-repeat' learning procedure can significantly increase productive speaking skills in an unfamiliar language, with a large effect size, when the groups are well matched for individual differences factors (especially previous language learning and

[2]Self-perceived success learning the English phrases and change in motivation were also significant using MANCOVA in Chapter 4, but data on motivation was only collected in the second experiment and therefore these factors were not included in the regression analyses.

productive musical skills).

8.2 Limitations of the research findings

This project was conducted by an individual researcher operating under time and funding constraints. The two experimental studies and the corresponding pilot studies used self-selecting adult participants who were randomly assigned to a learning condition. Unfortunately in the second experiment, new participants had to be recruited for the Listen to Speech condition due to technical problems during data collection for the first twenty individuals who took part (missing the Hungarian Production Test). Unexpectedly, the second group of 20 participants in the Listen to Speech condition also had the highest performance on the Hungarian tests, perhaps in part because they did not complete the study while they were revising for exams. Thus, it was not possible to determine whether active (listen-and-repeat) or passive (listening only) learning would show greater improvements in Hungarian test performance. While the learning procedure for the two experimental studies was modelled on those used in previous studies that explored the effects of music on verbal memory in the native language (Wallace, 1994; Racette & Peretz, 2007), modifications to the learning and testing procedures were made in order to investigate the effects of listening to songs and singing on learning L2 material (because learners did not already know the meaning of the phrases they heard). The learning procedure involved learning short, paired-associate phrases in English and an unfamiliar language, rather than listening to entire songs with rhyme and meter (which would have been possible in the native language), but the duration and rate of presentation for the English and Hungarian stimuli were carefully controlled for, in line with Kilgour et al. (2000).

For the French study, a convenient sample of two pre-existing classroom groups was chosen rather than randomly assigning a sample of children to different learning conditions. While every attempt was made to control as many factors as possible without sacrificing the ecological validity of conducting a classroom-based study, the learning procedures were not identical in the two classroom groups for each of the two-week periods. The curriculum required the regular classroom teacher to complete French testing with the classroom group that was learning a song during the second two-week period of the arts intervention (S1 class), and this testing had a detrimental effect by increasing children's anxiety during class and also decreased the total amount of time the children spent learning the song. In addition, it was not possible to gather data about the children's IQ, language, or musical abilities before the arts intervention study began, which would have provided valuable information about the role of these ID factors on L2 learning with adolescents over a longer time period than was observed in the experimental studies with adult learners.

For all three research studies, it was necessary to use and adapt tests which had been used previously with relatively large groups of participants, but which were not standardised. In all other respects, the research studies were conducted using rigorous, controlled procedures and every attempt was made to keep the learning experience as similar as possible for all participants. Thus, it is hoped that these limitations do not strongly affect the findings in the S2 class (which showed large increases in French grammar and vocabulary translation test performance for both the song and dramatic dialogue) and in the first experimental study (where a benefit was observed in the listen-and-repeat Singing condition for saying the phrases in Hungarian).

8.3 Relationship to existing literature

Several questions were raised at the start of this thesis, some of which have been addressed, while others will require additional research. Previous work has shown a benefit for receptive L2 vocabulary learned through songs (e.g., Medina, 1993), but the present studies show that several other L2 skills can also be improved through listening to songs and singing: listening, speaking, translation, grammar and vocabulary.

The suggestion that music and songs may be particularly supportive for the beginning stages of language learning, as put forth by Schön et al. (2008) in their statistical learning study of word-boundary learning, may be true. The present evidence suggests that the memory benefit found for a sung presentation of verbal material can be harnessed for learning phrases and their meanings – rather than only learning word boundaries in a continuous stream of nonsense speech sounds – and this could be very helpful for building up vocabulary and grammatical knowledge at the start of second language learning.

Another question is whether any memory benefits might be observed after learning verbal material through a sung presentation several times, rather than immediately after the initial exposure; in other words, is longer-term, verbatim verbal memory more supported by L2 learning through singing than through speaking? In these three research studies, participants' verbal memory was not tested immediately after hearing the L2 stimuli. This decision was made because experimental previous work has not shown a benefit for verbal memory (in the native language) after one presentation of a song, or for speed of initial learning through music compared to speech (Calvert & Tart, 1993; McElhinney & Annett, 1996; Rainey & Larsen, 2002; Wallace, 1994). However, the fact that a verbal memory benefit was observed in the first experimental study for the listen-and-repeat Singing condition after three presentations of the L2 material (particularly for oral production), both immediately after the learning period and after a short delay, suggests that the benefits observed in the native language for musical features to better support and sustain verbal learning and memory over time (Tillmann & Dowling, 2007) can also occur for verbal material in an unfamiliar foreign language.

The experiments by Tillmann and Dowling (2007) suggest that the rhythm and rhyme structure of poetry (and similar features in music) are important at the initial encoding stage, perhaps through chunking that helps to bind different features of the stimuli into a coherent whole, so that memory does not decrease as precipitously over time (although their study does not distinguish which of these factors contributes most). In the experimental studies conducted for this project, the sung phrases contained both rhythm and melody, but there were no higher-level rhythmic or rhyme structures such as would be found in poetry or complete songs (e.g., AABB or ABAB rhyme structure with equal numbers of syllables in each line). Thus, the current evidence suggests that the larger-scale structure of a musically presented verbal stimulus is not the only contributing factor to the memory benefit for singing.

The answer to whether musical mnemonic devices might only enhance surface-level processing and memory at the expense of deeper, semantic processing appears to be negative. Grammar improved as much for singing as for speech (both in the first Hungarian experiment, based on the multiple-choice vocabulary post-test scores, and in the S2 class in the French study). While in the two experimental studies it appears that the sung stimuli were not as good for surface-level detail as the spoken stimuli were, the lower scores on the Hungarian Recognition Test for the sung stimuli groups could be due to having heard the sung stimuli while learning, but the spoken stimuli for that test (with a greater memory load than for participants who heard spoken or rhythmic stimuli during the learning process).

The question of whether active L2 learning is better than passive learning (Iwata, 2005; Hannaford, 2005), or if concentrated observation without behavioural rehearsal is sufficient (Colvin Clark & Mayer, 2008; Michael, 2006) could not be answered because the groups in the second experiment were not well matched. However, it is possible that one reason the S1 class had lower performance on the grammar and vocabulary translation post-test scores for the song was because they spent less time singing the words during class (in addition to less time hearing the song). Answering the question of whether songs provide only active or also passive learning benefits would be very valuable for practitioners, especially since some learners (and teachers) may not feel comfortable singing in the classroom. If there are similar benefits of listening to L2 songs as for singing them, this could encourage more use of songs in the L2 classroom.

The French study also showed that incorporating authentic L2 songs into the curriculum can also set a positive classroom tone and may provide affective or motivational benefits for learning. Consistent with findings for adults who were learning Spanish (Smith Salcedo, 2002), the songs used in the French study were also more effective for increasing mental rehearsal ('din') than the dramatic dialogues, which may have encouraged the children to practise French after class. The role of song-related 'din' on L2 learning would be a useful question to answer in future experimental and practical work, since involuntary mental repetition of words in the new language might be one mechanism through which a benefit for musical L2 learning can appear.

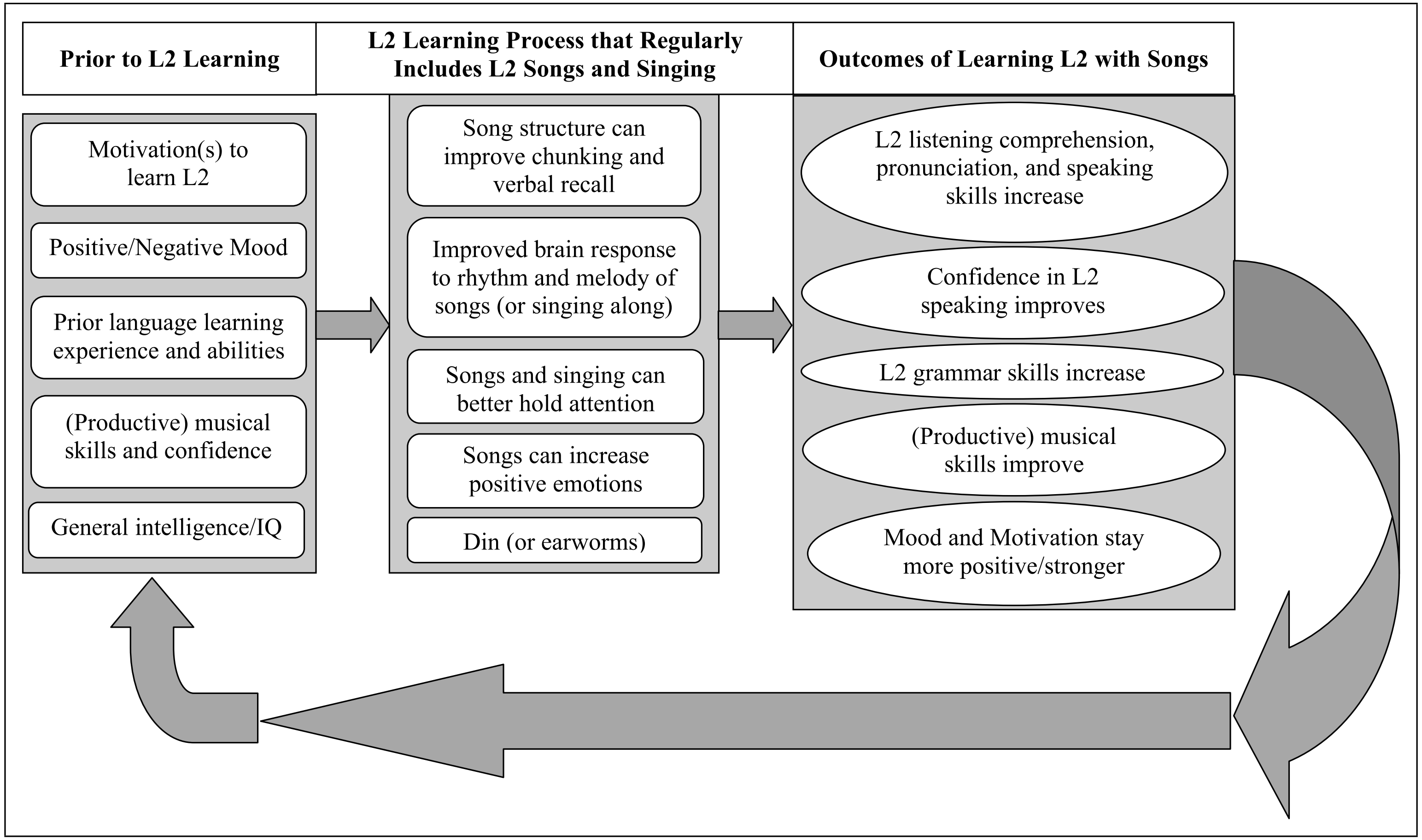
This research project also showed that several individual differences had a large influence on L2 learning, particularly participants' previous language learning experience, mood at the start of the experiment session (see also Schellenberg, 2006), and productive musical skills. Similarly, Gilleece (2006) found that children's productive rhythm skills were correlated with second language learning ability, even after controlling for the effects of IQ. Gilleece concludes that her research study '...highlights the fact that a link between music and language aptitude does indeed exist. It is not simply training in music which has an effect on language development but even basic aptitude in music may be helpful for the language learner' (p. 243).

Taken together, the results of these empirical studies do support the idea that learning a new language through music can support modern language teachers' goals for second language education.

8.4 Framework for future research

The stepwise regression results presented in the last chapter, along with findings from the existing literature in this area, have formed the basis of a framework that can help direct future research. Each learner brings his or her language learning experience (in the native language(s) and new language(s) as well), motivation to learn the new language, mood, musical experience and abilities, and intelligence to the L2 learning environment. These five factors were incorporated into a framework for future research exploring how songs might support L2 learning (see Figure 8.1). The model proposes that, through one or more underlying mechanisms, incorporating L2 songs and singing into the curriculum may increase a learner's L2 listening comprehension, pronunciation and speaking skills; confidence to speak in the new language; grammatical skills; productive musical skills; and potentially sustain the learner's mood and motivation to learn the new language. While at this stage it is clear that many unanswered questions remain, it is hoped that this framework can provide useful directions for future investigations into the effects of singing on L2 learning. This section describes the framework and the basis upon which each component was included.

Figure 8.1: Framework for future research exploring how singing might support modern foreign language learning



Firstly, this framework proposes that there are five important factors that are brought to the L2 classroom or learning context: (1) the learner's motivation(s) to learn the new language; (2) the learner's mood or emotional state at the beginning of the learning session, which is to some extent positive and negative; (3) the learner's previous language learning experience and abilities (both in the native language(s) and new languages); (4) the learner's musical skills, in particular productive musical skills; and (5) general intelligence or IQ. Below, the reasons for including each of these factors is described in turn.

Motivation. Learners' motivation (and change in motivation) to learn the material in the new language was only included as a factor in the second Hungarian study of this research project. In that study, change in motivation did appear as a marginal factor (p = .07) that was predictive of overall Hungarian test performance in the MANCOVA analysis. Previous work and testing has also shown that motivation is an important factor in L2 learning success (Pimsleur et al., 2004).

Mood. Emotions have long been recognised as important to the learning process, with strong emotions (whether positive or negative) at the time of learning (or of an event) dramatically increasing the likelihood that detailed information will be stored in long-term memory (McGaugh, 2004; J. R. Anderson, 1999). Krashen argues that a learner's emotional state is very important for second language acquisition and that teachers must aim to lower students' 'affective filters' by creating a classroom atmosphere that relaxes learners, including through playing classical music (2008). For the two Hungarian studies, stepwise regression analyses showed that positive mood at the start of the experiment session was an important predictor of L2 learning on the Delayed-Recall Hungarian Conversation test (after Phonological Working Memory and Pitch Discrimination (or Productive MAT) scores); MANCOVA analyses for each of the Hungarian studies also showed that positive and negative mood were important predictors of Hungarian test performance.

Previous L1 and L2 learning experience. Researchers have found that first and second language reading and writing skills are positively correlated (Ito, 2009; Bernhardt & Kamil, 1995). Studies have also shown benefits for bilingual individuals who have a high level of competence in both languages on subsequent foreign language writing skills (Sagasta Errasti, 2003) and overall L3 skills (Garate & Iragui, 1993), although other factors such as having high motivation to learn the new language were also important. Having positive experiences, feelings of success, and good strategies for learning new languages would also be likely to support subsequent language learning. In addition, MANCOVA and stepwise regression analyses consistently showed that participants' previous language learning experience was an important factor predicting L2 learning in the three research studies described in this thesis. In addition, in the French arts intervention study, the six bilingual children who did not also have learning difficulties tended to show higher performance on the French tests.

Musical skills (and Productive musical skills). Some work has shown that musical skills and second language abilities are positively correlated (Slevc & Miyake, 2006; Pastuszek-Lipinska, 2008), even when IQ is taken into account (Gilleece, 2006). Preliminary results of another study showed that intensive musical training in childhood can significantly enhance monolingual adult students' second language learning skills and expressive fluency compared to matched nonmusicians, even after only one semester of L2 instruction (Petitto, 2008). Results of the stepwise regression analyses from the two Hungarian studies frequently extracted productive musical skills as an important factor predicting performance on the Hungarian language tests. For certain Hungarian tests, productive musical skills was an even more important factor than previous language learning experience, so this is an essential factor to be included at the beginning of the learning process.

General intelligence (or IQ). As previously mentioned, intelligence (also called

IQ or *g*) is an important factor that can predict learning success in many different areas (Mayes et al., 2009), including L2 learning (Deary et al., 2007). It is likely that general intelligence, or perhaps verbal intelligence, was one of the important predictors in the regression analysis that were not measured in the studies conducted for this research project. This possibility could be explored in future studies.

The second stage is to teach a new language through songs and singing activities. These musical activities should introduce or reinforce L2 vocabulary and grammatical structures that are at an appropriate level for the students (challenging but not too difficult), thereby enabling the implicit and explicit learning of many different aspects of the new language at once. In addition, the songs should be enjoyable for the students to hear and fairly easy to learn, meaning that the rhythmic and melodic complexity and the rate of presentation of the lyrics in the song should take the students' current musical and second language skills into account. The framework proposes five possible means through which listening to songs and singing can support L2 learning. One or all of these explanations may be relevant (to a larger or smaller extent), and they are primarily based on other theories and research findings, with some supporting evidence emerging from the research studies presented in this thesis. The reasons for including each of these five potential explanations of how singing might support L2 learning are outlined below.

Structure of the song. It has been suggested that learning material through songs can enhance verbal memory in the native language because of the way the structure of pitches and rhythmic patterns are aligned with the syllables and words, allowing better initial memory encoding (Wallace, 1994; Yalch, 1991). Researchers have also suggested that learning a song (in the native language) may encode the melody and lyrics together in memory, improving later recall for both elements (Rainey & Larsen, 2002; Wallace, 1994; Calvert & Tart, 1993; Yalch, 1991). It is also possible that each musical event is more salient than each speech event, in part because of the rhythm and meter (Dowling et al., 2002; Tillmann & Dowling, 2007; Thaut et al., 2008), and that the rhyme, rhythmic and melodic structures of a song can provide a 'hook' for retrieving chunks of verbal material (Wolfe & Hom, 1993; Wallace, 1994; McElhinney & Annett, 1996). The components of a song's structure that may support verbal memory in the native language are also likely to benefit second language learning.

Brain's response to rhythm and melody. A number of neuroscientists and researchers in the psychology of music have begun to investigate the possible overlaps between language and music processing (for a review, see Patel, 2008). Findings suggest that certain features of music and language are processed in similar areas of the brain, including syntax (Jentschke & Koelsch, 2009), pitch (Schön, Magné and Besson, 2004), and rhythm (Magné et al., in press). Whereas linguistic processing of songs (and of speech) are focused in the left hemisphere (Yelle & Grimshaw, 2009), melodies (and prosody in language) show greater activation in the brain's right hemisphere, especially in nonmusicians (Sridharan, Levitin, Chafe, Berger, & Menon, 2007); this cooperative dual-hemispheric processing may explain why having learned the lyrics of a song along with its melody can provide more pathways for memory retrieval. Thaut et al. (2008) have speculated that the beneficial effects of songs for learning sequences of verbal information may be due to rhythm's ability to increase the phase-locking of neuronal firing patterns in the brain, and that increasing the brain's efficiency during learning could positively influence memory storage. In addition, Wallace (1994) has argued that the musical structures in songs provide an effective way for learners to 'chunk' verbal material into phrases rather than processing each word separately, enabling later retrieval of phrases rather than individual words. To date, comparably fewer studies have investigated the neural processing of songs – complex stimuli that include melody and rhythm (and often harmony) in addition to lyrics because much of the research in this field has tended to keep music and language separate, investigating the effects of listening to tonal instrumental music compared to listening to speech. While the results are very interesting in terms of what they can reveal about the brain and cognition, as yet these studies do not provide sufficient information about how the potential links between music and language processing can be used to support learning. As this area of research, techniques, and neuroimaging equipment continue to develop, the neural processing of songs during learning and their integration into long-term memory will become clearer.

Attention. Some studies have shown that music is very effective for gaining and retaining attention (Janata, Tillmann, & Bharucha, 2002; Siegel, 1990), and that listening to preferred music even distracts participants from pain from cold significantly more than listening to humour or to silence in the control condition (L. Mitchell, MacDonald, & Brodie, 2006). Findings from a pre-post intervention study with young under-privileged children tested on six measures of language, IQ and numeracy skills showed that music training had a similar beneficial effect to attention training and to receiving more intense adult attention over an eight-week period on children's nonverbal IQ, numeracy, and spatial cognition (Neville et al., 2008). Investigating whether music and songs can more effectively engage attention than speech or other forms of presenting L2 material is a question well suited to exploration through psychological and neuroscience research.

Music can increase positive emotions during learning. Another possible reason that songs can support long-term memory is that positive (and negative) emotions can be induced by listening to music (Kreutz, Ott, Teichmann, Osawa, & Vaitl, 2008) and that singing or making music as part of a group can create a variety of positive responses in participants (Weinberger, 1998; Pitts, 2005). Thus, researchers have argued that including music and songs in the second language classroom may increase positive emotions and emotional arousal, or alternately lead to a state of greater relaxation and a decrease in feelings of self-doubt and self-consciousness, and hence lead to a more effective learning state (Krashen, 1983; Promislow, 2005). More positive emotional states were reported by musician participants who were taught 15 Japanese words through singing and ASL signs compared to participants who learned the same phrases through passive listening and watching the teacher's and another participant's speech or singing with signs (Iwata, 2005). In turn, long-term memory may be enhanced by greater activation of and cooperation between the amygdala and the hippocampus resulting from strong positive emotions during learning (Roozendaal, 2008; McGaugh, 2004; Richter-Levin & Akirav, 2000) through listening to music and singing. An fMRI study showed that when professional singers imagined singing a song in a foreign language, emotional processing areas were strongly activated, compared to lower activation during singing (Kleber et al., 2007). Although the Kleber et al. study did not investigate emotions while singers were learning the song, and in addition it is likely that nonmusicians' brains will react somewhat differently to the way trained singers' do, it is possible that actively taking part in singing activities to learn the words of a song in the target language may improve memory formation through greater emotional involvement during the learning process.

Effect of din or earworms. As previously mentioned, a language learner may involuntarily rehearse L2 words or phrases which repeat in his or her head after having heard, read, or said them, a phenomenon called 'din' in second language education research (Krashen, 1983) or 'earworms' when the repeating material is musical in nature (Kellaris, 2003, 2001). In the French arts intervention study conducted for this thesis, a higher incidence of 'din' was reported for the songs than for the dramatic dialogues (more than half of the children in each class), and a similar percentage of the pupils also reported a preference for listening to the songs relative to the dramatic dialogues. This fits with previous findings of greater reports of 'din' with adult Spanish language learners who heard songs compared to another group that heard a spoken version (poetry) of the Spanish texts (Smith Salcedo, 2002). If L2 sounds and structures presented through songs are more likely to repeat in the students' heads after class, this in turn may encourage them to engage in self-motivated practice of the new language outside of class.

These five possible reasons why singing and listening to songs can support second language learning are interesting, but the research conducted for this thesis was aimed at exploring whether a beneficial effect for songs could be found under rigorously controlled experimental and quasi-experimental classroom learning conditions. Thus, future work will be needed to establish which of these proposed reasons – or others – may lie behind the ability of songs to support L2 learning.

When songs and singing are incorporated into the second language learning process, at the present time there appear to be five main outcomes: (1) L2 listening and speaking skills improve; (2) L2 grammar skills improve; (3) confidence in L2 speaking ability increases; (4) productive musical skills improve; (5) mood and/or motivation to learn the L2 may remain more positive or stronger over time.

Second language listening comprehension, pronunciation, and speaking skills improve. Researchers have argued that folk songs can be particularly helpful for L2 learning (Spicher & Sweeney, 2007) because they best preserve the language's natural prosody (rhythm and intonational contours) for learners to hear and imitate. In the first Hungarian experiment, a listen-and-repeat singing method to learn L2 phrases was significantly more effective than learning the phrases through the speaking or rhythmic speaking method, suggesting that it is easy for participants to transfer chunks of L2 material learned through singing into spoken L2 production. Another study showed an L2 production benefit for material learned through songs, showing that the melody of a song can implicitly teach the intonation of the L2 phrases in the lyrics (Fomina, 2000).

Second language grammatical skills improve. In a previous arts intervention study in which one group of children learned French through songs and another group learned French through drama and visual art activities (Ludke, 2006), results showed that the largest benefit for the music group was on a productive written grammar test which asked children to write a French question that would result in the given response (e.g., for the sentence 'My name is Bart Simpson,' the child could write 'What is your name?' or 'My name is X, and yours?' in French). The potential for better 'chunking' and memory benefits observed when using songs as stimuli in the native language (Thaut et al., 2008; Wallace, 1994) could be particularly relevant because remembering a complete phrase – a sequence of words – is important for productive L2 grammar skills (e.g., quickly finding the correct form of a verb that agrees with its subject). While a significant benefit for grammar learning through a song compared to a dramatic dialogue was not shown in the French study conducted for this thesis, further investigation is warranted.

Confidence in second language speaking abilities increases. In the four-week French arts intervention study, many children reported that learning to sing French songs had increased their confidence to speak in French on the post-intervention opinion questionnaires. This finding of an increase in confidence for speaking is corroborated by a study with ESL learners at different proficiency levels, which de Guerrero (1987) attributes to the effects of increased 'din' for L2 songs (which was also found in the French study conducted for this thesis).

Productive musical skills improve. When song-related activities and singing are regularly incorporated into the L2 curriculum, language learners' musical skills will gradually increase because practising any skill is likely to improve it. In addition, in a previous arts intervention study that incorporated song-related activities and singing into the L2 French classroom for 10 hours in total (Ludke, 2006), the children's ability to sing 'Happy Birthday' improved over a six-week time period. Although the increase in singing abilities was not statistically significant, it is likely that improvements in productive musical skills would show greater increases over time.

Mood and Motivation may stay more positive or stronger over time. Some researchers have argued that the short-term benefits of music listening on spatio-temporal reasoning (the so-called 'Mozart effect,' see Rauscher et al., 1993) arise mainly due to an increase in mood or mental arousal (Schellenberg, 2006; Schellenberg & Hallam, 2005). Given other findings about music's ability to modulate mood (Weinberger, 1998), there is likely some validity to this claim, although in the first Hungarian experiment, the beneficial effects of singing to learn the L2 phrases over a 15-minute period were still present after using MANCOVA to control for the effects of mood. An individual's emotional state during learning can influence content retention (Hulse et al., 2007), so it is important to establish the extent to which mood may be influenced by second language learning through music over time.

Since activities that a language learner enjoys and which create positive emotional responses are more likely to be intrinsically motivating, it would follow that motivation can also be improved by learning new languages through singing and musical activities. Csikszentmihalyi and Nakamura (1989) advanced a theory that individuals are more motivated and enthusiastic about learning when they are taking part in an activity that is both highly enjoyable and challenging. Educators have reported that songs can provide this type of pedagogical material and hence can be used effectively in L2 education to support different language skills (Spicher & Sweeney, 2007; Fomina & Merkulova, 2000). MANCOVA results for the second Hungarian study (the only context in which motivation was investigated in this thesis) showed that change in motivation was a strong predictor of Hungarian test performance, although women's motivation also decreased more than men's in this study, particularly in the sung stimuli conditions.

Both experimental and educational studies should be used to explore the effects of mood and motivation on L2 learning through listening to songs and singing. It would be valuable to investigate how learning L2 songs can influence mood and motivation over short and longer periods of time, while also exploring possible interactions between these factors and a language learner's musical skills over time.

Finally, a crucial feature of this proposed framework for future research is that several of the observed or proposed reasons that singing could support L2 learning are the same components that the learner brings to the second language learning context: motivation, mood, previous L1 and L2 skills, and (productive) musical abilities. Thus, the structure of the current framework suggests that any benefits of singing on L2 learning will show greater improvements as songs are incorporated into L2 instruction over longer time periods.

8.5 Implications for modern foreign language education

This research project has provided support for the claim that singing can support L2 learning. Even at a very basic level of L2 proficiency, learning material in a new language through songs was at least as effective as listening to spoken materials when the individual differences between learners were small. Teachers should not be concerned that their students will have difficulty saying words and phrases learned through singing, since participants in the experimental studies were able to transfer the sung phrases to spoken ones immediately and without any trouble. There is evidence suggesting that songs may be particularly useful for L2 listening comprehension, L2 speaking skills (even after a 20-minute delay), and vocabulary and grammar learning. While the effect sizes found in the stepwise regression analyses did not provide a large amount of predictive power, nevertheless the current experimental evidence does indicate that songs can support L2 learning, particularly for productive skills in the new language (both speaking and writing). Thus, modern language teachers should consider integrating songs into the curriculum to support their students' learning and enjoyment of learning a new language.

It is hoped that modern language teachers who have not done so in the past will begin to include songs in their classroom. This could be simply playing a song in class and asking pupils to draw pictures of any words they recognise in the lyrics, or asking the school's music specialist (if there is one) to help with a musical collaboration. Another option is to use one of the lesson ideas in the workbook developed for the knowledge exchange workshop (available in Appendix D), which were designed to be accessible for teachers without any formal musical training. These activities can provide a useful starting point from which to gradually build a repertoire of creative musical activities to support L2 learning.

A pedagogical implication that arises from the proposed framework of how songs might support L2 learning is that it might be important for modern language teachers to regularly incorporate songs into the curriculum, rather than saving songs for a 'reward' for special occasions or at the end of class. While listening to and singing songs can be fun for language learners of all ages, the framework suggests that incorporating musical activities and singing into the L2 curriculum on a regular basis may provide even greater benefits for second language learning over time because these activities will also gradually increase students' productive musical skills, which appear to be a crucial factor in predicting some L2 learning outcomes.

8.6 Suggestions for future directions

Since the first experimental study showed a benefit for the 'listen-and-repeat' singing condition when the groups were well matched on the measures of individual differences, it would be interesting to take the results further and explore the effect of adding different combinations of modalities to the stimuli. For example, participants could be shown a native speaker's mouth moving as he or she speaks or sings phrases in the new language, or create learning conditions that include gestures or subtitles in the native language at the bottom of the screen. This might provide a more motivating learning experience and it would be an interesting experiment because presenting auditory material, and songs in particular, along with visual aids has been shown to enhance L2 learning in the classroom with young Spanish speakers who had limited English proficiency (Medina, 1993) and with adult student musicians learning Japanese words together with ASL signs (Iwata, 2005).

Because the groups were not well matched for productive musical skills in the second experimental study, it would be interesting to further investigate whether or not input (spoken phrases, poetry, or dialogues compared to sung phrases or lyrics) is sufficient for effective L2 learning, or if oral practice through either speaking or singing during the learning process is an essential element to improve L2 test performance. The results of Gfeller (1983) indicate that learners must repeat the sounds (at least sub- vocally) for musical mnemonics to support verbal retention in the native language. Similar findings were found for the long-term memory benefits of vocalisation on verbal retention in a study conducted by Gathercole and Conway (1988), and another study which showed that both children with developmental delays and control children were more likely to spontaneously produce new lexical items after the words were presented through a song (vs. through a spoken story) in the native language (Kouri & Winn, 2006). Despite the group differences on the individual differences factors, this question merits further exploration because the adult participants in the Listen to Speech condition performed at a very high level on the Hungarian language tests despite a learning method that involved passive listening to the paired-associate English-Hungarian phrases.

Since the stepwise regression analysis using the results of the two experimental studies revealed four factors that consistently had a strong influence on participants' success in Hungarian language learning over a short time period, it would also be useful to further explore the effects on short- and long-term L2 learning of an individual's mood, motivation (and change in motivation) to learn the new language, previous language learning experience, productive musical skills, and intelligence. One approach would be to use the framework in the previous section as a point of departure and to fill in and modify its structure as more details are discovered over time. In particular, three questions are worthy of investigation due to their potential to inform both educational practice and future work in psychology and neuroscience:

1. To what extent are the benefits of including singing in the modern foreign language curriculum due to the repetition of L2 sounds, words and phrases in the learner's mind after the stimuli were presented (e.g., 'din' or 'earworms')?
2. Do the benefits of listening to songs and singing in a new language emerge due to the ability of music to engage learners' attention more effectively?
3. Does singing and music in the L2 curriculum improve or maintain mood and/or motivation to learn the new language better than speech or other means of presenting material in a new language?

In addition, conducting experimental studies that are similar to the ones presented in this thesis, but with specially selected populations could provide information that would be useful for researchers, educators, and policy-makers. For example, the effects of a musical instructional method for language learning could compare performance by musicians vs. nonmusicians, or bilingual vs. monolingual learners of a new language. Work could also investigate the importance of productive musical skills on participants' mood and motivation to learn a new language through songs and singing.

Conclusion

This research project methodically explored the effects of using songs and singing to support modern foreign language learning.

An afternoon workshop and workbook developed to enable non-musician teachers to more effectively incorporate songs into the L2 curriculum showed that an interactive workshop format was useful for sharing knowledge between researchers and practitioners. The workshop facilitated a valuable exchange of information, needs, and ideas that can inform future educational practice and research.

A classroom-based quasi-experiment with adolescent learners showed that incorporating songs and dramatic dialogues into the French curriculum for 15 minutes per class session over a four-week period provided several benefits for beginning and beginning-intermediate level French learning. Vocabulary and grammar scores showed significant overall improvements over the four-week arts intervention period (in the S2 class). In addition, most children reported that they had enjoyed the new activities, especially the songs, and that they had improved their listening comprehension, speaking skills, and confidence to speak in French.

An experimental study which attempted to investigate the effects of active vs. passive and spoken vs. sung phrases showed that the effects of individual differences between learners are very important for paired-associate aural/oral L2 phrase learning. The IDs were especially important for participants assigned to a learning procedure with sung stimuli when they had low productive musical skills, perhaps due to an aptitude-treatment interaction.

The results of a controlled, randomised experimental study showed that, when individual differences between learners in the different groups were small, the 'listen-and-repeat' singing procedure provided significant support for paired-associate phrase learning in an unfamiliar language, particularly for productive L2 skills, both immediately after learning and after a 20-minute delay.

The framework for research proposed in this chapter can guide future educational and experimental investigations in this area. It raises the possibility of a mutually beneficial effect for linguistic and musical skills when songs and singing are incorporated into the modern foreign language learning process, adding to the growing body of evidence linking music and language abilities. This thesis also provides the first experimental evidence that singing during the learning process can support beginning-level foreign language learning.

References

AAMCBI. (2010). *Imaging technologies and techniques.* Athinoula A. Martinos Center for Biomedical Imaging, MIT/Harvard Medical School. URL http://www.nmr.mgh.harvard.edu/martinos/research/technologies.php (last accessed 5 March 2010).

Abello-Contesse, C., Chacón-Beltrán, R., López-Jiménez, M. D., & Torreblanca-López, M. M. (Eds.). (2006). *Age in L2 acquisition and teaching.* Bern, Switzerland: Peter Lang Verlagsgruppe.

Abu-Rabia, S. (2003). The influence of working memory on reading and creative writing processes in a second language. *Educational Psychology*, *23*(2), 209-222.

Abu-Rabia, S. (2004). Teachers' role, learners' gender differences, and FL anxiety among seventh-grade students studying English as a FL. *Educational Psychology*, *24*(5), 711-721.

Abu-Rabia, S., & Kehat, S. (2004). The critical period for second language pronunciation: Is there such a thing? *Educational Psychology*, *24*(1), 77-97.

Adkins, S. (1997). Connecting the powers of music to the learning of languages. *The Journal of the Imagination in Language Learning and Teaching*, *IV*, 1-10.

Ahlkvist, J. A. (2001). Sound and vision: Using progressive rock to teach social theory. *Teaching Sociology*, *29*, 471-482.

AHRC. (2008). *The Arts and Humanities Research Council's (AHRC) Knowledge Transfer (KT) Strategy 2008 - 2011.* [WWW document.] URL http://www.ahrc.ac.uk/About/Policy/Pages/KnowledgeTransferPolicy.aspx (last accessed 12 October 2009).

Albers, B., & Bach, R. (2003). Rockin' soc: Using popular music to introduce sociological concepts. *Teaching Sociology*, *31*, 237-245.

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, *106*(1), 20–29.

Anderson, E. (1993). *Positive use of rap music in the classroom.* ERIC Systems Publication. www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED353588 (last accessed 8 January 2010).

Anderson, J. R. (1999). *Learning and memory: An integrated approach* (2nd ed.). New York, NY: John Wiley & Sons.

Andreou, G., Vlachos, F., & Andreou, E. (2005). Affecting factors in second language learning. *Journal of Psycholinguistic Research*, *34*(5), 429-438.

Anton, R. J. (1990). Combining singing and psychology. *Hispania*, *73*(4), 1166–70.

Archibald, L. M. D., & Gathercole, S. E. (2006). Nonword repetition: A comparison of tests. *Journal of Speech, Language, and Hearing Research*, *49*, 970-983.

Arnold, K. M., & Jusczyk, P. W. (2002). Text-to-tune alignment in speech and song. In *Proceedings of Speech Prosody 2002, Aix-en-Provence, France* (p. 135-138).

Baddeley, A. D. (1990). *Human memory.* London, UK: Lawrence Erlbaum Associates.

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158–173.

Bancroft, W. J. (1995). *The two-sided mind: Teaching and Suggestopedia.*

Baumgartner, T., Esslen, M., & Jäncke, L. (2006). From emotion perception to emotion experience: Emotions evoked by pictures and classical music. *International Journal of Psychophysiology*, *60*, 34–43.

Bellezza, F. S. (1981). Mnemonic devices: Classification, characteristics, and criteria. *Review of Educational Research*, *51*(2), 247–275.

Bernhardt, E. B., & Kamil, M. L. (1995). Interpreting relationships between L1 and L2 reading: Consolidating the linguistic threshold and the linguistic interdepend- ence hypotheses. *Applied Linguistics*, *16*(1), 15–34.

Besson, M., Faïta, F., Peretz, I., Bonnel, A.-M., & Requin, J. (1998). Singing in the brain: Independence of lyrics and tunes. *Psychological Science*, *9*(6), 494–498.

Besson, M., Schön, D., Moreno, S., Santos, A., & Magne, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restorative Neurology and Neuroscience*, *25*(3-4), 399–410.

Bigand, E., & Poulin-Charronnat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition*, *100*, 100-130.

Bigand, E., Tillmann, B., Poulin, B., D'Adamo, D., & Madurell, F. (2001). The effect of harmonic context on phoneme monitoring in vocal music. *Cognition*, *81*, B11–B20.

Bishop, D. (2001). Individual differences in handedness and specific speech and language impairment: Evidence against a genetic link. *Behavior Genetics*, *31*(4), 339-351.

Bongaerts, T., Summeren, C. van, Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition*, *19*(4), 447–465.

Bottari, S. S., & Evans, J. R. (1982). Effects of musical context, type of vocal presentation, and time on the verbal retention abilities of visual-spatially oriented and verbally oriented learning disabled children. *Journal of School Psychology*, *20*, 329–338.

Bower, G. H., & Bolton, L. S. (1969). Why are rhymes easy to learn? *Journal of Experimental Psychology*, *82*(3), 453–461.

Brown, H. D. (2000). *Principles of language learning and teaching* (4th ed.). White Plains, NY: Addison Wesley Longman, Pearson Education.

Brown, N., & Lamb, D. (2004). Parallels between music learning and language acquisition: From fluency to literacy. *The ACIE Newsletter*, *8*(1).

Burman, D. D., Bitan, T., & Booth, J. R. (2008). Sex differences in neural processing of language among children. *Neuropsychologia*, *46*(5), 1349-1362.

Calvert, S. L., & Tart, M. (1993). Song versus verbal forms for very-long-term, long-term, and short-term verbatim recall. *Journal of Applied Developmental Psychology*, *14*(2), 245–260.

Carroll, J. B., & Sapon, S. M. (1959). *Modern Language Aptitude Test.* San Antonio, CA: The Psychological Corporation.

Castagnaro, P. J. (2006). Audiolingual method and behaviorism: From misunderstanding to myth. *Applied Linguistics*, *27*(3), 519–526.

Chan, A. S., Ho, Y.-C., & Cheung, M.-C. (1998). Music training improves verbal memory. *Nature*, *396*(6707), 128.

Chang, C. (2000). *Relationship between music learning and language reading? Review of literature.* ERIC Systems Publication. [WWW document.] URL www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED440375 (last accessed 8 January 2010).

Choksy, L. (1998). *The Kodaly Method I: Comprehensive music education* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155-159.

Colvin Clark, R., & Mayer, R. E. (2008). Learning by viewing versus learning by doing: Evidence-based guidelines for principled learning environments. *Performance Improvement*, *47*(9), 5-13.

Csikszentmihalyi, M., & Nakamura, J. (1989). The dynamics of intrinsic motivation: A study of adolescents. In R. Ames & C. Ames (Eds.), *Research on motivation in education: Goals and cognitions.* (pp. 45–71). New York, NY: Academic Press.

Daltrozzo, J., & Koelsch, S. (2009). Conceptual processing in music as revealed by N400 effects on words and musical targets. *Journal of Cognitive Neuroscience*, *21*(10), 1882–1892.

de Guerrero, M. (1987). The din phenomenon: Mental rehearsal in the second language. *Foreign Language Annals*, *20*, 537–548.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, *35*(1), 13–21.

Diller, K. C. (1981). *Individual differences & universals in language learning aptitude.* Rowley, MA: Newbury House.

Dowling, W. J., Tillmann, B., & Ayers, D. F. (2002). Memory and the experience of hearing music. *Music Perception*, *19*(2), 249-276.

Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology.* (H. A. Ruger & C. E. Bussenius (1964 translation of Über das Gedächtnis), Eds.). New York: Dover Publications.

Fadiga, L., & Craighero, L. (2006). Hand actions and speech representation in Broca's area. *Cortex*, *42*(4), 486–490.

Failoni, J. W. (1993). Music as means to enhance cultural awareness and literacy in the foreign language classroom. *Mid-Atlantic Journal of Foreign Language Pedagogy*, *1*, 97–108.

Fawcett, A., & Nicolson, R. (1998). *The Dyslexia Adult Screening Test.* San Antonio, CA: The Psychological Corporation.

Feld, S., & Fox, A. A. (1994). Music and language. *Annual Review of Anthropology*, *23*(1), 25–53.

Felix, U. (1989). *An investigation of the effects of music, relaxation and suggestion in second language acquisition in schools.* Unpublished doctoral dissertation, Flinders University, Adelaide, Australia.

Filley, C. M., Ramsberger, G., Menn, L., Wu, J., Reid, B. Y., & Reid, A. L. (2006). Primary progressive aphasia in a bilingual woman. *Neurocase*, *12*(5), 296–299.

Fischer, R. A. (1981). Measuring linguistic competence in a foreign language. *International Review of Applied Linguistics in Language Teaching*, *19*(1-4), 207–218.

Fomina, A. S. (2000). Song melody influence on speech intonation memorization. In C. Woods, G. Luck, R. Brochard, S. O'Neill, & J. A. Sloboda (Eds.), *Proceedings of the Sixth International Conference on Music Perception and Cognition.* Keele, Staffordshire, UK: SEMPRE.

Fomina, A. S. (2006). Using modern popular songs to teach a foreign language and culture: Enhancement of emotional perception when developing socio-cultural awareness. In M. Baroni, A. R. Addessi, R. Caterina, & M. Costa (Eds.), *Proceedings of the Ninth International Conference on Music Perception and Cognition.* Bologna, Italy: SEMPRE.

Fomina, A. S., & Merkulova, T. (2000). The use of songs in the EFL classroom. *IATEFL-Ukraine Newsletter*(18), 26–27.

Foster, K. R., Kersh, M. E., & Masztal, N. B. (1999). *An analysis of content delivery systems using speaking voice, speaking with repetition voice, chanting voice, and singing voice.* ERIC Systems Publication. [WWW document.] URL http://www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED437391 (last accessed 8 January 2010).

Fujioka, T., Ross, B., Kakigi, R., Pantev, C., & Trainor, L. J. (2006). One year of musical training affects development of auditory cortical-evoked fields in young children. *Brain*, *129*, 2593-2608.

Gan, L., & Chong, S. (1998). The rhythm of language: Fostering oral and listening skills in Singapore pre-school children through an integrated music and language arts program. *Early Child Development and Care*, *144*(1), 39–45.

Garate, J. V., & Iragui, J. C. (1993). *Bilingualism and third language acquisition.* ERIC Systems Publication. [WWW document.] URL www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED364118 (last accessed 3 February 2010).

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* New York, NY: Basic Books.

Gardner, H. (1999). *Intelligence reframed: Multiple intelligences for the 21st century*. New York, NY: Basic Books.

Gardner, R. C., Tremblay, P. F., & Masgoret, A.-M. (1997). Towards a full model of second language learning: An empirical investigation. *The Modern Language Journal*, *81*(3), 344-362.

Garner, R. (2002, 11 February). Europe tells UK: Improve teaching of our languages. *The Independent*.

Garner, R. (2007, 13 March). Foreign languages to be compulsory from age seven. *The Independent*.

Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, *27*(4), 513-543.

Gathercole, S. E., & Conway, M. A. (1988). Exploring long-term modality effects: Vocalization leads to best retention. *Memory and Cognition*, *16*(2), 110-119.

Gfeller, K. (1983). Musical mnemonics as an aid to retention with normal and learning disabled students. *Journal of Music Therapy*, *20*(4), 179-189.

Gilleece, L. F. (2006). *An empirical investigation of the association between musical aptitude and foreign language aptitude.* Unpublished doctoral dissertation, University of Dublin, Trinity College, Ireland.

Goral, M., Levy, E. S., & Kastl, R. (2007). Cross-language treatment generalization: A case of trilingual aphasia. *Brain and Language*, *103*(8), 203–204.

Goral, M., Levy, E. S., Obler, L. K., & Cohen, E. (2006). Cross-language lexical connections in the mental lexicon: Evidence from a case of trilingual aphasia. *Brain and Language*, *98*(2), 235–247.

Grafton, S., & Cross, E. (2008). Dance and the brain. In C. Asbury & B. Rich (Eds.), *Learning, arts, and the brain: The Dana Consortium report on arts and cognition.* (pp. 61–69). Washington, DC: Dana Foundation.

Halsband, U. (2006). Bilingual and multilingual language processing. *Journal of Physiology*, *99*(4-6), 355–369.

Hannaford, C. (2005). *Smart moves: Why learning is not all in your head* (2nd ed.). Great River Books.

Hannon, E. E. (2009). Perceiving speech rhythm in music: Listeners classify instrumental songs according to language of origin. *Cognition*, *111*(3), 404-410.

Hannon, E. E., & Trainor, L. J. (2007). Music acquisition: Effects of enculturation and formal training on development. *Trends in Cognitive Sciences*, *11*(11), 466–472.

Hanzeli, V. E. (1977). The effectiveness of Cloze tests in measuring the competence of students of French in an academic setting. *The French Review*, *50*(6), 865-874.

Hébert, S., & Peretz, I. (1997). Recognition of music in long-term memory: Are melodic and temporal patterns equal partners? *Memory and Cognition*, *25*(4), 518–533.

Hébert, S., & Peretz, I. (2001). Are text and tune of familiar songs separable by brain damage? *TENNET XI*, 169-175.

Heilenman, L. K. (1983). The use of a Cloze procedure in foreign language placement. *The Modern Language Journal*, *67*(2), 121–126.

Hernandez, A. E., & Li, P. (2007). Age of acquisition: Its neural and computational mechanisms. *Psychological Bulletin*, *133*(4), 638–650.

HMIE. (2008). *Developing the four capacities through modern languages: Focusing on successful learners in primary schools.* Her Majesty's Inspectorate of Education. URL http://www.hmie.gov.uk/documents/publication/dmlps.html#1 (last accessed 10 January 2010).

Ho, Y.-C., Cheung, M.-C., & Chan, A. S. (2003). Music training improves verbal but not visual memory: Cross-sectional and longitudinal explorations in children. *Neuropsychology*, *17*(3), 439-450.

Hulse, L. M., Allan, K., Memon, A., & Read, J. D. (2007). Emotional arousal and memory: A test of the poststimulus processing hypothesis. *American Journal of Psychology*, *120*(1), 73–90.

Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., et al. (2009). Musical training shapes structural brain development. *The Journal of Neuroscience*, *29*(10), 3019–3025.

Ito, F. (2009). Threshold to transfer writing skills from L1 to L2. *Kanto-Koshinetsu Association of Teachers of English Bulletin*, *23*, 1-10.

Iwata, K. (2005). *The effect of active and passive participation with music on the foreign language acquisition and emotional state of university music students.* Unpublished doctoral dissertation, The Florida State University School of Music, Florida State University, FL, USA.

Jakobson, L. S., Cuddy, L. L., & Kilgour, A. R. (2003). Time tagging: A key to musicians' superior memory. *Music Perception*, *20*(3), 307–313.

Jalongo, M. R., & Ribblett, D. M. (1997). Using song picture books to support emergent literacy. *Childhood Education*, *74*(1), 15–22.

Janata, P., Tillmann, B., & Bharucha, J. J. (2002). Listening to polyphonic music recruits domain-general attention and working memory circuits. *Cognitive, Affective, & Behavioral Neuroscience*, *2*(2), 121–140.

Jensen, E. P. (2005). *Top tunes for teaching: 977 song titles & practical tools for choosing the right music every time.* San Diego, CA: Corwin Press.

Jentschke, S., & Koelsch, S. (2009). Musical training modulates the development of syntax processing in children. *Neuroimage*, *47*(2), 735–744.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*(1), 60–99.

Karpicke, J. D., & Roediger III, H. L. (2008). The critical importance of retrieval for learning. *Science*, *319*(5865), 966–968.

Kaufman, A. S. (1994). Practice effects. In R. J. Sternberg (Ed.), *Encyclopedia of human intelligence* (Vol. II, p. 828-833). New York, NY: Macmillan Publishing Company.

Kellaris, J. J. (2001). Identifying properties of tunes that get 'stuck in your head': Toward a theory of cognitive itch. In *Proceedings of the Society for Consumer Psychology Winter 2001 Conference.* Miami, FL: American Psychological Society.

Kellaris, J. J. (2003). Dissecting earworms: Further evidence on the 'song-stuck-in-your-head' phenomenon. In C. Page & S. Posavac (Eds.), *Proceedings of the*

Society for Consumer Psychology Winter 2003 Conference. (p. 220-222). New Orleans, LA: American Psychological Society.

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

Kimmel, K. J. (1998). *The development and evaluation of a music mnemonic-enhanced multimedia computer-aided science instructional module.* Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University.

Kleber, B., Birbaumer, N., Veit, R., Trevorrow, T., & Lotze, M. (2007). Overt and imagined singing of an Italian aria. *Neuroimage*, *36*(3), 889–900.

Koelsch, S. (2009). A neuroscientific perspective on music therapy. *Annals of the New York Academy of Sciences*, *1169*, 374–384.

Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience*, *7*(3), 302–307.

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition*, *11*, 261–271.

Kouri, T. A., & Winn, J. (2006). Lexical learning in sung and spoken story script contexts. *Child Language Teaching and Therapy*, *22*(3).

Krashen, S. D. (1983). The din in the head, input, and the language acquisition device. *Foreign Language Annals*, *16*(1), 41–44.

Kreutz, G., Ott, U., Teichmann, D., Osawa, P., & Vaitl, D. (2008). Using music to induce emotions: Influences of musical preference and absorption. *Psychology of Music*, *36*(1), 101-126.

Lake, R. (2002). *Music and language learning.* [WWW document.] URL www.dtae.org/Adultlit/connections/music.html (last accessed 12 January 2010).

Lapidaki, E. (1996). *Consistency of tempo judgments as a measure of time experience in music listening*. Unpublished doctoral dissertation, Northwestern University, School of Music, Evanston, IL.

Larsen-Freeman, D. (2000). *Techniques and principles in language teaching* (2nd ed.). Oxford, UK: Oxford University Press.

Lê, M. H. (1999). The role of music in second language learning: A Vietnamese perspective. In *Combined 1999 Conference of the Australian Association for Research in Education and the New Zealand Association for Research in*

Education.

Levin, D. Z., & Cross, R. (2004). The strength of weak ties you can trust: The mediating role of trust in effective Knowledge Transfer. *Management Science*, *50*, 1477–1490.

Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception and Psychophysics*, *56*(4), 414–423.

Levitin, D. J., & Menon, V. (2003). Musical structure is processed in "language" areas of the brain: A possible role for Brodmann Area 47 in temporal coherence. *Neuroimage*, *20*, 2142–2152.

Liperote, K. A. (2006). Audiation for beginning instrumentalists: Listen, speak, read, write. *Music Educators Journal*, *93*(1), 46–52.

Lozanov, G. (1999, 3 January). Letter from Dr. G. Lozanov. *International Alliance for Learning*.

Ludke, K. M. (2006). *Using music in foreign language education: An exploratory study*. Unpublished master's thesis, University of Edinburgh.

Lutz, A., & Thompson, E. (2003). Neurophenomenology: Integrating subjective experience and brain dynamics in the neuroscience of consciousness. *Journal of Consciousness Studies*, *10*(9-10), 31-52.

MacLeod, F., & Ross, S. (2009, 22 May). University language classes face £200,000 budget cut. *The Scotsman*.

Marques, C., Moreno, S., Castro, S. L., & Besson, M. (2007). Musicians detect pitch violation in a foreign language better than nonmusicians: Behavioral and electrophysiological evidence. *Journal of Cognitive Neuroscience*, *19*(9), 1453–1463.

Martin, M. (1983). Success! Teaching spelling with music. *Academic Therapy*, *18*(4), 505-506.

Martin, P. (2004). Intonation of French songs: From text to tune. In *Proceedings of Speech Prosody 2004, Nara, Japan.*

Masoura, E. V., & Gathercole, S. E. (1999). Phonological short-term memory and foreign vocabulary learning. *International Journal of Psychology*, *34*, 383-388.

Matsumoto, M., & Nishimura, T. (1998). Mersenne twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Transactions on Modeling and Computer Simulation*, *8*(3), 3-30.

Mayes, S. D., Calhoun, S. L., Bixler, E. O., & Zimmerman, D. N. (2009). IQ and neuropsychological predictors of academic achievement. *Learning and Individual Differences*, *19*(2), 238–241.

McColl, H. (2005). Foreign language learning and inclusion: Who? Why? What? –

and How? *Support for Learning*, *20*(3), 103–108.

McElhinney, M., & Annett, J. (1996). Pattern of efficacy of a musical mnemonic on recall of familiar words over several presentations. *Perceptual and Motor Skills*, *82*(2), 395–400.

McGaugh, J. L. (2004). The amygdala modulates the consolidation of memories of emotionally arousing experiences. *Annual Review of Neuroscience*, *27*, 1-28.

Medina, S. L. (1993). The effects of music upon second language vocabulary acquisition. *National Network for Early Language Learning*, *6*(3), 1–26.

Michael, J. (2006). Where's the evidence that active learning works? *Advances in Physiology Education*, *30*, 159 –167.

Midgley, K. J., Holcomb, P. J., & Grainger, J. (2009). Language effects in second language learners and proficient bilinguals investigated with event-related potentials. *Journal of Neurolinguistics*, *22*, 281–300.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97.

Milovanov, R., Huotilainen, M., Esquef, P., Alku, P., Välimäki, V., & Tervaniemi, M. (2009). The role of musical aptitude and language skills in preattentive duration processing in school-aged children. *Neuroscience Letters*, *460*(2), 161–165.

Milovanov, R., Huotilainen, M., Välimäki, V., Esquef, P., & Tervaniemi, M. (2008). Musical aptitude and second language pronunciation skills in school-aged children: Neural and behavioral evidence. *Brain Research*, *1194*, 81–89.

Mitchell, L., MacDonald, R., & Brodie, E. (2006). A comparison of the effects of preferred music, arithmetic and humour on cold pressor pain. *European Journal of Pain*, *10*(4), 343–351.

Mitchell, R., & Martin, C. (1997). Rote learning, creativity and 'understanding' in classroom foreign language teaching. *Language Teaching Research*, *1*(1), 1–27.

Morrongiello, B. A., & Roes, C. L. (1990). Children's memory for new songs: Integration or independent storage of words and tunes? *Journal of Experimental Child Psychology*, *50*(1), 25–38.

Murphey, T. (1989). The when, where, and who of pop lyrics: The listener's prerogative. *Popular Music*, *8*(2), 185–193.

Murphey, T. (1990). The song stuck in my head phenomenon: A melodic din in the LAD? *System*, *18*(1), 53–64.

Murphey, T. (1992). *Music and song.* Oxford, UK: Oxford University Press.

Natsopoulos, D., Kiosseoglou, G., Xeromeritou, A., & Alevriadou, A. (1998). Do the hands talk on mind's behalf? Differences in language ability between left- and right-handed children. *Brain and Language*, *64*, 182–214.

Neville, H., Andersson, A., Bagdade, O., Bell, T., Currin, J., Fanning, J., et al. (2008). Effects of music training on brain and cognitive development in under-privileged 3- to 5-year-old children: Preliminary results. In C. Asbury & B. Rich (Eds.), *Learning, arts, and the brain: The Dana Consortium report on arts and cognition.* Washington, DC: Dana Foundation.

Overy, K. (2003). Dyslexia and music. From timing deficits to musical intervention. *Annals of the New York Academy of Sciences*, *999*, 497–505.

Overy, K., Nicolson, R. I., Fawcett, A. J., & Clarke, E. F. (2003). Dyslexia and music: Measuring musical timing skills. *Dyslexia*, *9*(1), 18–36.

Overy, K., Norton, A., Ozdemir, E., Helm-Estabrooks, N., & Schlaug, G. (2004). Activation of the left anterior inferior frontal gyrus after melodic intonation therapy in a Broca's aphasia patient. In *Proceedings of the Annual Meeting of the Society for Neuroscience.* San Diego, CA.

Oxford, R. L. (1996). Employing a questionnaire to assess the use of language learning strategies. *Applied Language Learning*, *7*(1-2), 25–45.

Oxford, R. L. (Ed.). (1999). *Language learning motivation: Pathways to the new century.* Honolulu, HI: University Of Hawai'i Press.

Pantev, C., Ross, B., Fujioka, T., Trainor, L. J., Schulte, M., & Schulz, M. (2003). Music and learning-induced cortical plasticity. *Annals of the New York Academy of Sciences*, *999*, 438–450.

Parker, M. (2000). Pronunciation & grammar: Using video and audio activities. *Forum English Teaching*, *38*(1), 24–28.

Pastuszek-Lipinska, B. (2008). Musicians outperform nonmusicians in speech imitation. In *Computer music modeling and retrieval: Sense of sounds.* (Vol. 4969, pp. 56–73). Berlin/Heidelberg, Germany: Springer.

Patel, A. D. (1998). Syntactic processing in language and music: Different cognitive operations, similar neural resources? *Music Perception*, *16*(1), 27–42.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*(7), 674–681.

Patel, A. D. (2007). *Music, language, and the brain.* Oxford, UK: Oxford University Press.

Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition*, *87*(1), B35–B45.

Patel, A. D., Foxton, J. M., & Griffiths, T. D. (2005). Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech. *Brain and Cognition*, *59*, 310–313.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, *10*(6), 717–733.

Peterson, D. A., & Thaut, M. H. (2007). Music increases frontal EEG coherence during verbal learning. *Neuroscience Letters*, *412*, 217–221.

Petitto, L.-A. (2008). Arts education, the brain, and language. In C. Asbury & B. Rich (Eds.), *Learning, arts, and the brain: The Dana Consortium report on arts and cognition.* Washington, DC: Dana Foundation.

Pickering, S. J. (2006). Working memory in dyslexia. In T. P. Alloway & S. E. Gathercole (Eds.), *Working memory and neurodevelopmental disorders.* (pp. 7–40). Hove, East Sussex, UK: Psychology Press.

Pimsleur, P., Reed, D. J., & Stansfield, C. W. (2004). *Pimsleur Language Aptitude Battery: Manual 2004 Edition.* Bethesda, MD: Second Language Testing, Inc.

Pitts, S. (2005). *Valuing musical participation.* Aldershot, Hants, UK: Ashgate Publishing.

Pring, L., & Walker, J. (1994). The effects of unvocalized music on short-term memory. *Current Psychology*, *13*(2), 165–171.

Promislow, S. (2005). *Making the brain-body connection: A playful guide to releasing mental, physical & emotional blocks to success.* (Revised ed.). Access Publishers Network.

Purnell-Webb, P., & Speelman, C. (2008). Effects of music on memory for text. *Perceptual and Motor Skills*, *106*(3), 958–962.

Racette, A., Bard, C., & Peretz, I. (2006). Making non-fluent aphasics speak: Sing along! *Brain*, *129*, 2571–2584.

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

Rainey, D. W., & Larsen, J. D. (2002). The effect of familiar melodies on initial learning and long-term memory for unconnected text. *Music Perception*, *20*(2), 173–186.

Rauschecker, J. P. (1999). Auditory cortical plasticity: A comparison with other sensory systems. *Trends in Neurosciences*, *22*(2), 74–80.

Rauscher, F. H., Shaw, G. L., & Ky, C. N. (1993). Music and spatial task performance. *Nature*, *365*(6447), 611.

Rees, A. L. W. (1977). Techniques for presenting songs. *ELT Journal*, *XXXI*(3), 226–233.

Reilly, H. (2002, 27 February). Teaching a foreign language can be fun – albeit totally pointless. *The Scotsman*.

Richter-Levin, G., & Akirav, I. (2000). Amygdala-hippocampus dynamic interaction in relation to memory. *Molecular Neurobiology*, *22*(1-3), 11–20.

Rifkin, B. (2005). A ceiling effect in traditional classroom foreign language instruction: Data from Russian. *The Modern Language Journal*, *89*(i), 3–18.

Ritter, F. E., Nerb, J., Lehtinen, E., & O'Shea, T. M. (Eds.). (2007). *In order to learn: How the sequence of topics influences learning*. Oxford, UK: Oxford University Press.

Rivers, W. M. (1964). *The psychologist and the foreign language teacher.* Chicago, IL: University of Chicago Press.

Roediger III, H. L., & Karpicke, J. D. (2006). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science*, *1*(3), 181–210.

Roehm, M. L. (2001). Instrumental vs. vocal versions of popular music in advertising. *Journal of Advertising Research*, *41*(3), 49–58.

Roozendaal, B. (2008). Adrenal stress hormones, amygdala activation, and memory for emotionally arousing experiences. *Progress in Brain Research*, *167*, 79–97.

Rubin, D. C. (1995). *Memory in oral traditions: The cognitive psychology of epic, ballads, and counting-out rhymes.* New York, NY: Oxford University Press.

Rubin, D. C., & Wallace, W. T. (1989). Rhyme and reason: Analyses of retrieval cues. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*, 698–709.

Sagasta Errasti, M. P. (2003). Acquiring writing skills in a third language: The positive effects of bilingualism. *International Journal of Bilingualism*, *7*(1), 27–42.

Salamé, P., & Baddeley, A. D. (1989). Effects of background music on phonological short-term memory. *Quarterly Journal of Experimental Psychology*, *41A*, 107–122.

Sammler, D., Koelsch, S., Ball, T., Brandt, A., Elger, C., Friederici, A. D., et al. (2009). Overlap of musical and linguistic syntax processing: Intracranial ERP evidence. *Annals of the New York Academy of Sciences*, *1169*, 494–498.

Sammler, D., Koelsch, S., & Friederici, A. D. (2006). Processing of music syntactic information in brain lesioned patients. In M. Baroni, A. R. Addessi, R. Caterina, & M. Costa (Eds.), *Proceedings of the Ninth International Conference on Music Perception and Cognition.* Bologna, Italy: SEMPRE.

Schacter, D. L. (2002). *The seven sins of memory: How the mind forgets and remembers.* (G. E. McPherson, Ed.). New York, NY: Houghton Mifflin.

Schellenberg, E. G. (2006). Exposure to music: The truth about the consequences. In G. E. McPherson (Ed.), *The child as musician: A handbook of musical* *development.* (pp. 111–134). New York, NY: Oxford University Press.

Schellenberg, E. G., & Hallam, S. (2005). Music listening and cognitive abilities in 10 and 11 year olds: The Blur effect. *Annals of the New York Academy of Sciences*, *1060*, 202–209.

Scherag, A., Demuth, L., Rösler, F., Neville, H. J., & Röder, B. (2004). The effects of late acquisition of L2 and the consequences of immigration on L1 for semantic and morpho-syntactic language aspects. *Cognition*, *93*(3), B97–B108.

Schlaug, G., Norton, A., Overy, K., & Winner, E. (2005). Effects of music training on the child's brain and cognitive development. *Annals of the New York Academy of Sciences*, *1060*, 219–230.

Schofield, K. (2006, 25 March). Schools must do more to boost language learning. *The Scotsman*.

Schön, D., Boyer, M., Moreno, S., Besson, M., Peretz, I., & Kolinsky, R. (2008). Songs as an aid for language acquisition. *Cognition*, *106*(2), 975–983.

Schunk, H. A. (1999). The effect of singing paired with signing on receptive vocabulary skills of elementary ESL students. *Journal of Music Therapy*, *XXXVI*(2), 110–124.

Schuster, D. H. (1985). The effects of background music on vocabulary learning. *The Journal of the Society for Accelerative Learning and Teaching*, *10*(1), 21–42.

Scovel, T. (1979). Review of Suggestology and outlines of Suggestopedy by Georgi Lozanov. *TESOL Quarterly*, *13*(2), 55–66.

Serafine, M. L., Crowder, R. G., & Repp, B. H. (1984). Integration of melody and text in memory for songs. *Cognition*, *16*, 285–303.

Shtakser, I. (2001). *Using music and songs in the foreign language classroom.* University of Texas. [WWW document.] URL http://www.laits.utexas.edu/hebrew/music/music.html (last accessed 10 January 2010).

Siegel, L. S. (1990). The potential for therapeutic applications of music on problems related to memory and attention. *Journal of Music Therapy*, *27*(4), 195–208.

Silverman, M. J. (2007). The effect of paired pitch, rhythm, and speech on working memory as measured by sequential digit recall. *Journal of Music Therapy*, *44*(4), 415–427.

Skehan, P. (1989). *Individual differences in second-language learning*. London, UK: Edward Arnold.

Slevc, L. R., & Miyake, A. (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science*, *17*(8), 675-681.

Slevc, L. R., Rosenberg, J., & Patel, A. D. (2009). Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin and Review*, *16*(2), 374–381.

Sloboda, J. A. (2005). *Exploring the musical mind: Cognition, emotion, ability, function.* New York, NY: Oxford University Press.

Smith, M. C., & Phillips Jr, M. R. (2001). Age differences in memory for radio advertisements: The role of mnemonics. *Journal of Business Research*, *53*, 103–109.

Smith Salcedo, C. (2002). *The effects of songs in the foreign language classroom on text recall and involuntary mental rehearsal.* Unpublished doctoral dissertation, Louisiana State University and Agricultural and Mechanical College.

Snyder, B. (2001). *Music and memory: An introduction.* Cambridge, MA: The MIT Press.

Sommera, I. E., Alemanc, A., Somersa, M., Boksa, M. P., & Kahna, R. S. (2008). Sex differences in handedness, asymmetry of the planum temporale and functional language lateralization. *Brain Research*, *1206*, 76–88.

Spicher, L., & Sweeney, F. (2007). Folk music in the L2 classroom: Development of native-like pronunciation through prosodic engagement strategies. *Connections*, *1*, 35–48.

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

Sridharan, D., Levitin, D. J., Chafe, C. H., Berger, J., & Menon, V. (2007). Neural dynamics of event segmentation in music: Converging evidence for dissociable ventral and dorsal networks. *Neuron*, *55*, 521–532.

Stansell, J. W. (2005). *The use of music for learning languages: A review of the literature.* Unpublished master's thesis, University of Illinois at Urbana-Champaign.

Steinbeis, N., & Koelsch, S. (2008a). Comparing the processing of music and language meaning using EEG and fMRI provides evidence for similar and distinct neural representations. *PLoS One*, *3*(5), 22–26.

Steinbeis, N., & Koelsch, S. (2008b). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex*, *18*, 1169–1178.

Stern, E. (2005). Pedagogy meets neuroscience. *Science*, *310*(5749), 745.

Szpunar, K. K., Schellenberg, E. G., & Pliner, P. (2004). Liking and memory for musical stimuli as a function of exposure. *Journal of Experimental Psychology:*

Learning, Memory, and Cognition, *30*(2), 370–381.

Thaut, M. H., Peterson, D. A., & Mcintosh, G. C. (2005). Temporal entrainment of cognitive functions: Musical mnemonics induce brain plasticity and oscillatory synchrony in neural networks underlying memory. *Annals of the New York Academy of Sciences*, *1060*(1), 243–254.

Thaut, M. H., Peterson, D. A., Sena, K. M., & Mcintosh, G. C. (2008). Musical structure facilitates verbal learning in multiple sclerosis. *Music Perception*, *25*(4), 325–330.

The Independent. (2007, 13 March). Leading article: Revolutionary approach. *The Independent*.

Thomas, W. P., & Collier, V. P. (2002). *A national study of school effectiveness for language minority students' long-term academic achievement.* Santa Cruz, CA: Center for Research on Education, Diversity and Excellence, University of California – Santa Cruz.

Thomas-Hunt, M. C., Ogden, T. Y., & Neale, M. A. (2003). Who's really sharing? Effects of social and expert status on knowledge exchange within groups. *Management Science*, *49*, 464–477.

Thorndike, R., Hagen, E., & Jerome, M. (1986). *The Stanford-Binet Intelligence Scale.* Chicago, IL: Riverside Publishing.

Tillmann, B., & Dowling, W. J. (2007). Memory decreases for prose, but not for poetry. *Memory and Cognition*, *35*(4), 628–639.

Ting, Y. L. T. (2002). An in-house training session on the use of songs. *Humanising Language Teaching*, *4*(2), 1–11.

Tumanov, A. (1986). Music in L2 pedagogy: A comparative analysis of music and language. *Russian Language Journal*, *40*(136-137), 35–54.

Turner, P. R. (1974). Why Johnny doesn't want to learn a foreign language. *The Modern Language Journal*, *58*(4), 191–196.

Twomey, A., & Esgate, A. (2002). The Mozart effect may only be demonstrable in nonmusicians. *Perceptual and Motor Skills*, *95*(3), 1013–1026.

Van Voorhis, C. (2002). Stat jingles: To sing or not to sing. *Teaching of Psychology*, *29*(3), 249–250.

Visser, B. A., Ashton, M. C., & Vernon, P. A. (2006). Beyond g: Putting multiple intelligences theory to the test. *Intelligence*, *34*(5), 487–502.

Vuust, P., Roepstorff, A., Wallentin, M., Mouridsen, K., & Østergaard, L. (2006). It don't mean a thing... Keeping the rhythm during polyrhythmic tension, activates language areas (BA47). *Neuroimage*, *31*(2), 832–841.

Walczak, D., & Reuter, M. (1994). Using popular music to teach sociology: An evaluation by students. *Teaching Sociology*, *22*(3), 266–269.

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

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070.

Weinberger, N. M. (1998). Understanding music's emotional power. *MuSICA*, *V*(2).

Welch, C. (2007). Choral music in the junior high/middle school: Teaching choral literature with foreign language texts. *Choral Journal*, *47*(10), 59–60.

White, L. (2005, 3 September). Fascinating rhythm. *New Scientist*.

Wolfe, D., & Hom, C. (1993). Use of melodies as structural prompts for learning and retention of sequential verbal information by preschool students. *Journal of Music Therapy*, *30*(2), 100–118.

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, *10*(4), 420–422.

Wong-Fillmore, L. W. (1976). *The second time around: Cognitive and social strategies in second language acquisition.* Unpublished doctoral dissertation, Stanford University, Palo Alto, CA.

Yalch, R. F. (1991). Memory in a jingle jungle: Music as a mnemonic device in communicating advertising slogans. *Journal of Applied Psychology*, *76*(2), 268–275.

Yasui, T., Kaga, K., & Sakai, K. (2009). Language and music: Differential hemispheric dominance in detecting unexpected errors in the lyrics and melody of memorized songs. *Human Brain Mapping*, *30*(2), 588–601.

Yelle, S., & Grimshaw, G. (2009). Hemispheric specialization for linguistic processing of sung speech. *Perceptual and Motor Skills*, *108*(1), 219–228.

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science*, *256*(5058), 846–849.

Zetterberg, C., Backlund, H., Karlsson, J., Werner, H., & Olsson, L. (1998). Musculoskeletal problems among male and female music students. *Medical Problems of Performing Artists*, *13*(4), 160.

Ziegler, A. (2007). Rhythm and rhyme – the effects of music on vocabulary development of kindergarten students. In D. A. McAllister & S. R. Deaver (Eds.), *Culminating experience action research projects* (Vol. 10, pp. 490–506). Chattanooga, TN: The University of Tennessee.

Zimmermann, L. S. (2006). *A descriptive study of the relationship between language reading proficiency and the age at which music lessons are begun.* Unpublished master's thesis, Silver Lake College, Manitowoc, Wisconsin, USA.

Appendix A

Supplemental Materials for Hungarian Study 1

A.1 Informed Consent Form

Figure A.1: Informed Consent Form: Information sheet

University of Edinburgh
Auditory Memory Study
Researcher: Karen M. Ludke

Information for research study participants

You are invited to participate in a research study that investigates how individuals' auditory memory functions when learning to say phrases in a foreign language.

If you choose to participate in this study, you will be asked to learn a series of 20 short words and phrases in a language you do not know. You will be asked to repeat the phrases aloud as best you can, while trying to remember the English meaning of each new phrase you are learning. We will make a high-quality audio recording of your session. We would like your permission to use the audio we will record, which includes your voice, for later analysis and comparison with performance by other participants. Your individual privacy will be maintained in any published, educational, written, and electronic reports and data resulting from the study.

You are free to call a halt at *any* stage

You have no obligation to take part. Your participation is voluntary and you are free to withdraw your consent or to discontinue participation at any time — before, during, or after the session — without penalty. You have the right to refuse to answer particular questions.

What do we need from you?

For this study, the most important task is for you to listen to a recording of English words followed by the same phrases in an unfamiliar language, and to repeat the new phrases aloud as best you can while trying to remember the English meaning. We also need you to do a few tests related to what you have heard and practised, to complete a few additional computer-based activities, and finally to respond to a brief questionnaire. The total amount of time we will need your participation is approximately 60 minutes. Beyond the remuneration that you will receive (£6), there are no long-term benefits to participation.

What is this research for and what will we do with the data?

We are interested in the effects that different auditory presentations of material will have on people who are learning to speak a foreign language.

The audio recordings of all participants' voices will be used as data in this research. Recordings will be used only for research and educational purposes. On the Consent Form, you may choose whether or not to allow your anonymous recording to be added to the recordings archive for possible use for future research and teaching purposes.

We hope to publish the results of the research in the future in books, journals, and on the Internet, where appropriate. In publications, we will not identify individual participants in any way, and your Consent Form will be kept securely. Complete confidentiality will be respected at all times.

If you are willing to participate in this study, please fill in and sign the attached Consent Form and return it to the researcher. Please keep this information sheet. If you have further questions, you are encouraged to contact me at k.ludke@sms.ed.ac.uk or at the address below.

Karen M. Ludke
University of Edinburgh, IMHSD
12 Nicolson Square, Alison House
Edinburgh EH8 9DF

Figure A.2: Informed Consent Form: Signature sheet

University of Edinburgh
Auditory Memory Study
Researcher: Karen M. Ludke

Informed Consent Form for Study Participant

By signing below, I confirm and agree to the following statements and conditions:

- Δ I agree to participate in this research.
- Δ I have carefully read this Consent Form and the accompanying information sheet, which together give information about the research, the session, and the recording. I have had the opportunity to ask questions about the study.
- Δ I understand that I am under no obligation to take part in this study and that a decision not to participate will not be a problem.
- Δ I understand that I have the right to withdraw from this study at any stage.
- Δ I understand that this is educational, non-therapeutic research from which I cannot expect to derive any long-term benefit.
- Δ I am a native English speaker with normal hearing who is over the age of 18, or the parent/guardian of a participant younger than 18 years old.
- Δ I permit the researcher to keep my computerised and audio-recorded data, with all personally identifying information removed, in a digital archive for possible future research and/or teaching purposes.

Name: ____

Email address: ____

Signature: ____

Today's Date: ______________________

Please address any questions regarding this consent form to:

Karen M. Ludke
University of Edinburgh, IMHSD
12 Nicolson Square, Alison House
Edinburgh EH8 9DF
k.ludke@sms.ed.ac.uk

A.2 Script for Hungarian Study 1

Introduction and brief small talk to put participant at ease.

Hi, I'm Karen. It's nice to meet you. Thank you for coming. How are you? [Beautiful/Awful] weather today, isn't it?

We begin walking down stairs to the room where participant will do experiment session.

I'm not sure if you remember what this study is about, but it's the auditory memory study. You will hear some words and phrases in a language which you hopefully don't already know, and also the English translation of those phrases. What we need you to do is to repeat what you hear in the new language aloud, as best you can, and to try to remember the English meaning of the phrases you hear. There are a few other activities on the computer as well, but that's the main point.

We enter the room where the participant will do the experiment session.

Have a seat there and make yourself comfortable. If you could read through that sheet and then fill in the second page, that would be great.

Participant fills in the Informed Consent Form and the researcher answers any questions.

That's great, thank you. So, first I'm going to turn this on.

Researcher presses 'record' on the digital audio recorder.

Great, it looks like that is working properly. Now, the first thing we're going to do is I'm going to read some fake words out loud to you. You don't need to remember these words, and they don't mean anything. Just try to repeat them as best you can. Since we have a different accent, try to say the same word, even if it has a slightly different pronunciation from the way I've said it. Does that make sense?

Researcher waits for confirmation/agreement.

Okay, so the first word is...

Researcher reads the 20 low-wordlike items on the childrens non-word repetition task, pausing after each one to allow the participant to repeat the word; repeat individual words again, once, if requested.

That's fantastic, thank you very much. So, next we're going to turn on the computer screen and if you could just read the instructions there – it basically says the same thing as the sheet you just read – and then start going through the pages. At some point you'll reach a screen which asks you to stop and get me, so when you reach that point, let me know. It's just to make sure you understand the instructions for the next part. I'll be right here so you can feel free to ask me any questions.

Researcher looks at some papers in the corner where the computer screen is not visible while the participant begins the computer-based component of the experiment, filling in basic information, the pre-session PANAS mood questionnaire, and the multiple-choice Hungarian vocabulary pre-test; participant stops when he/she finishes the pre-test and says they've completed that part.

Okay, great. If you could click to the next page, that has the instructions on it.

Participant reads the instructions and indicates he/she understands or else asks questions until everything is clear and he/she is ready to proceed.

All right, if you could just put on these headphones now and make sure they feel comfortable.

Participant does so, and when ready, clicks to the next page (practice session with written words on screen). Participant repeats the three practice phrases aloud and the researcher gives additional explanation if necessary.

Extra explanation if participant is in the Rhythmic condition:

Yes, that's great, and try to repeat what you hear with the same rhythm that you hear in the recording. You're not being judged on how perfect it sounds. Just try your best.

Extra explanation if participant is in the Singing condition:

Yes, that's great, and try to repeat what you hear with the same rhythm and the same notes that you hear in the recording. You're not being judged on how perfect it sounds. Just try your best.

Great, that's right. Now, the next practice page will be like the second and third listening/learning sessions. It's exactly the same as before, except that this time you won't see the words written on the screen, you'll only hear them.

When ready, participant clicks to the next page (practice session without words).

Great, so, there will be three learning sessions, which will be just like what you just did here. While you're learning the phrases, you can feel free to use any strategies you can think of. For example, you could say the English phrase out loud if you want to, or you could repeat the phrase in the new language more than once – whatever you think might help you remember.

The learning sessions will be followed by three tests which become easier and easier as you go along. There's a lot to remember, so if you can only remember one word or part of the phrase, then just say that – it's not all-or-nothing. Just try your best. The tests will be followed by some other activities on the computer, so follow the instructions on the screen until you reach the end. I'll be in the room just outside if you have any questions. Otherwise, just come and find me when you finish.

Researcher waits for agreement, then stands up to leave.

Great, best of luck, and thank you very much.

Researcher leaves the room.

After participant has finished the experiment and come to get the researcher, we re-enter the room.

Finished? Great, thanks. Now, if you wouldn't mind just filling in this brief questionnaire, and I will get your money.

The participant fills in the Debriefing questionnaire and the researcher puts their money on the desk nearby.

Great, thank you very much. You were probably wondering what the language was: it's Hungarian. Here is a debriefing sheet with more information about the study. If you know other people who might be interested in participating, please feel free to give them my email address. Thanks again.

The researcher directs them to the exit or if they seem lost, walks with them upstairs to the lobby.

A.3 Stimuli for Hungarian studies 1 and 2

Learning condition	**Material to learn**	
Spoken	**Hungarian**	**English**
	20 phrases 43 words 87 syllables	20 phrases 66 words 79 syllables
Rhythmic	Same as spoken + 20 rhythms	Same as spoken
Singing	Same as rhythmic + 20 melodies	Same as spoken

Figure A.3: Music for rhythmic and sung stimuli in the learning sessions

Table A.1: Order of stimuli presentation for the Hungarian learning sessions

Item	Hungarian	English
Practice 1	Hi.	Szia.
Practice 2	Excuse me.	Elnézést.
Practice 3	Sorry.	Bocsánot.
1	Where is it?	Hol van?
2	I don't understand.	Nem értem.
3	Can I help you?	Segíthetek?
4	You're welcome.	Szívesen.
5	How are you?	Hogy van?
6	Yes, thank you.	Igen, köszönöm.
7	I don't know.	Nem tudom.
8	I am well.	Jól vagyok.
9	I only want water.	Csak vizet akarok.
10	I am Maria.	Márja vagyok.
11	Where can I buy tickets?	Hol lehet jegyet váltani?
12	Could you repeat that, please?	Megismételné, kérem?
13	Good day.	Jó napot kívánok.
14	Help.	Segítség.
15	Do you speak English?	Beszélsz angolul?
16	How much does it cost?	Mennyibe kerül?
17	Goodbye.	Viszontlátásra.
18	Call a doctor, please.	Hívjan orvost, kérem.
19	I understand a little, yes.	Értem egy kicsit, igen.
20	How can I find the train station?	Merre van a pályaudvar?

A.4 Hungarian test item presentation

Table A.2: Items for Hungarian Production Test

Item	Test Prompt	Correct Answer
1	Can I help you?	Segíthetek?
2	Good-bye.	Viszontlátásra.
3	I don't understand.	Nem értem.
4	How can I find the train station?	Merre van a pályaudvar?
5	Help.	Segítség.
6	Where is it?	Hol van?
7	Could you repeat that, please?	Megismételné, kérem?
8	Yes, thank you.	Igen, köszönöm.
9	Good day.	Jó napot kívánok.
10	I understand a little, yes.	Értem egy kicsit, igen.
11	Where can I buy tickets?	Hol lehet jegyet váltani?
12	I am well.	Jól vagyok.
13	I only want water.	Csak vizet akarok.
14	Call a doctor, please.	Hívjan orvost, kérem.
15	You're welcome.	Szívesen.
16	How are you?	Hogy van?
17	I am Maria.	Márja vagyok.
18	Do you speak English?	Beszélsz angolul?
19	How much does it cost?	Mennyibe kerül?
20	I don't know.	Nem tudom.

Table A.3: Items for English Recall Test

Item	Test Prompt	Correct Answer
1	Értem egy kicsit, igen.	I understand a little, yes.
2	Márja vagyok.	I am Maria.
3	Beszélsz angolul?	Do you speak English?
4	Segíthetek?	Can I help you?
5	Hogy van?	How are you?
6	Hívjan orvost, kérem.	Call a doctor, please.
7	Nem értem.	I don't understand.
8	Hol lehet jegyet váltani?	Where can I buy tickets?
9	Segítség.	Help.
10	Szívesen.	You're welcome.
11	Merre van a pályaudvar?	How can I find the train station?
12	Mennyibe kerül?	How much does it cost?
13	Jó napot kívánok.	Good day.
14	Igen, köszönöm.	Yes, thank you.
15	Megismételné, kérem?	Could you repeat that, please?
16	Csak vizet akarok.	I only want water.
17	Nem tudom.	I don't know.
18	Viszontlátásra.	Good-bye.
19	Jól vagyok.	I am well.
20	Hol van?	Where is it?

Table A.4: Items for Hungarian Recognition Test

Item	Test Prompt	Correct Answer
1	Jó napot kívánok.	Same
2	Sehegítek?	Different
3	Jól vagyok.	Same
4	Hol lehet jegyet taválni?	Different
5	Hogy van?	Same
6	Márja vagyok.	Same
7	Szíseven.	Different
8	Mennyibe kerül?	Same
9	Beszélsz golanul?	Different
10	Nem tudom.	Different
11	Értem egy csitki, igen.	Different
12	Segítség.	Same
13	Zetvi csak akarok.	Different
14	Nem értem.	Same
15	Hívjan vostor, kérem.	Different
16	Merre van a pályaudvar?	Same
17	Igen, könöszöm.	Different
18	Viszontláslátra.	Different
19	Hol van?	Same
20	Megisteménlné, kérem?	Different

Table A.5: Items for Hungarian Multiple-Choice Vocabulary test

Item				
1. váltani	*A. tickets*	B. thank you	C. doctor	D. week
2. nem	A. I know	B. *not*	C. a little	D. I speak
3. pályaudvar	A. thank you	B. can I help you?	C. evening	*D. train station*
4. kérem	A. doctor	*B. please*	C. find	D. excuse me
5. értem	A. call	*B. I understand*	C. please	D. I know
6. kerül	A. what	B. excuse me	C. week	*D. cost*
7. beszélsz	*A. you speak*	B. call	C. tickets	D. how much
8. kiscit	A. evening	B. please	*C. a little*	D. I would like
9. orvost	A. only	B. tickets	C. visit	*D. doctor*
10. vizet	*A. water*	B. please	C. name	D. you speak
11. jó	A. yes	B. speak	C. I	*D. good*
12. napot	A. good-bye	B. name	*C. day*	D. tickets
13. csak	A. visit	*B. only*	C. coffee	D. water
14. segítség	*A. help*	B. I understand	C. how much	D. a little
15. hol	*A. where*	B. are	C. buy	D. is
16. tudom	A. some	B. only	*C. know*	D. you
17. van	A. water	*B. is*	C. how	D. yes
18. szívesen	A. tickets	*B. you're welcome*	C. doctor	D. please
19. mennyibe	A. you say	B. want	C. a little	*D. how much*
20. vagyok	*A. I am*	B. good	C. water	D. excuse me

Table A.6: Items for Delayed-Recall Hungarian Conversation

Item	**Test Prompt**	**Appropriate Answer**
1	Jó napot kívánok.	Jó napot kívánok.
2	Hogy van?	Jól vagyok.
3	Segíthetek?	Igen, köszönöm. / Merre van a pályaudvar? / etc.
4	Nem értem. Megismételné, kérem?	[Repeat previous response.]
5	Viszontlátásra.	Viszontlátásra.

A.5 Debriefing questionnaire – Study 1

1. Before beginning this study, did you know the meaning of any of the words you heard in this new language? If so, could you say how many (or which) words you already knew?
2. What do you think the experiment was about?
3. Would you suggest any changes or any ways of improving the learning experience for people participating in future studies?
4. Do you have any other comments?

A.6 Debriefing sheet – Study 1

Auditory Memory Study – Debriefing

This sheet explains the purpose of the study you just participated in.

Please do not tell others what you did during the session or share this information about the research questions, because it could affect their performance if they also wish to take part.

Thank you very much!

The purpose of this study was to try to measure how well people could learn English phrases that were paired with phrases in an unfamiliar language. The words you heard and tried to imitate as part of this study are real phrases in Hungarian, which were recorded by a native speaker of that language.

We were interested in whether or not learning might be improved by hearing different auditory presentations of the material. Previous studies have indicated that melodies may help people remember words in their native language, so we investigated whether or not memory can be improved when learning words and phrases in a foreign language. By looking at people's performance in different listening conditions, we hope to explore whether a musical presentation of material might facilitate memory when learning a new language, compared to hearing phrases presented as speech.

We were also investigating whether performance on this short-term auditory learning and memory task would be influenced by certain differences between the people who participated. We were interested in the extent to which differences such as mood, working memory, previous language learning experience, and musical background might influence learning and performance.

If you have any further questions about this study, please feel free to ask the lead researcher, Karen Ludke, or to send her an email (k.ludke@sms.ed.ac.uk).

Thank you again for your participation!

A.7 Hungarian test histograms and density plots – Study 1

Figure A.4: Histograms for Hungarian Production Test

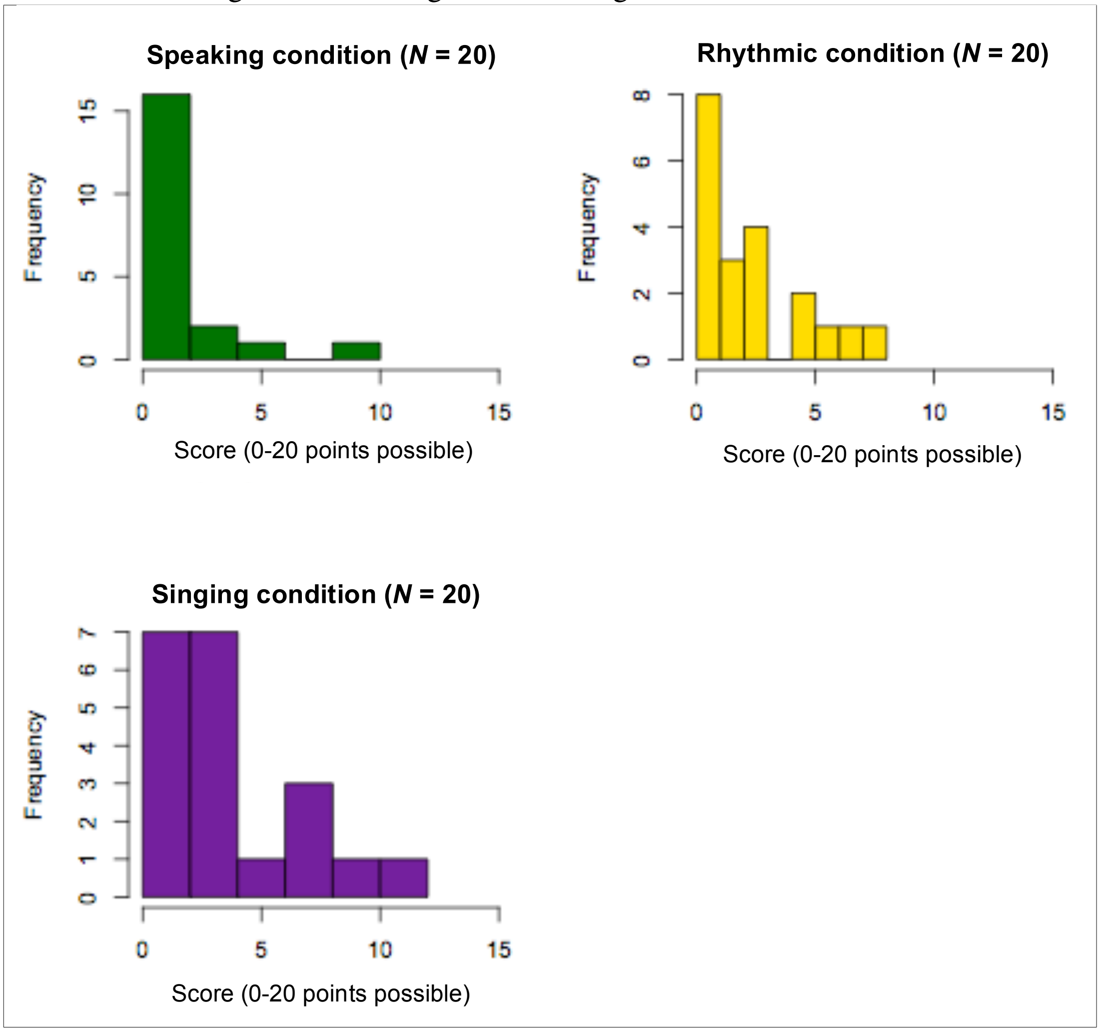


Figure A.5: Density plots for Hungarian Production Test

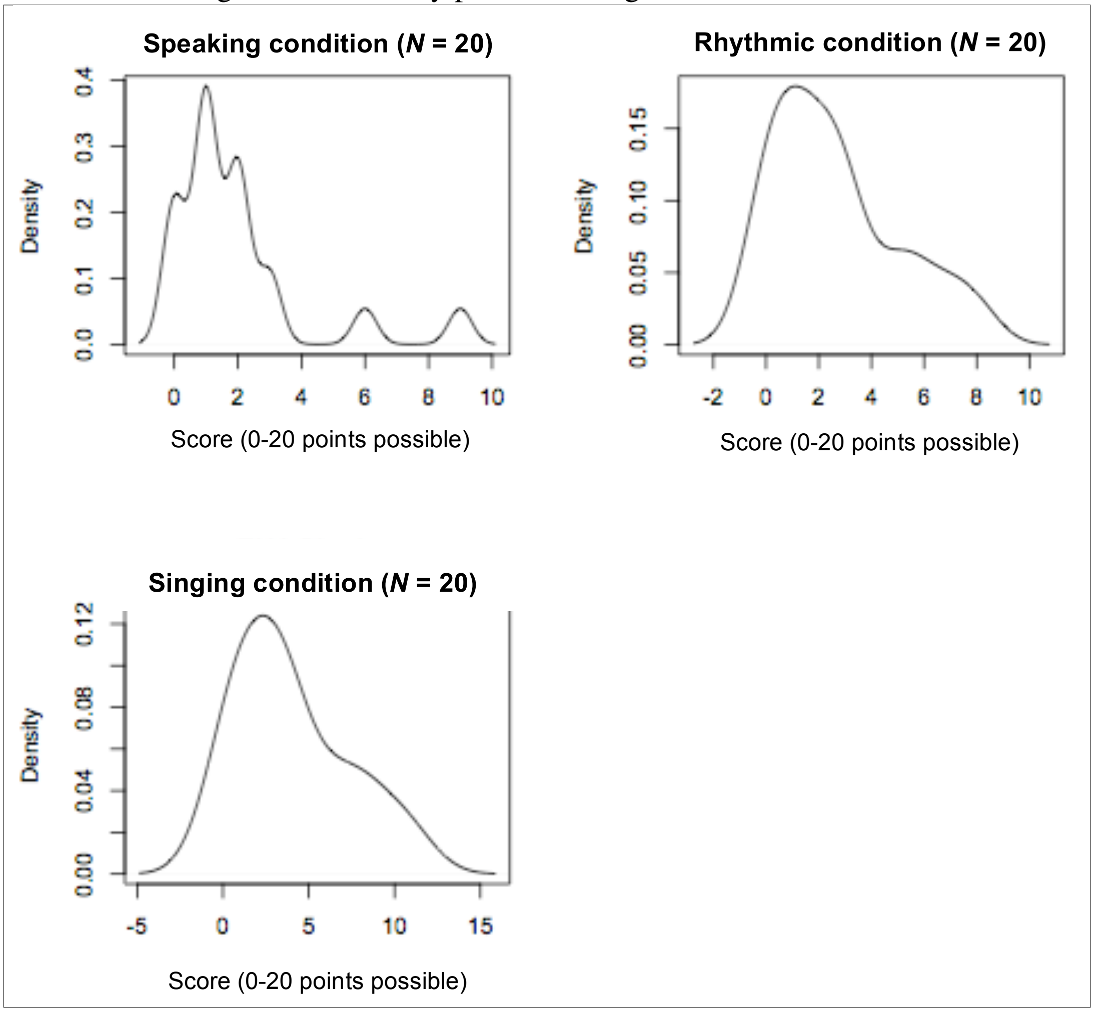


Figure A.6: Histograms for English Recall Test

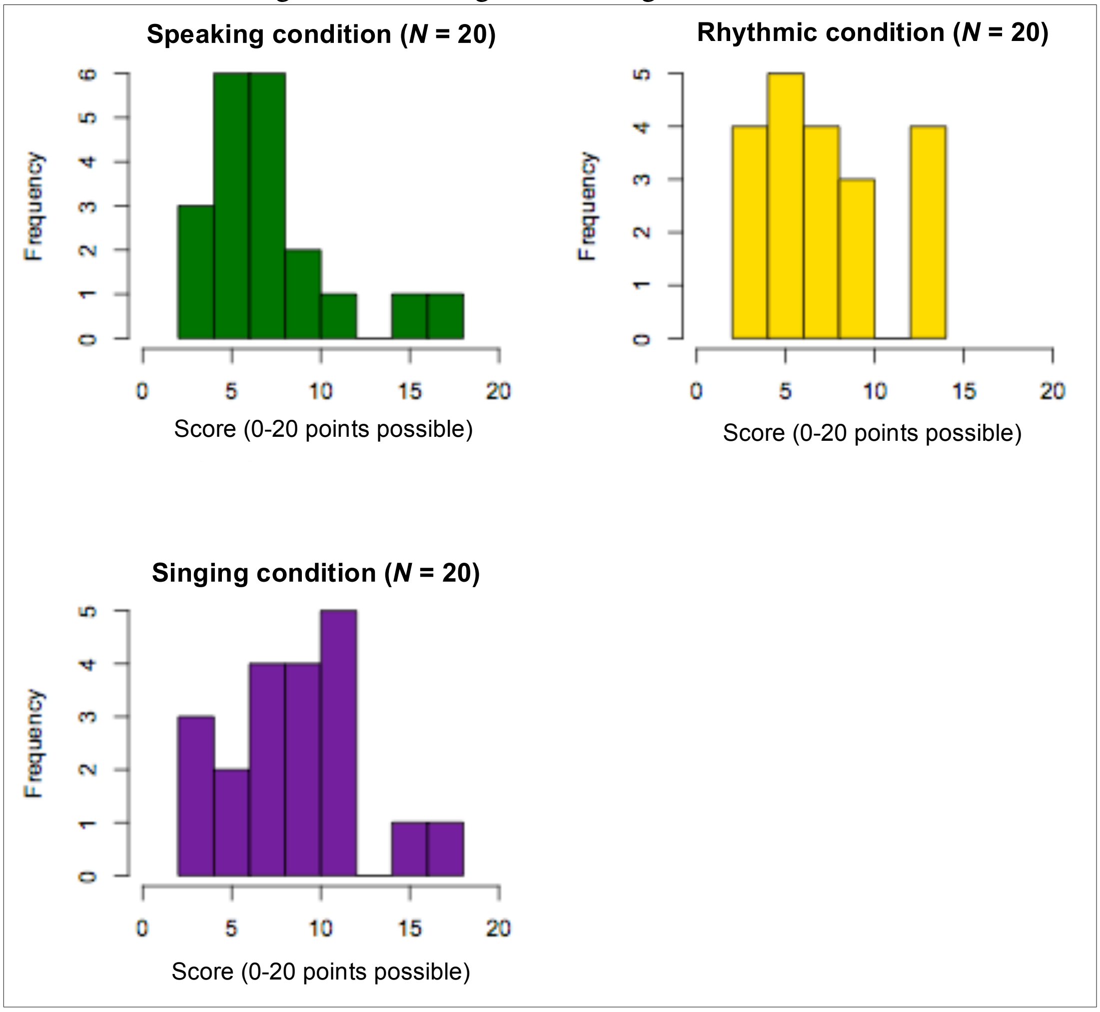


Figure A.7: Density plots for English Recall Test

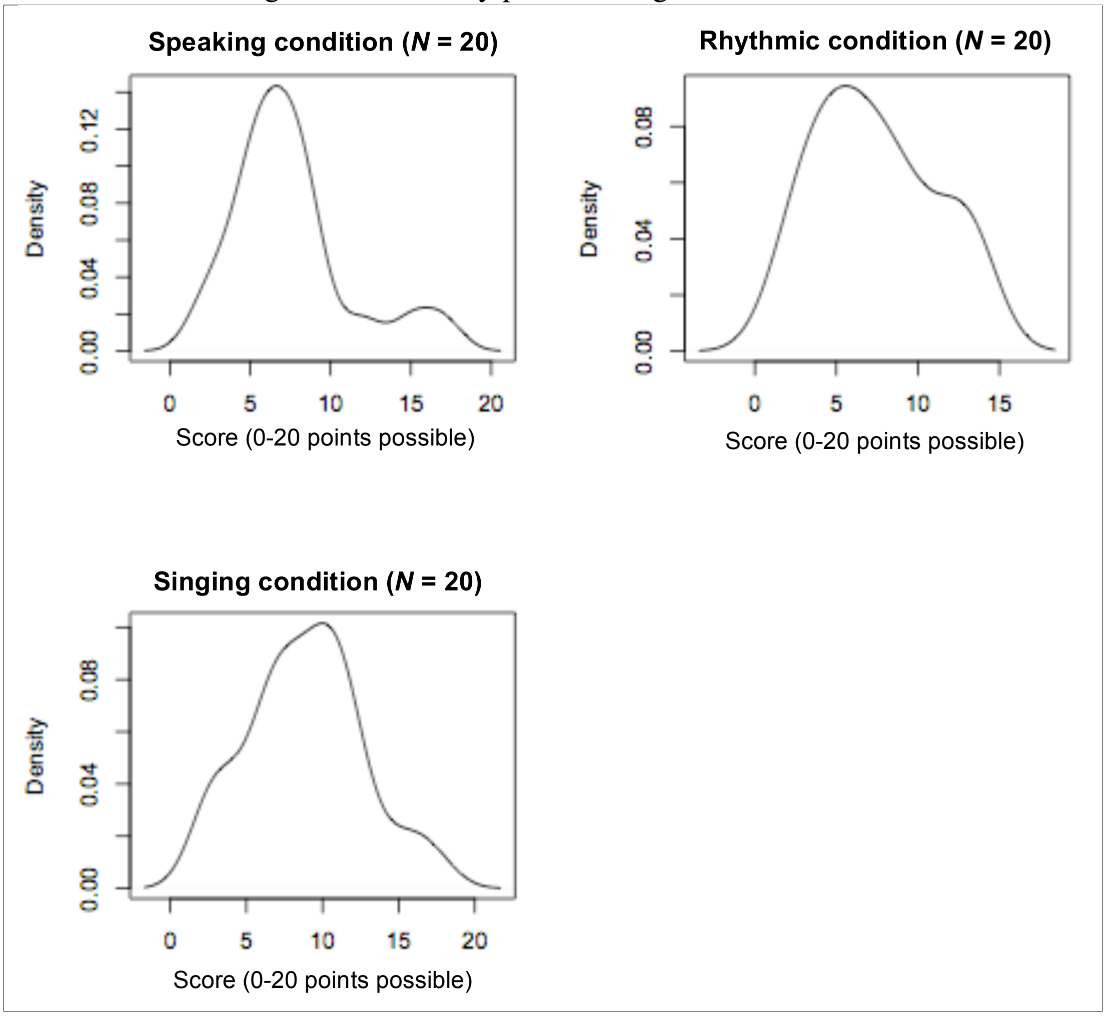


Figure A.8: Histograms for Hungarian Recognition Test

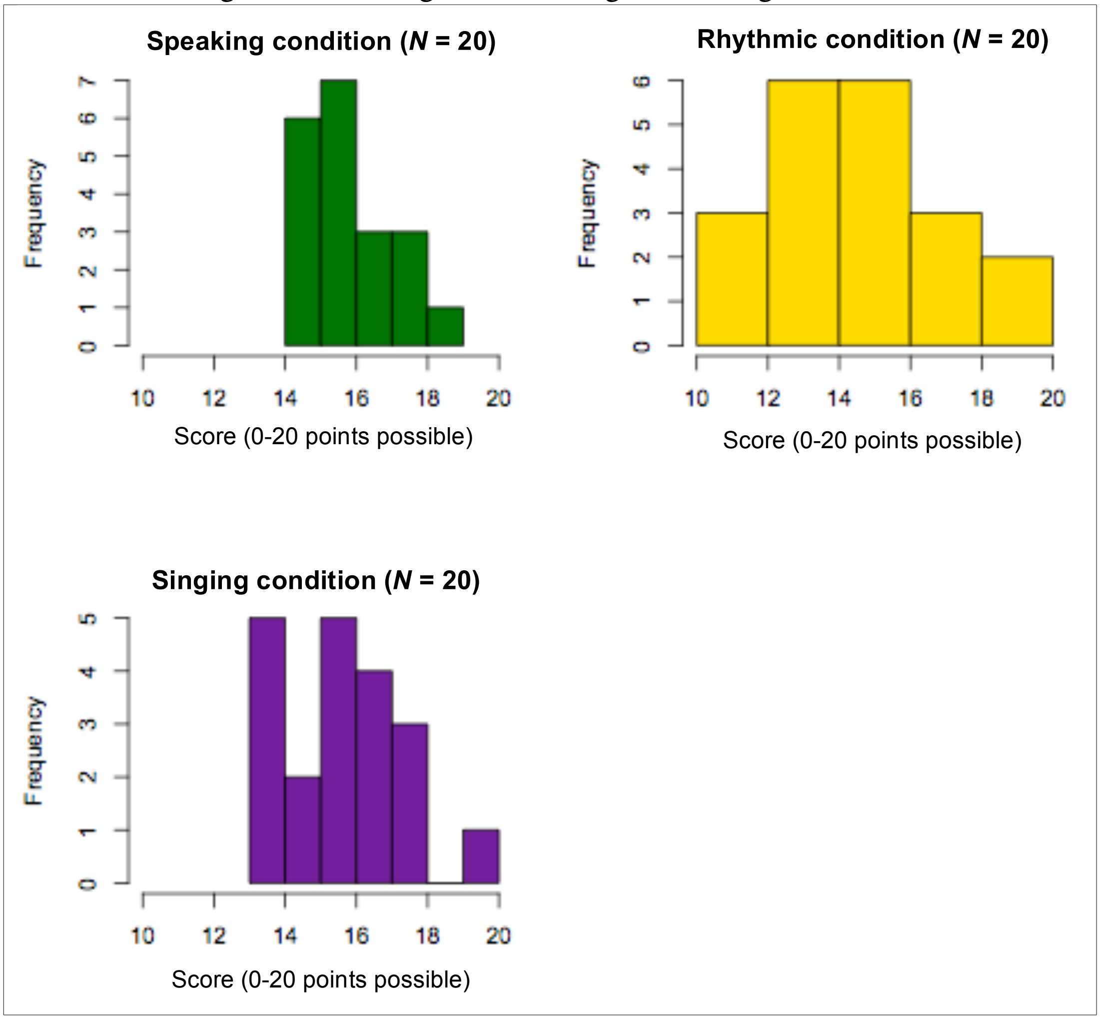


Figure A.9: Density plots for Hungarian Recognition Test

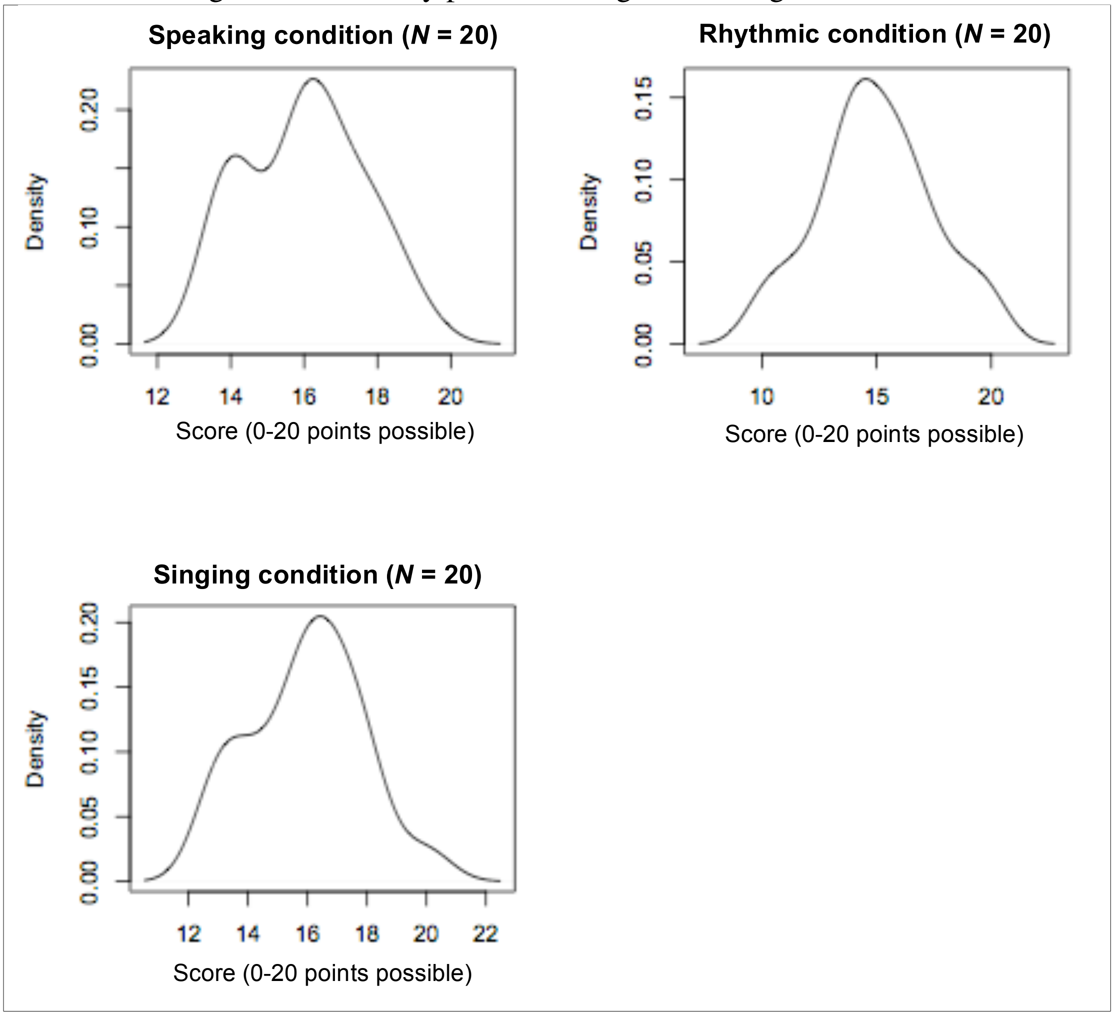


Figure A.10: Histograms for Multiple-Choice Hungarian Vocabulary Post-test

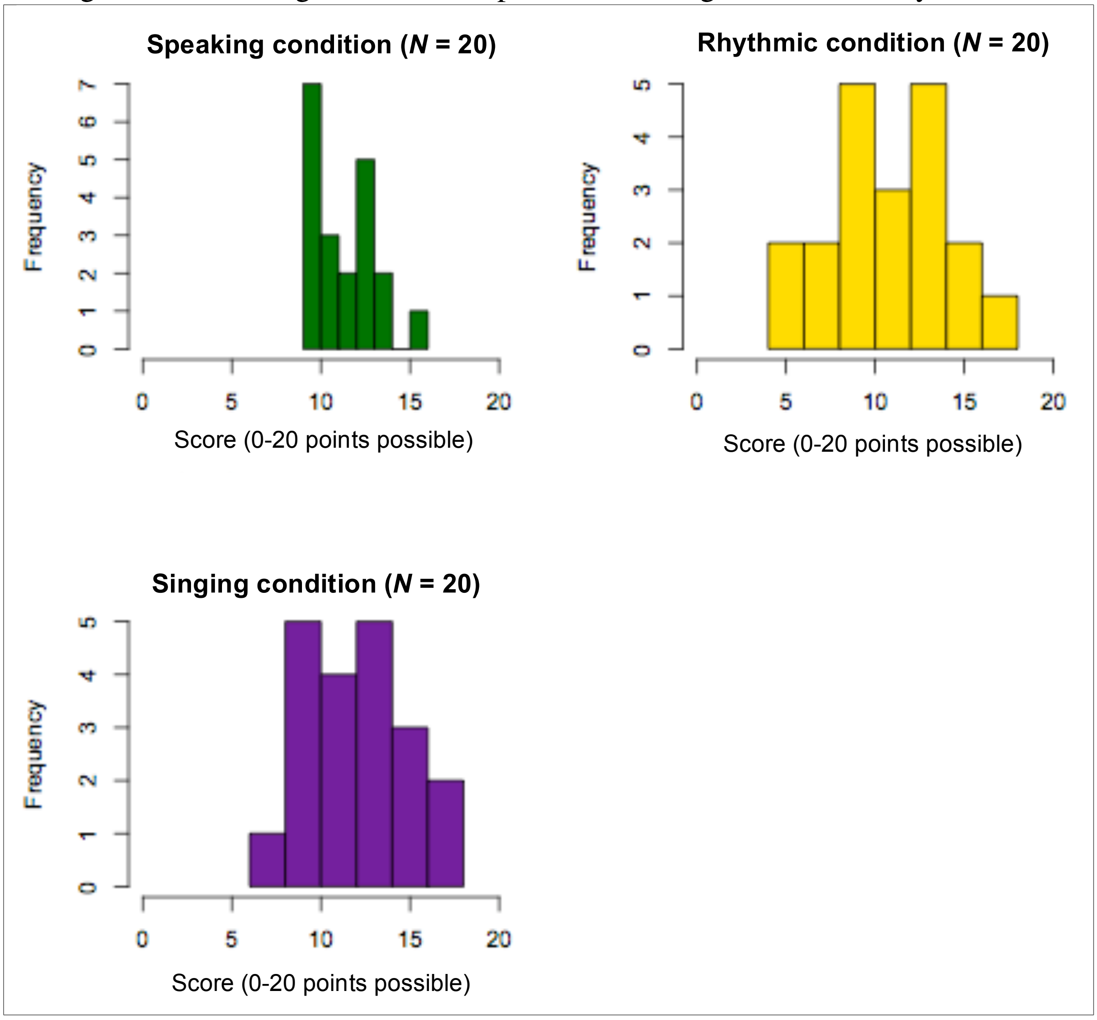


Figure A.11: Density plots for Multiple-Choice Hungarian Vocabulary Post-test

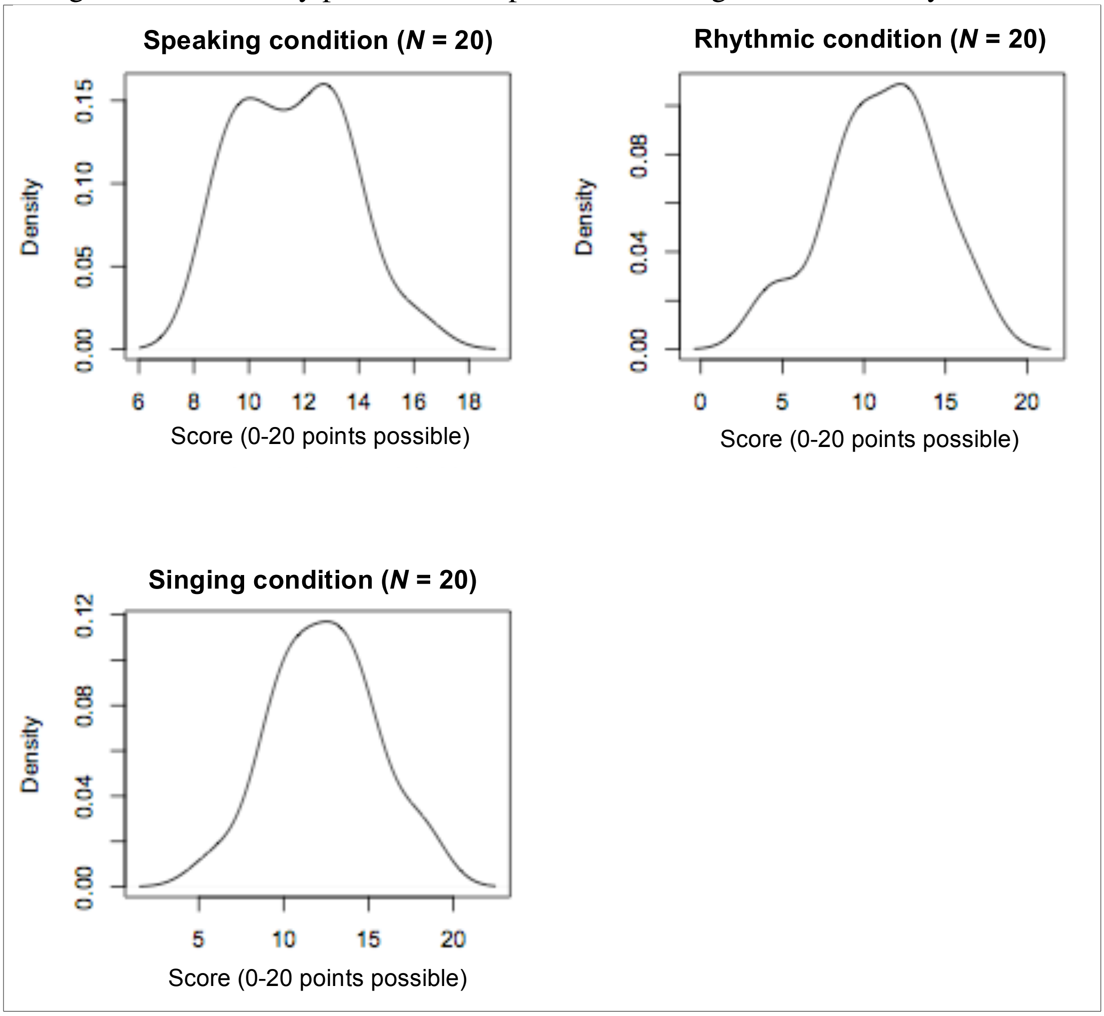


Figure A.12: Histograms for Delayed-Recall Hungarian Conversation

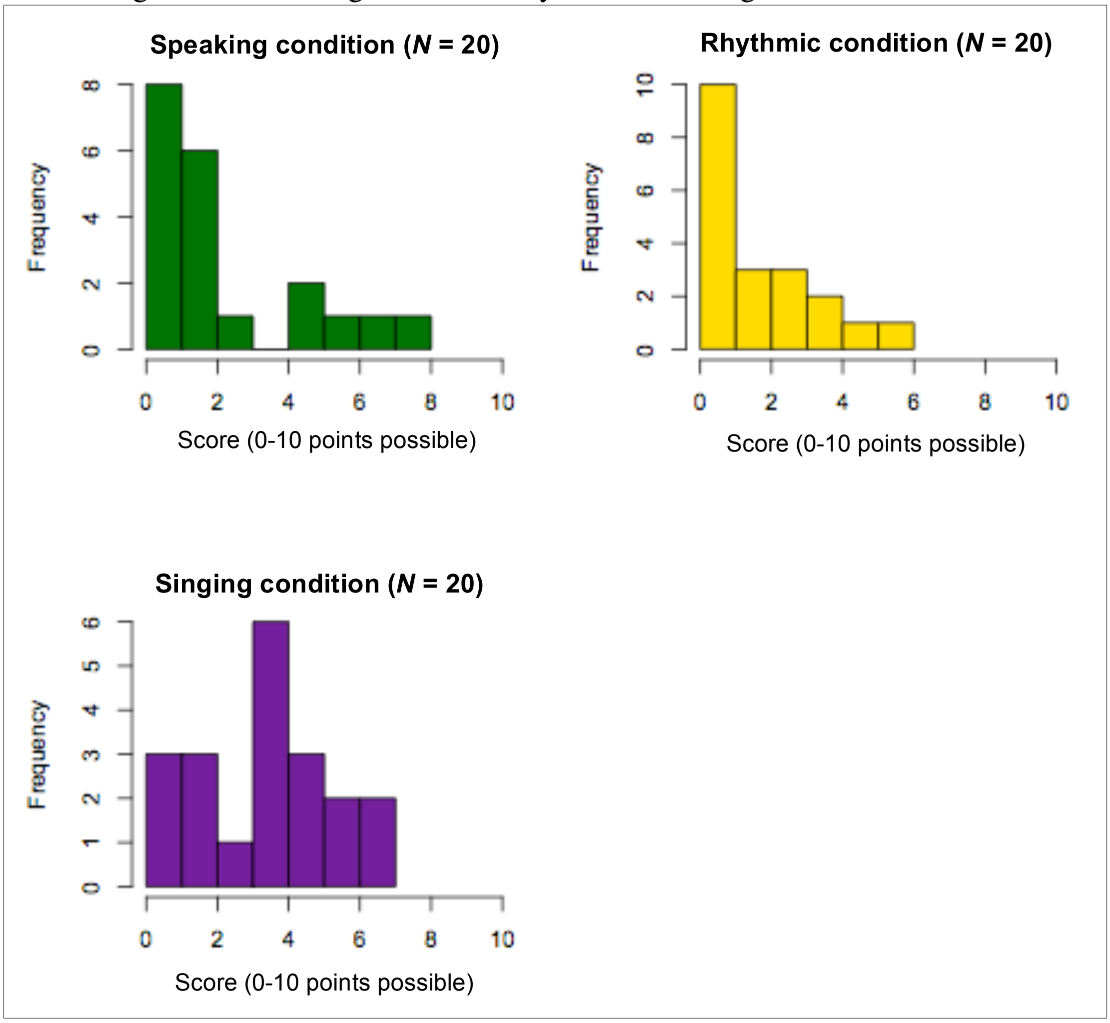
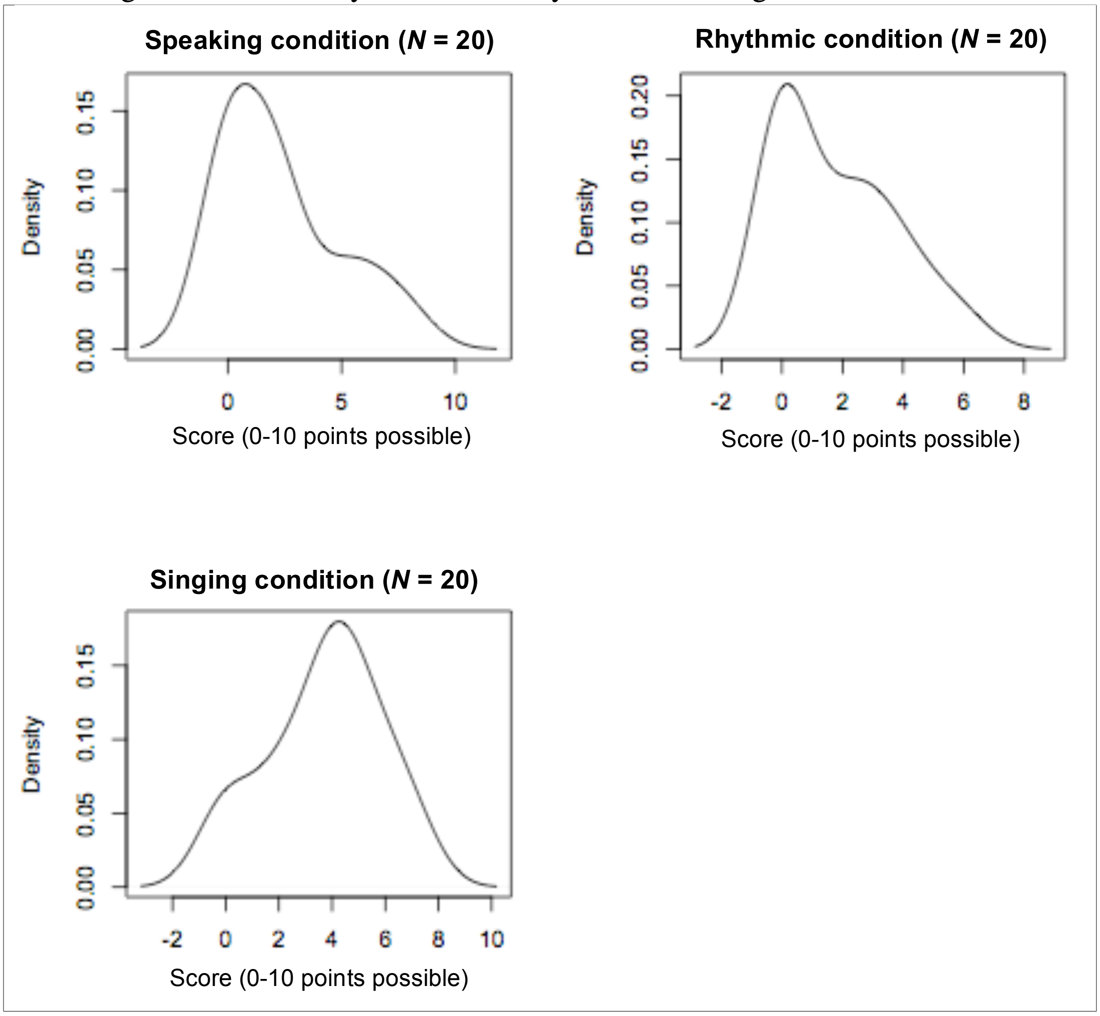


Figure A.13: Density Plots for Delayed-Recall Hungarian Conversation



Appendix B

Supplemental Materials for Hungarian Study 2

B.1 Informed Consent Form

Figure B.1: Informed Consent Form: Information sheet

University of Edinburgh
Auditory Memory Study
Researcher: Karen M. Ludke

Information for research study participants

You are invited to participate in a research study that investigates how individuals' auditory memory functions when learning phrases in a foreign language.

If you choose to participate in this study, you will be asked to learn a series of 20 short words and phrases in a language you do not know. You will be asked to listen to the phrases while trying to remember the English meaning of each phrase you hear. We will make a high-quality audio recording of your session. We would like your permission to use the audio we will record, which will include your voice, for later analysis and comparison with performance by other participants. Your individual privacy will be maintained in any published, educational, written, and electronic reports and data resulting from the study.

You are free to call a halt at *any* stage

You have no obligation to take part. Your participation is voluntary and you are free to withdraw your consent or to discontinue participation at any time — before, during, or after the session — without penalty. You have the right to refuse to answer particular questions.

What do we need from you?

For this study, the most important task is for you to listen to a recording of English words followed by the same phrases in an unfamiliar language and to try to remember the English meaning. We also need you to do a few tests related to what you have heard and practised, to complete a few additional computer-based activities, and finally to respond to a brief questionnaire. The total amount of time we will need your participation is approximately 75 minutes. Beyond the remuneration that you will receive (£7), there are no long-term benefits to participation.

What is this research for and what will we do with the data?

We are interested in the effects that different auditory presentations of material will have when people are learning to say phrases in a foreign language.

The audio recordings of all participants' voices will be used as data in this research. Recordings will be used only for research and educational purposes. On the Consent Form, you may choose whether or not to allow your anonymous recording to be added to the recordings archive for possible use for future research and teaching purposes.

We hope to publish the results of the research in the future in books, journals, and on the Internet, where appropriate. In publications, we will not identify individual participants in any way, and your Consent Form will be kept securely. Complete confidentiality will be respected at all times.

If you are willing to participate in this study, please fill in and sign the attached Consent Form and return it to the researcher. Please keep this information sheet. If you have further questions, you are encouraged to contact me at k.ludke@sms.ed.ac.uk or at the address below.

Karen M. Ludke
University of Edinburgh, IMHSD
12 Nicolson Square, Alison House
Edinburgh EH8 9DF

Figure B.2: Informed Consent Form: Signature sheet

University of Edinburgh
Auditory Memory Study
Researcher: Karen M. Ludke

Informed Consent Form for Study Participant

By signing below, I confirm and agree to the following statements and conditions:

- Δ I agree to participate in this research.
- Δ I have carefully read this Consent Form and the accompanying information sheet, which together give information about the research, the session, and the recording. I have had the opportunity to ask questions about the study.
- Δ I understand that I am under no obligation to take part in this study and that a decision not to participate will not be a problem.
- Δ I understand that I have the right to withdraw from this study at any stage.
- Δ I understand that this is educational, non-therapeutic research from which I cannot expect to derive any long-term benefit.
- Δ I am a native English speaker with normal hearing who is over the age of 18, or the parent/guardian of a participant younger than 18 years old.
- Δ I permit the researcher to keep my computerised and audio-recorded data, with all personally identifying information removed, in a digital archive for possible future research and/or teaching purposes.

Name: ____

Email address: ______________________________________

Signature: ____

Today's Date: ___________________

Please address any questions regarding this consent form to:

Karen M. Ludke
University of Edinburgh, IMHSD
12 Nicolson Square, Alison House
Edinburgh EH8 9DF
k.ludke@sms.ed.ac.uk

B.2 Script for Hungarian Study 2

Introduction and brief small talk to put participant at ease.

Hi, I'm Karen. It's nice to meet you. Thank you for coming. How are you? [Beautiful/Awful] weather today, isn't it?

We begin walking downstairs to the experiment room.

I'm not sure if you remember what this study is about, but it's the auditory memory study. You will hear some words and phrases in a language which you hopefully don't already know, and also the English translation of those phrases. What we need you to do is to try to learn the phrases in the new language and to try to remember the English meaning of the phrases you hear, as best you can. There are a few other activities on the computer as well, but that's the main point.

We enter the room where the participant will do the experiment session.

Have a seat there and make yourself comfortable. If you could read through that sheet and then fill in the second page, that would be great.

Participant fills in the Informed Consent Form and the researcher answers any questions.

That's great, thank you. So, first I'm going to turn this on.

Researcher presses 'record' on the digital audio recorder.

Great, it looks like that is working properly. Now, the first thing we're going to do is I'm going to read some fake words out loud to you. You don't need to remember these words, and they don't mean anything. Just try to repeat them as best you can. Since we have a different accent, just try to say the same word, even if it has a different pronunciation from the way I've said it. Does that make sense?

Researcher waits for confirmation/agreement.

Okay, so the first word is...

Researcher reads the 20 low-wordlike items on the children's non-word repetition task, pausing after each one to allow the participant to repeat the word; and repeats individual words again, once, if requested.

Great, thank you very much. Next is a nonverbal reasoning task. You can fill in your answers on this sheet.

Researcher conducts the nonverbal reasoning DAST sub-test (8 items) using the official procedure (Fawcett & Nicolson, 1998).

That's great, thank you. So, next we're going to turn on the computer screen and if you could just read the instructions there – it basically says the same thing as the sheet you read before – and then start going through the pages. At some point

you'll reach a screen which asks you to stop and get me, so when you reach that point, let me know. It's just to make sure you understand the instructions for the next part. I'll be right here so you can feel free to ask me any questions.

Researcher looks at some papers in the corner where the computer screen is not visible while the participant begins the computer-based component of the experiment, entering basic information, the pre-session PANAS mood questionnaire, language and musical ability pre-tests, and the multiple-choice Hungarian vocabulary pre-test; participant stops when he/she finishes the pre-test and says they've completed that part..

Okay, great. If you could click to the next page, that one has the instructions on it.

Participant reads the instructions and indicates that he/she understands or else asks questions until everything is clear and he/she is ready to proceed.

All right, if you could just put on these headphones now and make sure they feel comfortable.

Participant does so, and when ready, clicks to the next page (practice session with written words on screen). Participant repeats the three practice phrases aloud and the researcher gives additional explanation if necessary.

Extra explanation if participant is in the Listen-and-Repeat Speaking condition:

Yes, that's great, just try to repeat what you hear in the recording. You're not being judged on how perfect it sounds. Just try your best.

Extra explanation if participant is in the Listen-and-Repeat Singing condition:

Yes, that's great, and try to repeat what you hear with the same rhythm and the same notes that you hear in the recording. You're not being judged on how perfect it sounds. Just try your best.

Extra explanation if participant is in the Listen to Speech or Listen to Singing conditions:

This is a listening task, so it's very important that you don't vocalise – don't try to whisper or move your mouth as you learn the phrases.

Great, that's right. Now, the next practice page will be like the second, third, and fourth listening/learning sessions. It's exactly the same thing, except that this time you won't see the words written on the screen, you'll only hear them.

When ready, participant clicks to the next page (practice session without words).

Great, so, there will be four learning sessions, which will be just like what you just did here. The learning sessions will be followed by three tests. There's a lot

to remember, so if you can only remember one word or part of the phrase, then just say that – it's not all-or-nothing. Just try your best.

The tests will be followed by some other activities on the computer, so follow the instructions on the screen until you reach the page at the end that says that you're finished. I'll be in the room just outside if you have any questions. Otherwise, just come and find me when you're done.

The researcher waits for agreement, then stands up to leave.

Great, best of luck, and thank you very much.

Researcher leaves the room.

After participant has finished the experiment and come to get the researcher, we re-enter room.

Finished? Great, thanks. Now, if you wouldn't mind, could you please read this sheet and then fill in this brief questionnaire. Meanwhile, I'll get your money.

The participant reads the Debriefing form and fills in the final questionnaire, and the researcher puts their money on the desk nearby.

Great, thank you very much.

The researcher leads the participant upstairs to the lobby.

B.3 Debriefing sheet – Study 2

Auditory Memory Study Spring 2009

Auditory Memory Study Debriefing

This sheet explains the purpose of the study you just participated in.

Please do not tell others what you did during the session or share this information about the research, because it could affect their performance if they also wish to take part.

Thank you very much!

You have just taken part in a *very* difficult auditory memory study. The average score on the tests is less than 20%. The 20 phrases you heard were spoken by a native speaker of Hungarian. Hungarian is one of the hardest languages in the world to learn and this experiment was made very difficult on purpose, to investigate how people's auditory learning functions when phrases in this challenging language are paired with the English meanings.

In particular, we were interested in whether or not learning might be improved by hearing different auditory presentations of the material and whether practising the phrases aloud might support learning when compared to only listening to the phrases. Some research studies have indicated that melodies may help people remember words in their native language, so we investigated whether or not memory can be improved when learning words and phrases in a foreign language. By looking at people's performance in the different learning conditions, we hope to explore whether a musical presentation of material might facilitate memory when learning a new language, compared to hearing the same phrases presented as speech.

We were also investigating whether performance on this short-term auditory learning and memory task would be influenced by certain differences between the people who participated. We were interested in the extent to which previous language learning experience, musical background, mood, and nonverbal reasoning might influence learning and performance.

If you have any questions about this study, please feel free to ask the lead researcher, Karen Ludke, or send her an email (k.ludke@sms.ed.ac.uk).

Thanks a lot for your participation!

B.4 Debriefing questionnaire – Study 2

1. Before beginning this study, did you know the meaning of any of the words you heard in this new language? If so, could you say how many (or which) words you already knew?
2. What do you think the experiment was about?
3. Overall, how motivated did you feel to learn the phrases in this new language? (Please circle one and/or write a few words describing your level of motivation.)
4. Did you notice any change in your motivation to learn the phrases at the beginning compared to your motivation in the middle or at the end of the learning phase? If so, could you describe the difference?
5. How successful do you think you were at learning to say the Hungarian phrases? (Please circle one and/or write a few words describing your overall opinion.)
6. How successful do you think you were at learning the English meanings of the phrases? (Please circle one and/or write a few words describing your overall opinion.)
7. Do you believe that music and songs can support learning and/or memory?
8. Would you suggest any changes or any ways of improving the learning experience for people participating in future studies?
9. Do you have any other comments or suggestions?

B.5 Hungarian test histograms and density plots – Study 2

Figure B.3: Histograms for Hungarian Production Test – Study 2

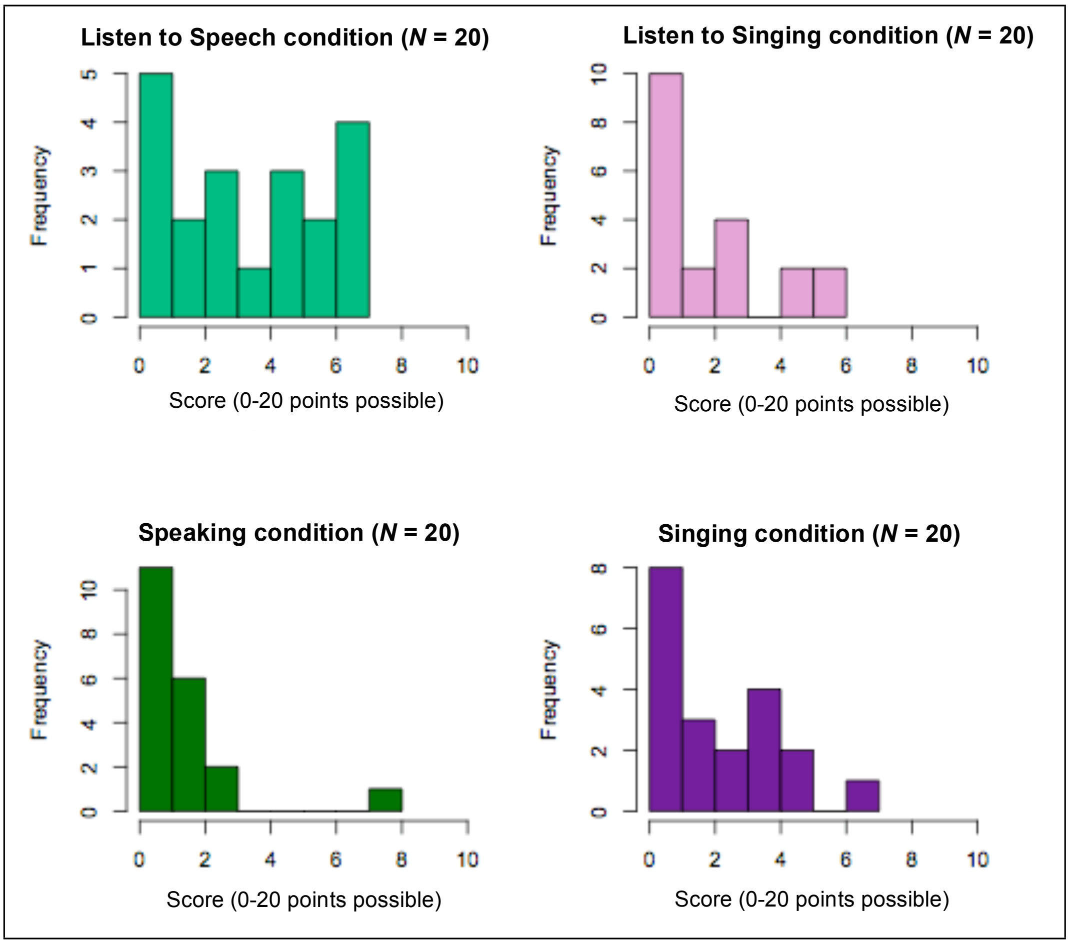


Figure B.4: Density plots for Hungarian Production Test – Study 2

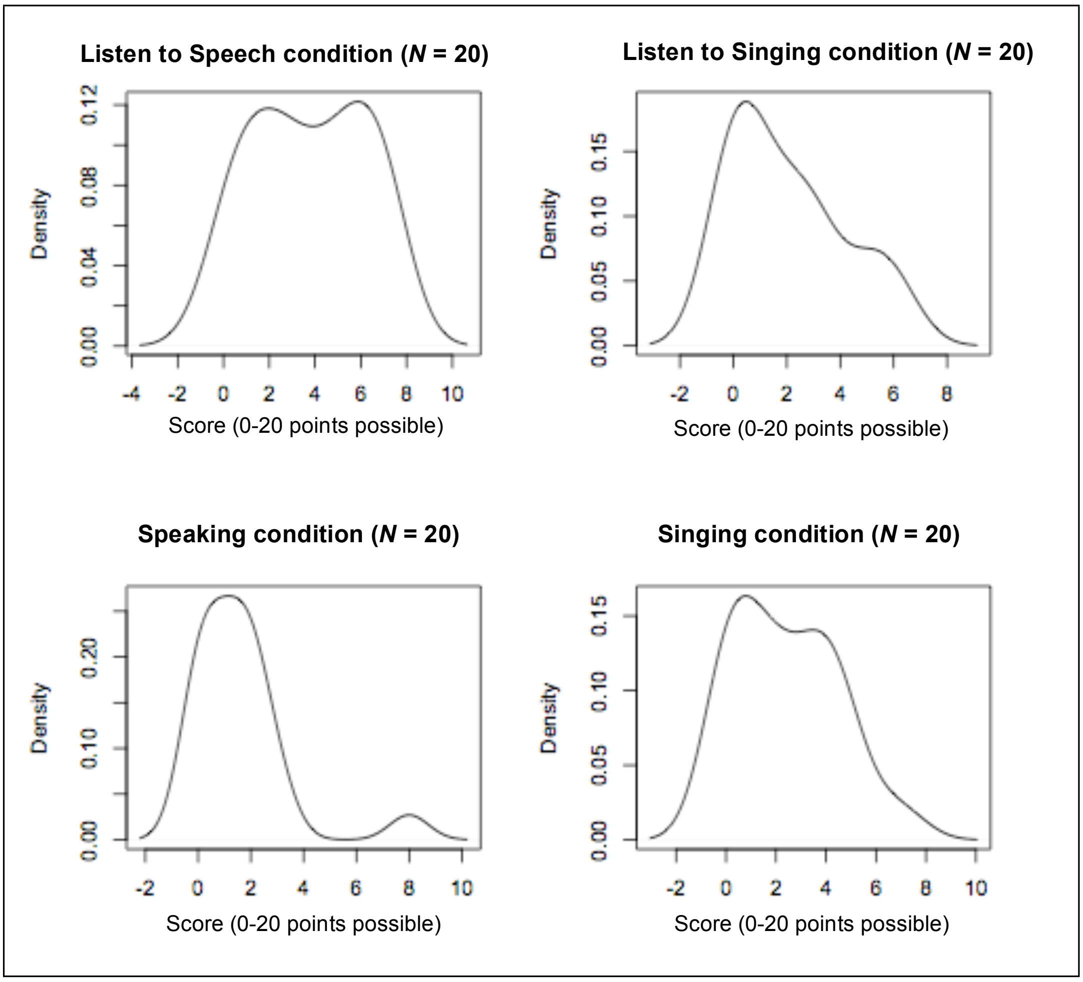


Figure B.5: Histograms for English Recall Test – Study 2

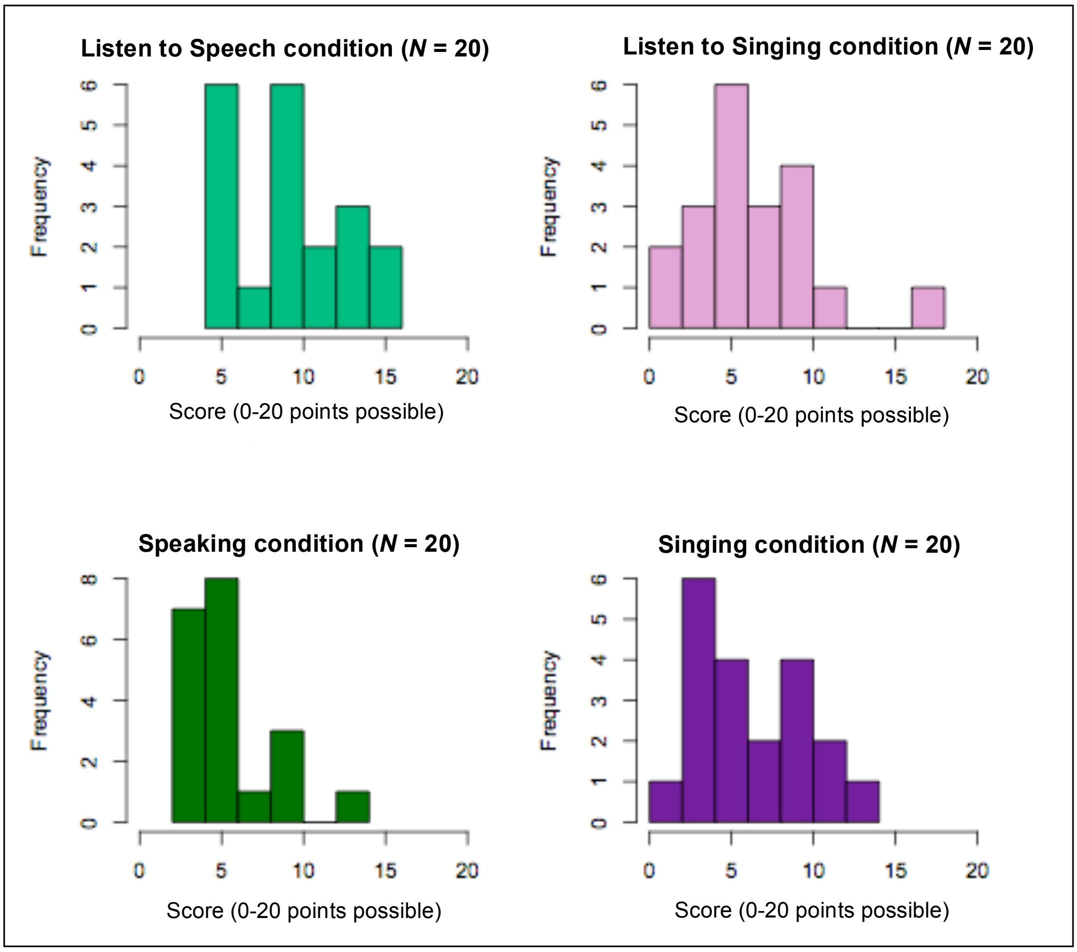


Figure B.6: Density plots for English Recall Test – Study 2

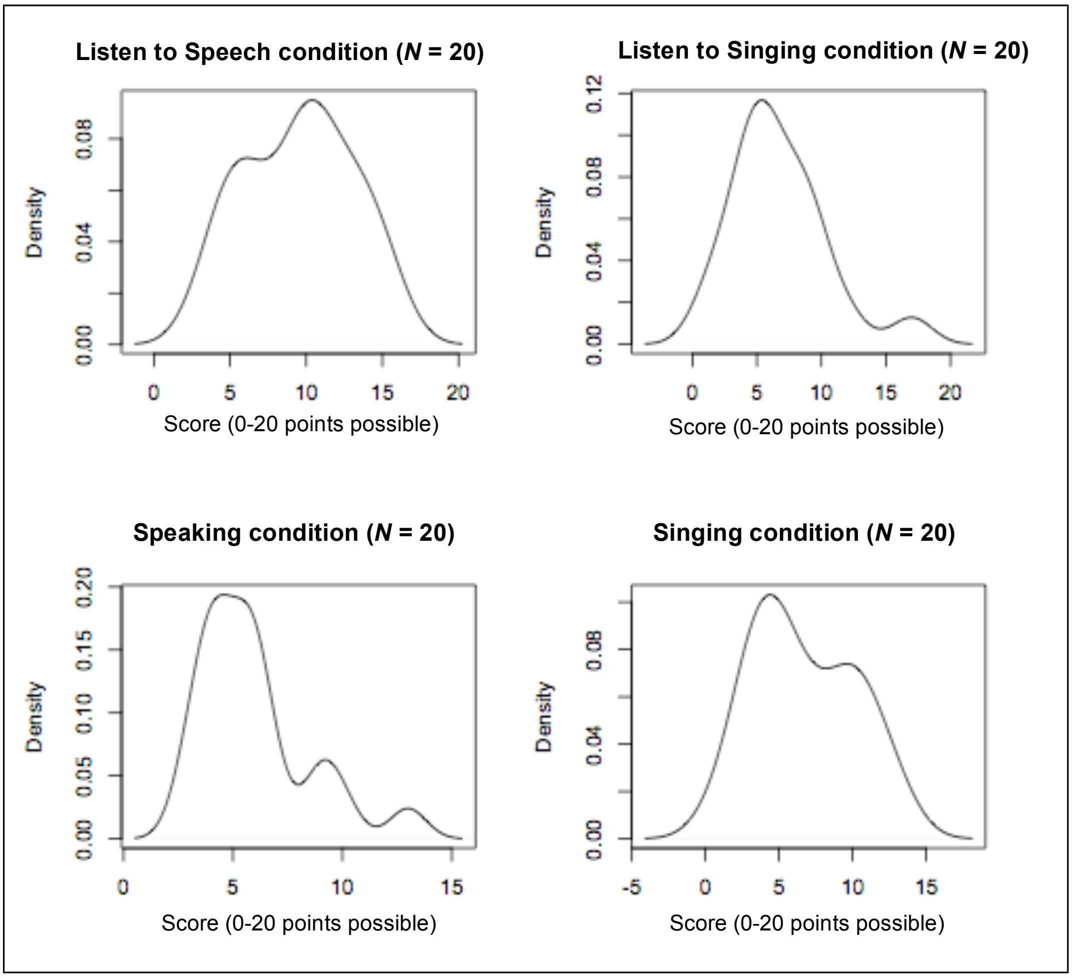


Figure B.7: Histograms for Hungarian Recognition Test – Study 2

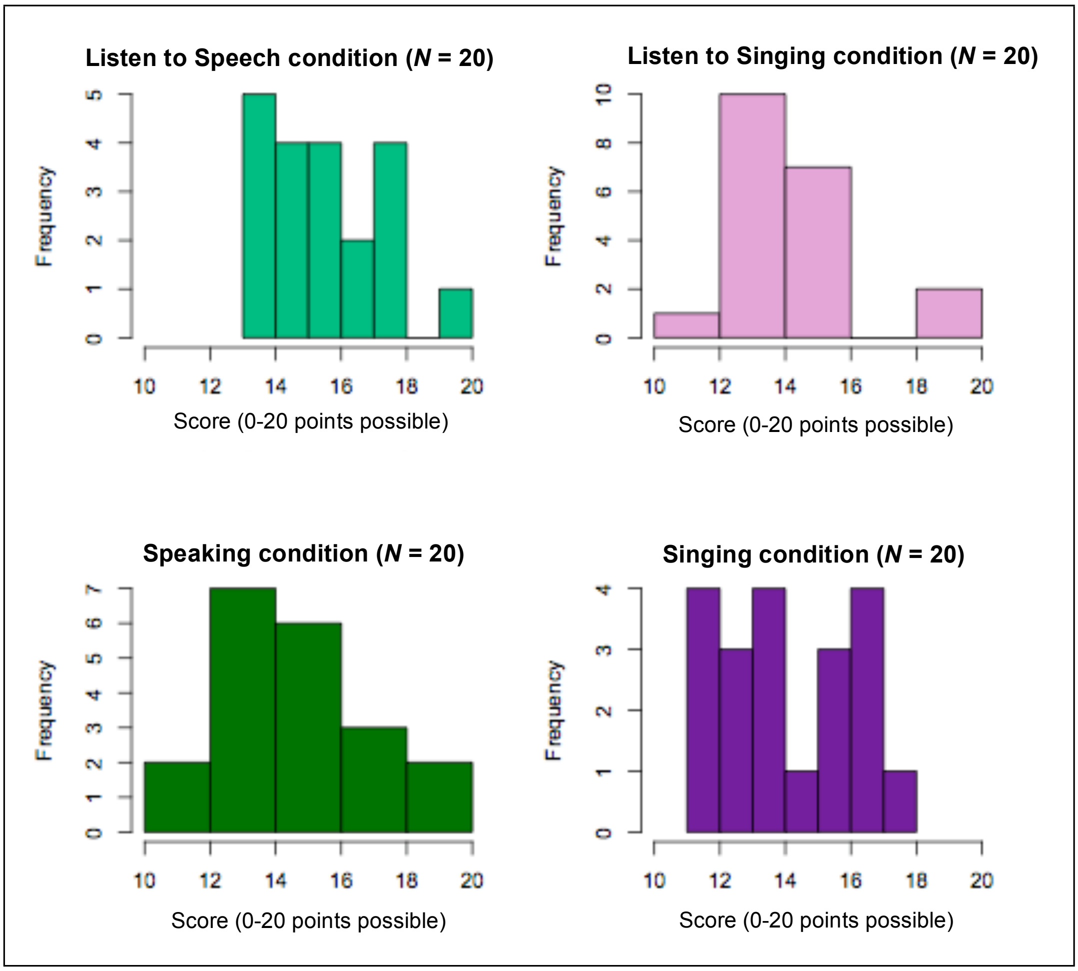


Figure B.8: Density plots for Hungarian Recognition Test – Study 2

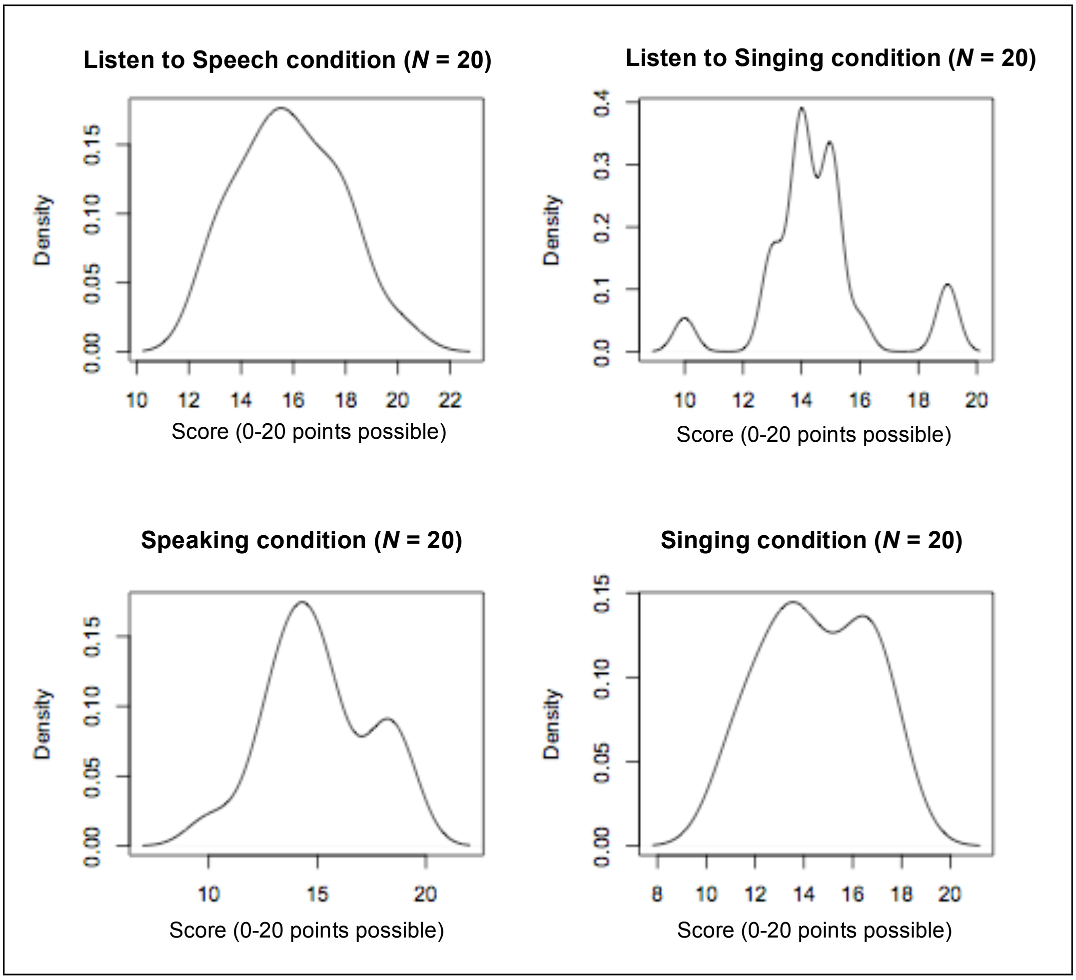


Figure B.9: Histograms for Delayed-Recall Hungarian Conversation – Study 2

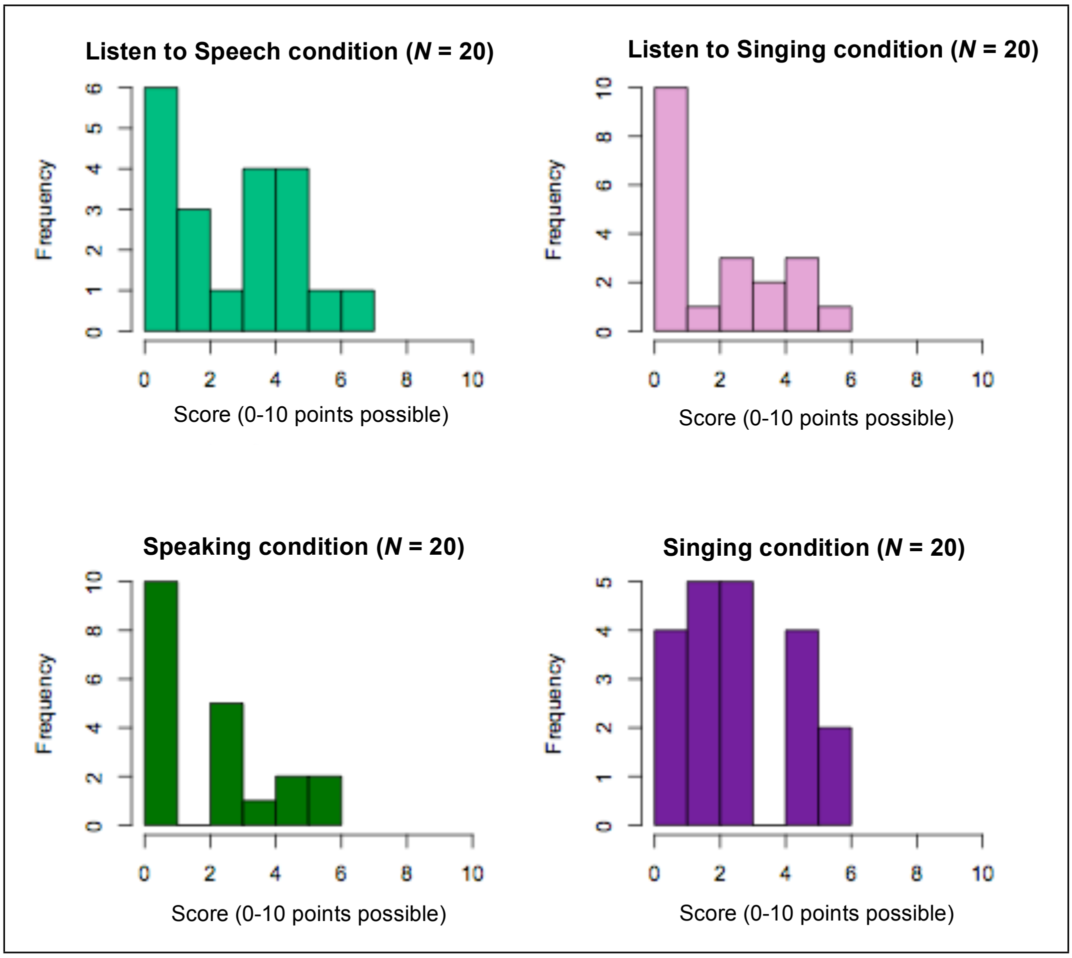
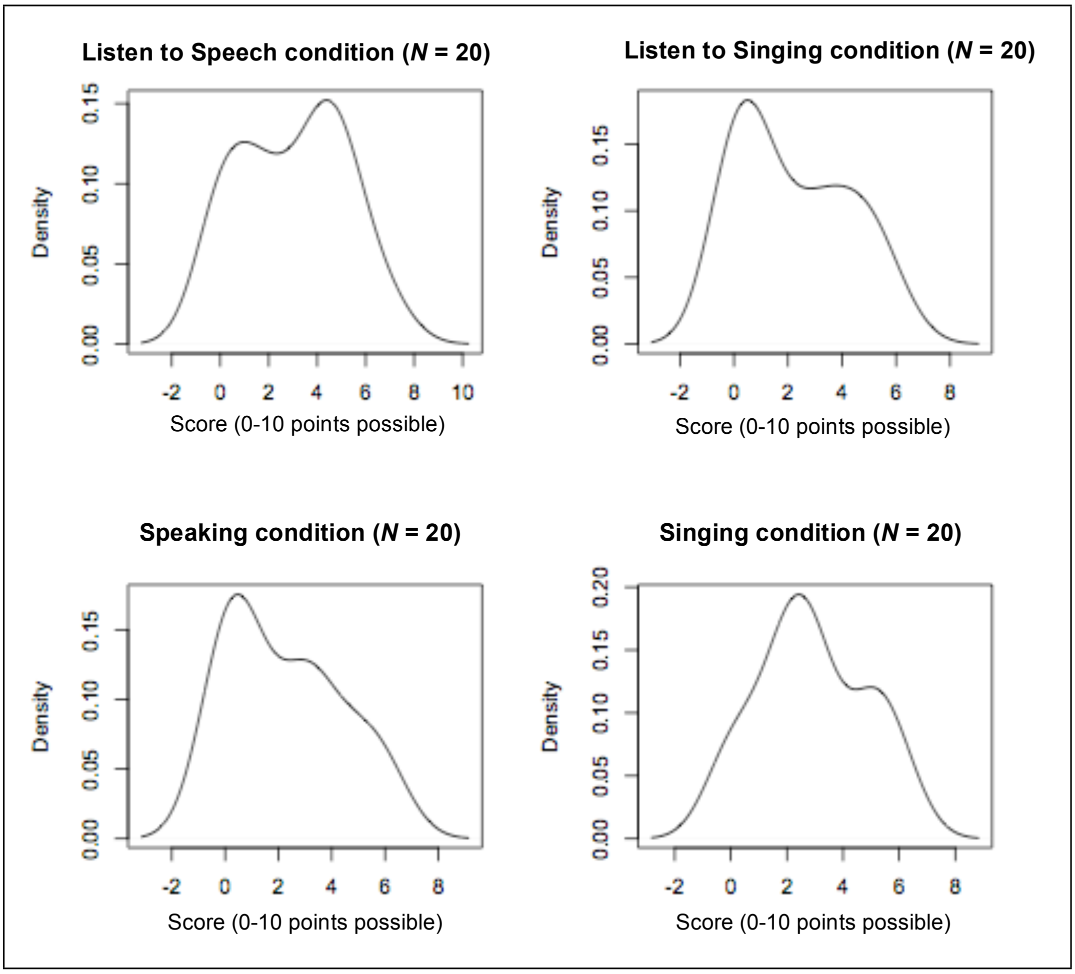


Figure B.10: Density plots for Delayed-Recall Hungarian Conversation – Study 2



B.6 ANOVA for the four groups, comparing the results of the two Listen to Speech groups – Study 2

This section provides a summary of the results of the passive Listen to Speech group that only completed three (rather than all four) Hungarian tests in the second experimental study. These 20 participants took part in the experiment during the exam period – which was the same as for most participants the other three groups, and different from the participants in the passive Listen to Speech condition whose results were reported in Chapter 4.

Similar to the ANOVA table found for the four groups that completed all four Hungarian tests, when comparing ID scores in the four groups that only completed three Hungarian tests, the listen-and-repeat Singing condition again tended to have lower performance compared to the other three groups for productive music skills, $p = .06$. Table B.1 shows the ANOVA for the ID measures in the four groups that only completed three Hungarian tests and for comparison, Table B.2 shows the ANOVA scores for the four groups that completed all four Hungarian tests (as presented in Chapter 4).

Table B.1: ANOVA for ID measures in the four learning conditions with an incomplete data set (only three Hungarian tests, which did not include the Hungarian Production Test)

ID Measure	N	df	Sum Sq.	Mean Sq.	F-stat.	p-value
Age	80	3, 76	17.90	5.97	1.009	.39
Nonverbal Reasoning (DAST)	80	3, 76	3.70	1.23	0.826	.40
Phonological WM	80	3, 76	9.20	3.07	1.155	.33
Language Experience Quest.	80	3, 76	0.07	0.02	1.278	.29
MLAT-style Pre-test Total	80	3, 76	77.10	25.70	1.453	.23
MLAT-style Pre-test Structure	80	3, 76	1.10	0.37	0.140	.94
MLAT-style Pre-test Memory	80	3, 76	65.70	21.90	1.629	.19
MLAT-style Post-test Total	80	3, 76	22.05	7.35	0.494	.69
MLAT-style Post-test Structure	80	3, 76	5.70	1.90	1.343	.27
MLAT-style Post-test Memory	80	3, 76	14.05	4.68	0.400	.75
Musical Experience Quest.	80	3, 76	0.01	0.00	0.097	.96
Musical Ability Test Total	80	3, 76	48.54	16.18	2.082	.11
MAT Receptive Total	80	3, 76	27.25	9.08	1.931	.13
MAT Rhythm Discrimination	80	3, 76	8.24	2.75	1.606	.20
MAT Pitch Discrimination	80	3, 76	4.30	1.43	2.029	.12
MAT Melody Discrimination	80	3, 76	3.74	1.25	1.373	.26
MAT Productive	80	3, 76	15.60	5.20	2.530	**.06**
Happy Birthday: Sing	80	3, 76	4.50	1.50	2.123	.10
Happy Birthday: Tap	80	3, 76	4.50	1.50	1.534	.21
Positive Mood Pre-session	80	3, 76	27.80	9.27	0.203	.89
Negative Mood Pre-session	80	3, 76	92.85	30.95	1.267	.29
Positive Mood Post-session	80	3, 76	121.84	40.61	0.735	.53
Negative Mood Post-session	80	3, 76	26.64	8.88	0.453	.72
Motivation to learn phrases	80	3, 76	1.84	0.61	0.684	.56
Change in Motivation	80	3, 76	3.48	1.15	1.968	.13
Success: Hungarian phrases	80	3, 76	2.25	0.75	1.691	.18
Success: English phrases	80	3, 76	1.25	0.42	0.749	.53

Table B.2: ANOVA for ID measures in the four learning conditions with a complete data set (completed all four Hungarian tests)

ID Measure	*N*	df	Sum Sq.	Mean Sq.	*F*-stat.	*p*-value
Age	80	3, 76	20.1	6.7	1.179	.32
Nonverbal Reasoning (DAST)	80	3, 76	3.1	1.0	0.594	.62
Phonological WM	80	3, 76	15.8	5.3	2.104	.11
Language Experience Quest.	80	3, 76	187.3	62.4	1.597	.20
MLAT-style Pre-test Total	80	3, 76	66.7	22.2	1.161	.33
MLAT-style Pre-test Structure	80	3, 76	1.1	0.4	0.179	.91
MLAT-style Pre-test Memory	80	3, 76	51.5	17.2	1.227	.31
MLAT-style Post-test Total	80	3, 76	31.8	10.6	0.622	.60
MLAT-style Post-test Structure	80	3, 76	8.7	2.9	1.681	.18
MLAT-style Post-test Memory	80	3, 76	11.5	3.8	0.308	.82
Musical Experience Quest.	80	3, 76	13.7	4.5	0.105	.96
Musical Ability Test Total	80	3, 76	40.0	13.3	1.945	.13
MAT Receptive	80	3, 76	4.5	1.5	0.369	.78
MAT Rhythm Discrimination	80	3, 76	1.5	0.5	0.301	.83
MAT Pitch Discrimination	80	3, 76	4.3	1.4	1.803	.15
MAT Melody Discrimination	80	3, 76	0.3	0.1	0.150	.93
MAT Productive	80	3, 76	22.2	7.4	3.671	**.016 ***
Happy Birthday: Sing	80	3, 76	2.000	.18	0.667	**.035 ***
Happy Birthday: Tap	80	3, 76	5.0	1.7	1.998	.12
Positive Mood Pre-session	80	3, 76	173.4	57.8	1.510	.22
Negative Mood Pre-session	80	3, 76	19.5	6.5	0.492	.69
Positive Mood Post-session	80	3, 76	93.4	31.1	0.593	.62
Negative Mood Post-session	80	3, 76	0.8	0.5	0.012	.99
Motivation to learn phrases	80	3, 76	1.6	0.5	0.714	.55
Change in Motivation	80	3, 76	1.34	0.45	0.761	.52
Success: Hungarian phrases	80	3, 76	1.9	0.6	1.238	.30
Success: English phrases	80	3, 76	1.9	0.6	1.061	.37

Graphs of the mean scores on the three Hungarian tests and overall Hungarian test scores are shown in Figures B.11 to B.18, with data from the two Listen to Speech groups shown on the same page (data from the other three groups are identical).

For the 80 participants who completed the experiment during the exam period, performance on the three Hungarian tests and overall raw Hungarian test scores showed a similar pattern to the results found in the first experiment:

- On the English Recall Test, the sung stimuli conditions (both active and passive learning) tended to slightly outperform the spoken stimuli conditions, but the difference was not significant (see Figure B.11).
- On the Hungarian Recognition Test, the spoken stimuli conditions tended to slightly outperform the sung stimuli conditions, but the difference was not significant (see Figure B.13).
- On the Delayed-Recall Hungarian Conversation, the listen-and-repeat Singing group had higher mean scores compared to the other three groups. This was the only test of verbatim, spoken Hungarian production for the 80 participants who completed the experiment during the exam period. Although the difference in performance on this test was not significant, it was in the direction observed in the first experimental study, with the listen-and-repeat Singing condition outperforming the other groups (see Figure B.15). This suggests that there might be a longer-term verbal memory benefit (at least for a 20-minute delay) for singing during the learning process, even when the learners have weaker productive musical skills.
- For overall Hungarian scores, performance was approximately the same in the four groups. The mean for the Singing condition was slightly higher than in the other three groups, but the difference was not significant (see Figure B.17).

Figure B.11: Mean scores for the English Recall Test (20 points possible) – 4 groups that completed only 3 Hungarian tests

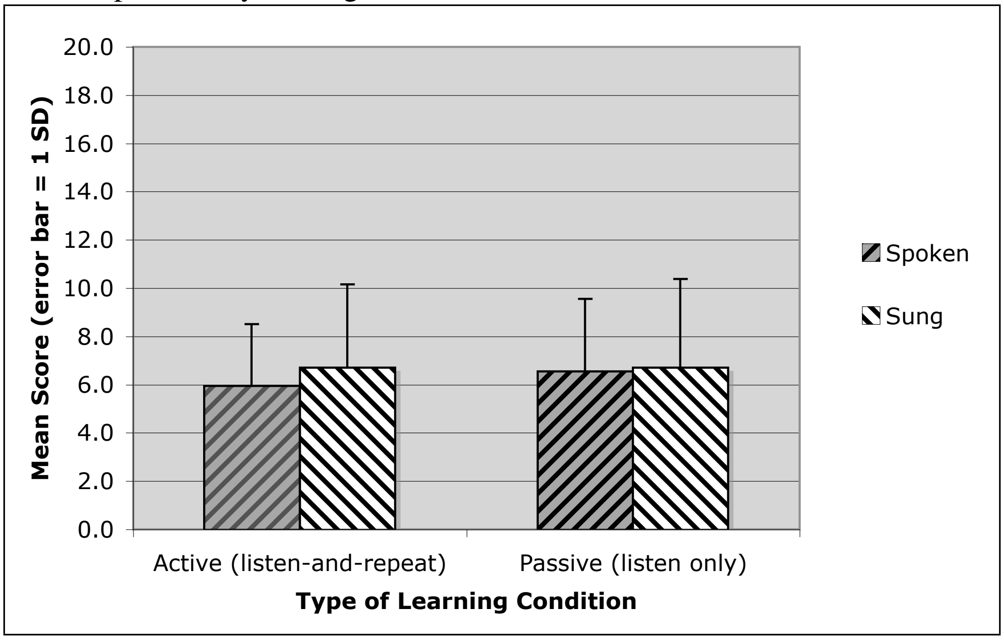


Figure B.12: Mean scores for the English Recall Test (20 points possible) – 4 groups that completed all 4 Hungarian tests

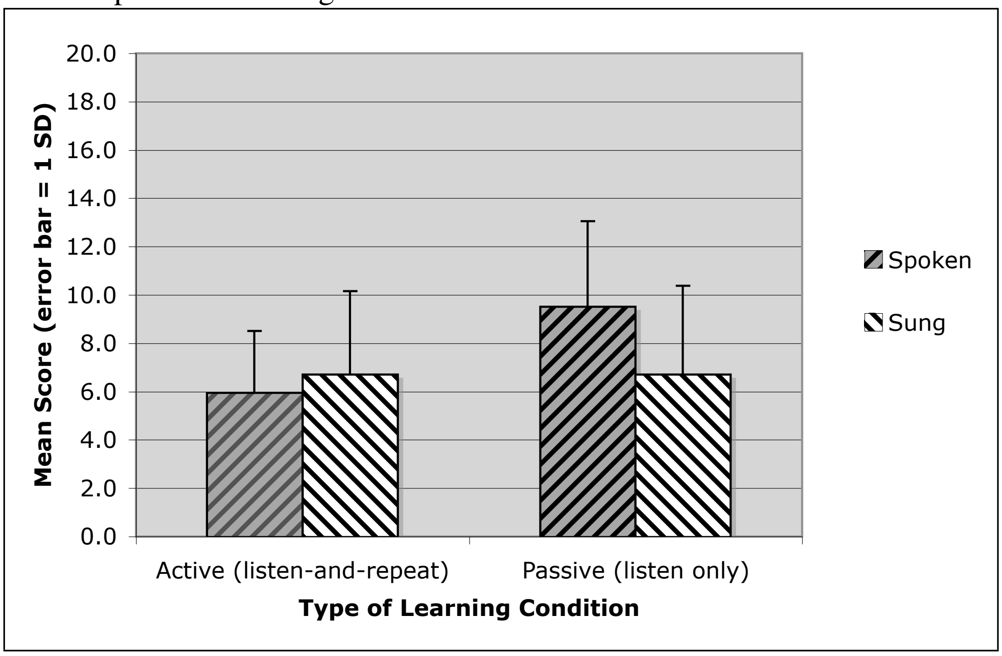


Figure B.13: Mean scores for the Hungarian Recognition Test (20 points possible) – 4 groups that completed only 3 Hungarian tests

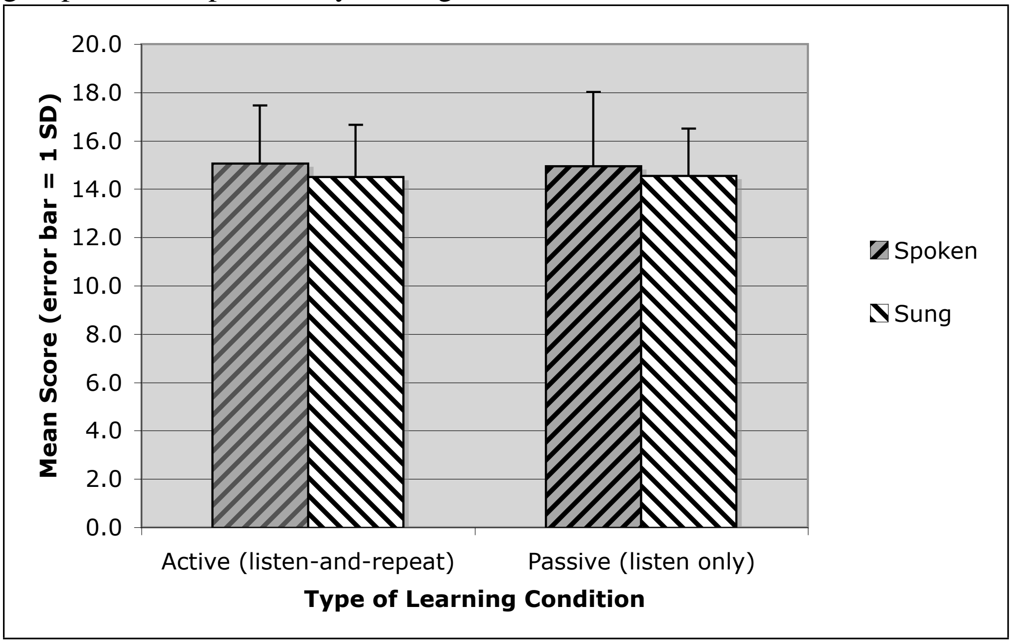


Figure B.14: Mean scores for the Hungarian Recognition Test (20 points possible) – 4 groups that completed all 4 Hungarian tests

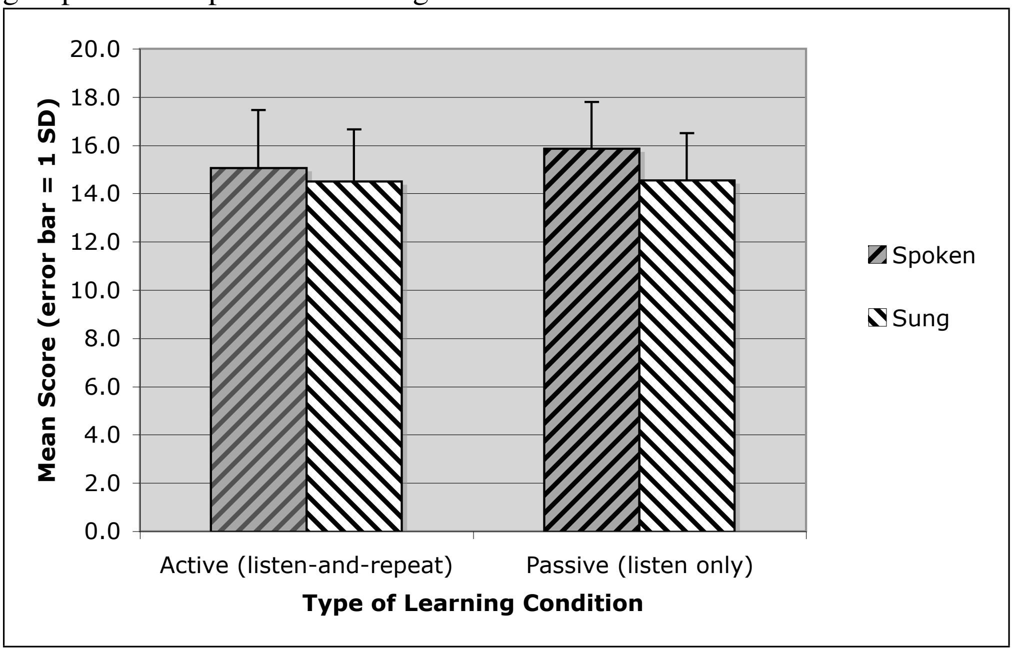


Figure B.15: Mean scores for the Delayed-Recall Hungarian Conversation (10 points possible) – 4 groups that completed only 3 Hungarian tests

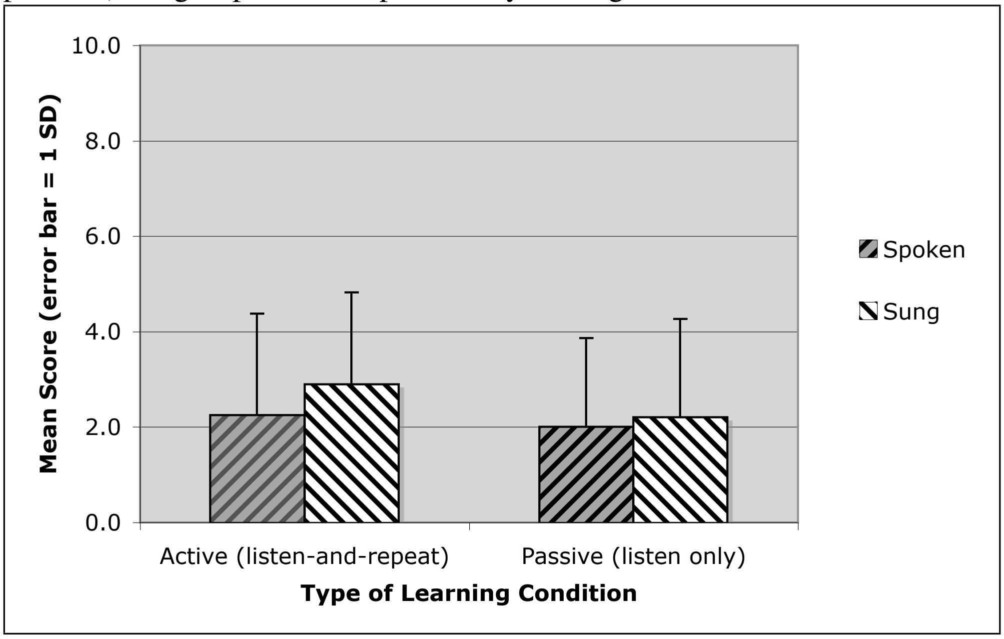


Figure B.16: Mean scores for the Delayed-Recall Hungarian Conversation (10 points possible) – 4 groups that completed all 4 Hungarian tests

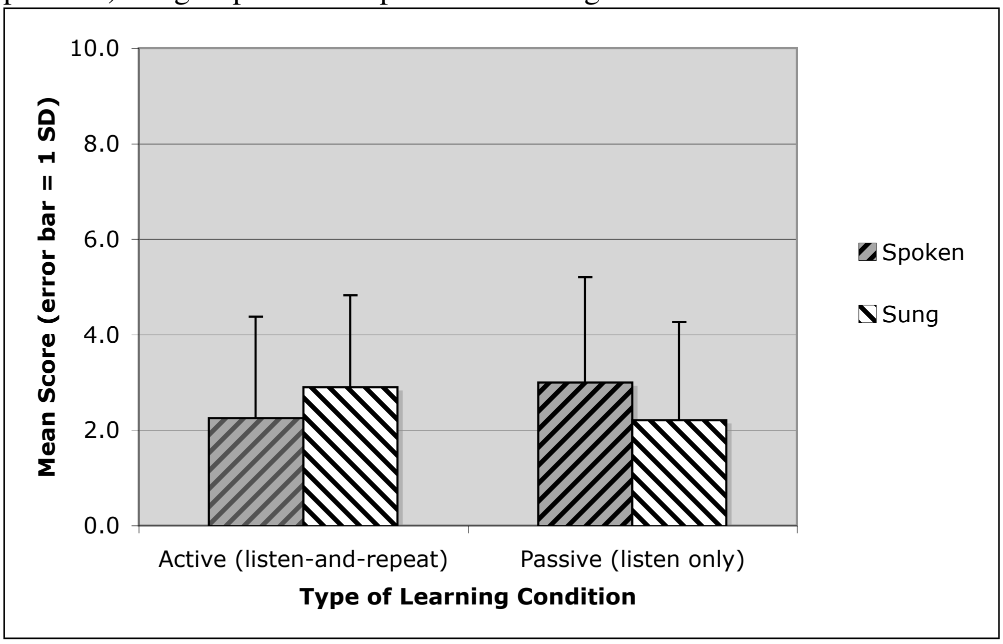


Figure B.17: Mean scores for overall raw Hungarian test score (70 points possible) – 4 groups that completed only 3 Hungarian tests

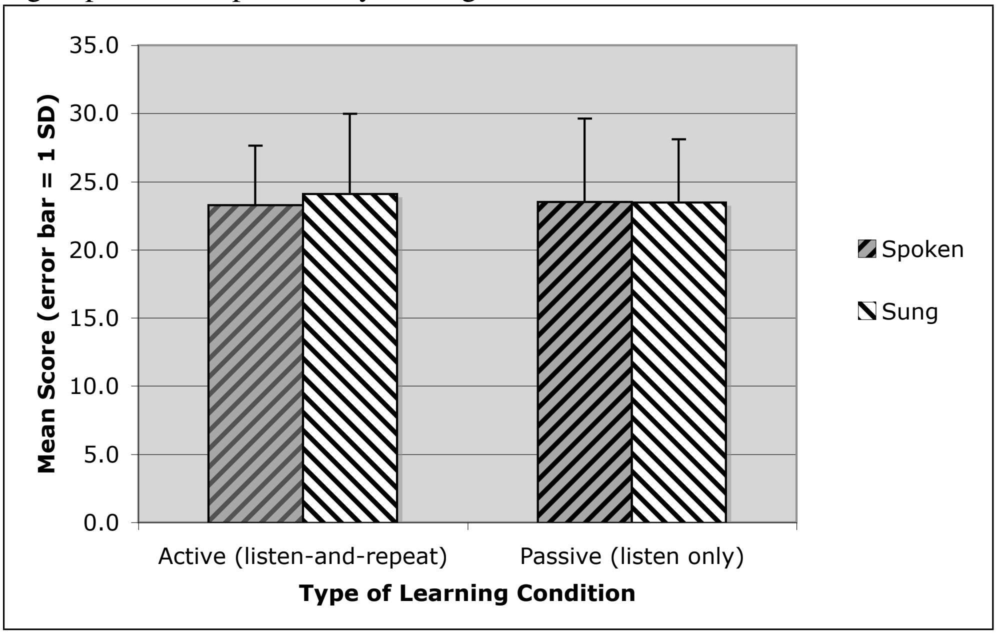
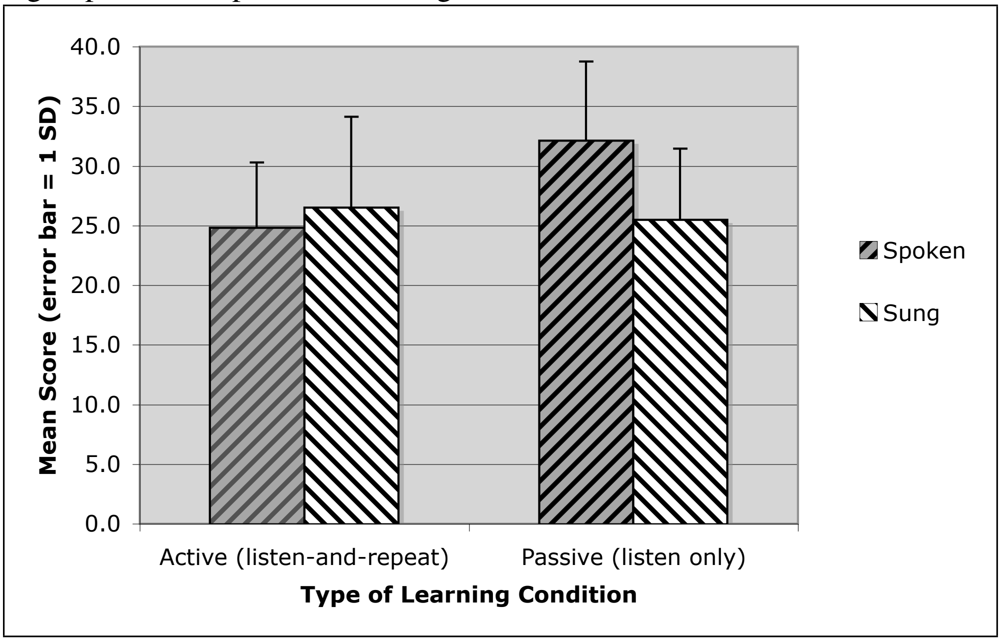


Figure B.18: Mean scores for overall raw Hungarian test score (70 points possible) – 4 groups that completed all 4 Hungarian tests



Taken together, the comparison of results of the two passive Listen to Speech groups with the other three groups suggests that the large differences in Hungarian test performance observed between the passive Listen to Speech group and the other groups in Chapter 4 may have arisen due to factors related to timing and anxiety, or due to other factors which were not captured by the ID measures included in the study.

Appendix C

Supplemental Materials for the Four-Week French Study

C.1 Full lessons and Cloze exercises

C.1.1 Field notes from the French lessons

WEEK 1

Day 1: Lundi 17-Nov-2008 S2 class – Song 'Les Bonbons'

Learning outcome: Practise French listening skills by filling in words missing from the text.

I introduced the song and provided pupils with a short description about Jacques Brel (in French) and talked them through it, mostly in French. Then the children listened twice to the song (filling in the blanks the second time) and I said we would correct it next time. Then I collected their worksheets.

For the rest of class, the children wrote and worked on their posters/games, dramatic dialogues, or French rap. They will practice more tomorrow and then present these projects to the class on Friday. The homework was p. 16 in their workbooks for Friday.

Day 1: Lundi 17-Nov-2008 S1 class – Dramatic 'Dialogue au parc'

Learning outcome: Practise French listening skills by filling in words missing from the text.

I introduced the French dialogue using a bit less detail than I did for the song with the S2 class, and I did not provide a short written description about the characters. The children listened twice to the dialogue (filling in the blanks the second time) and then I said we can correct it tomorrow. Then I collected their Cloze exercise sheets.

During the rest of class, the children completed several activities about different nationalities in their textbook. The homework was p. 3 in their workbooks for Friday.

Day 2: Mardi 18-Nov-2008 S2 class – Song 'Les Bonbons'

Learning outcome: Practise French listening skills and identify themes and facts about the text through listening and reading.

The children listened once to the song while filling in the missing words. Then we went through line by line as I typed the correct answers on the SmartBoard. They had to write the words quickly as I switched back and forth between the words and the recording to play and then pause the recording again. Tonight I will double-check that they have all the words in the correct locations on their Cloze exercise sheets.

Afterward, the children worked in small groups on their dramatic dialogues, games, rap, poster, or comic strips. They will give presentations on Friday during class.

Day 2: Mercredi 19-Nov-2008 S1 class – Dramatic 'Dialogue au parc'

Learning outcome: Practise French listening skills and identify themes and facts about the text through listening and reading.

The children listened once to the dialogue, then I went through the meaning line by line. I asked volunteers to describe what was happening in the dialogue and to give me the English translations of words that they already knew or could guess from the context. Then we listened one final time as I collected their Cloze worksheets. We also chose someone to play the role of the dog for next time.

This was followed by revision of adjectives (sportif/ve, etc.) and 'je suis... et...' and 'non, je ne suis pas...'. The teacher wrote on the board that 'Adjectives are describing words (big, green, funny).' Then they did p. 122 on colour adjectives and an exercise on p. 123. Their homework was to revise colour adjectives and to complete pp. 13-14 in their workbooks.

Day 3: Vendredi 21-Nov-2008 S2 class – Song Les Bonbons'

Learning outcome: Understand overall themes and specific details about the French text.

The children answered some comprehension questions in French about the song that I wrote on the board. Then we did a quick warm-up for our bodies and our voices: shake body around to loosen up, then say 'Bonjour' in different voices (happy, sad, quiet, loud) as a group. We sang each chorus of the song together as they read the words on their sheets, standing next to their desks.

Their teacher took pictures and video while they did their oral presentations in front of the class. At the end of class, there were a few minutes of extra time, so they filled out the teacher's evaluation sheets about the presentations they had given.

Day 3: Vendredi 21-Nov-2008 S1 class – Dramatic 'Dialogue au parc'

Learning outcome: Understand overall themes and specific details about the French text.

Before class started, I wrote a few comprehension questions in French on the SmartBoard about the dialogue and they answered them by writing down their responses and then volunteers told me what they had written. Then we did a quick warm-up for our bodies and our voices: shake body around to loosen up, then say 'Bonjour' in different ways (happy, sad, quiet, loud) as a group. Then we said the words of the dialogue (girls in a line on my left, boys in a line on the right) in a sort of chorus. There was a bit of talking that wasn't relevant to the activity. However, for some reason, when they sat down after we said the dialogue together, the children seemed more focused and quieter, working better than I have seen them do. Perhaps they are becoming accustomed to my presence in the class.

They did a review of colour adjectives and rules for forming them. Then the teacher presented grammar relating to countries: 'J'habite en France' versus 'J'habite au Luxembourg,' for example, and 'aux E-U.' Then they did an exercise on p. 123 in their textbooks, Écoute (1-5 words they have been learning) and Écrire (5 words in French). They had a hard time completing the textbook exercises and the teacher reminded them that it is important to review and revise for 10 minutes each day.

WEEK 2

Day 4: Lundi 24-Nov-2008 S2 class – Song 'Les Bonbons'

Learning outcome: Obtain and provide information by reading the French text and practise French speaking skills.

I asked the children to write three French comprehension questions about the text to exchange with the person sitting next to them, and several questions that I wrote on the board. Then we discussed the correct answers to the questions on the board. However, the children seemed very quiet and tired today. The warm-up exercise before we sang did not go particularly well – even the the 'loud' Bonjour was quiet. The left side of the class sang very well, but the right side of the class was struggling and very quiet. The teacher and I tried to encourage them to sing louder for the chorus.

Then they did revision of passé composé, je n'ai pas/rien fait ce week-end. They completed pp. 6, 7-9 in their workbooks, and had to do page 11 for homework (vacances, futur proche). They also did an exercise about 'que sais-je' (which was due on Friday), textbook exercises on pp. 36-37, 1b, 1c; p. 38, 2a; p. 39, 1 & 2.

They will watch a film next week when the two-week drama section starts, 'Malabar Princesse,' relating to the films unit in their textbook.

Day 4: Lundi 24-Nov-2008 S1 class – Dramatic 'Dialogue au parc'

Learning outcome: Obtain and provide information by reading the French text and practise French speaking skills.

We listened once to the dialogue while they answered three French reading comprehension questions their classmates had written last time. Then they responded to some questions I asked them about the dialogue in French. We did a quick warm-up, but without the loud 'Bonjour' due to a prelim exam in the class next door. A girl played the role of the dog today; they all laugh and enjoy that part. I split the class down the middle; again, the left side (Rachelle) was a bit better than the right side (Sébastian) in terms of saying the words.

They revised how to say what they like and dislike ('j'aime, j'adore...mais je n'aime pas, je déteste...') in writing and using the dictionary. The teacher explained the speaking test instructions and passed out a sample letter to help them prepare for the upcoming writing test, which will be very similar to the speaking test. Their homework for Friday is to practise for their speaking tests, using a handout. The children were talkative today, but not too bad.

Day 5: Mardi 25-Nov-2008 S2 class – Song 'Les Bonbons'

Learning outcome: Practise French speaking skills.

We listened to the song while they came into class and their worksheets were distributed. Then I had them stand up to do the warm-up activity, and I divided them into girls and boys for the singing. The girls started with the first verse, and they were more confident and louder than the boys were at first.

Then they did several Lire/Écrire activities on pp. 36-38 in their textbooks.

Day 5: Mercredi 26-Nov-2008 S1 class – Dramatic 'Dialogue au parc'

Learning outcome: Practise French speaking skills.

We listened once to the dramatic dialogue while the Cloze worksheets were distributed. Then we said the dialogue together while standing up; I divided them into boys for the role of Sébastian and girls for Rachelle, on different sides of the room. I challenged them to say their lines as correctly and loudly as they could. The boy playing the role of the dog walked from one side of the room to the other, which was funny.

Then they reviewed likes and dislikes, practised for their speaking tests and finished a first draft of the letter for their writing test next week. They did some speaking practice about their family, pets, nationality, appearance and personality, 'j'aime...' (etc.), and how to ask 5 questions in French. Their homework for Friday was to practise for their speaking tests. At the end of class, they sang 'Joyeux Anniversaire' to one of the pupils.

Day 6: Vendredi 28-Nov-2008 S2 class – Song 'Les Bonbons' Mid-point test

Learning outcome: Demonstrate recall of the text practised in class and French pronunciation skills.

We started with a 'petite quiz' at the very beginning of class (Cloze post-test and French to English translations). Then I recorded them singing along with the audio recording, but they were very quiet so I stood just in front of them with the microphone while they were singing, even though they were standing together in a group. The opinion questionnaire took almost 10 minutes, which was longer than expected. Then the children started to watch the 'Malabar Princesse' film until the end of class.

Day 6: Vendredi 28-Nov-2008 S1 class – Dramatic 'Dialogue au parc' Mid-point test

Learning outcome: Demonstrate recall of the text practised in class and French pronunciation skills.

We started with a 'petite quiz' (Cloze post-test and translations). Then we listened to the recording while I passed out their Cloze exercise sheets and then I audio recorded them saying the dramatic dialogue twice. There was no time for the opinion questionnaires today, so they can do the evaluation at the start of class on Monday. (The teacher said they need time today to practise for their speaking tests on Monday.)

Their homework for Monday (1/12) is to prepare for their speaking tests and to have their monitors signed by their parent. They also have a listening test on Friday (5/12) and were asked to revise vocabulary beforehand.

WEEK 3

Day 1: Lundi 1-Dec-2008 S2 class – Dramatic 'Conversation au lycée'

Learning outcome: Practise French listening skills by filling in words missing from the text.

I introduced the dialogue, saying that it is about two high school students making plans and talking about past plans. Then we listened to the dialogue and they drew pictures of any words they recognised. I showed a few of their pictures to the class. Then we listened again as they filled in the missing words on their Cloze worksheets. I collected the worksheets afterward.

The children quietly watched more of the film 'Malabar Princesse.'

Day 1: Lundi 1-Dec-2008 S1 class – Song 'Le Tourbillon'

Learning outcome: Practise French listening skills by filling in words missing from the text.

First, we did the evaluation (since there was not enough time on Friday), which took approximately 10 minutes. Next, I gave a brief introduction about the song's theme and singer, and I asked them to draw pictures of any words they could recognise, or to draw a picture of how the song made them feel. Several of the boys started dancing in their seats while the song played, but the teacher asked them to stop dancing. After we listened to the song once, I asked them to fill in the missing blanks. I collected their Cloze worksheets afterward.

Then the teacher did 10 speaking tests (of 30) today. Meanwhile everyone wrote and practiced their speaking tests; the children seem quite anxious about all the tests.

Also, their teacher wasn't in a great mood, and after class she asked me to not to do the song at the start of Friday's class so there is enough time for the listening comprehension and writing tests. In the end we compromised – I will play the song once at the start of class on Friday, without passing out the Cloze worksheets.

Day 2: Mardi 2-Dec-2008 S2 class – Dramatic 'Conversation au lycée'

Learning outcome: Practise French listening skills and identify themes and facts about the text through listening and reading.

I reminded them what the dialogue was about and played it twice for them to finish filling in the blanks for the dialogue. We went through the English meaning of about half of the lines while filling in the missing blanks, and the teacher helped by writing the words in French on the SmartBoard.

They finished watching the 'Malabar Princesse' film. For homework, they were asked to answer 6 questions in French.

At the very end of class, they played a game similar to 'Who wants to be a millionaire?' in PowerPoint on the SmartBoard screen, which the children seemed to enjoy.

Day 2: Mercredi 3-Dec-2008 S1 class – Song 'Le Tourbillon'

Learning outcome: Practise French listening skills and identify themes and facts about the text through listening and reading.

I introduced and played the song twice as the children tried to fill in the words. Then I went through the English meaning of the first two sections with them and we filled in the rest of the missing French words.

Then they did exercises in their Métro textbooks on pp. 52-53 and a word search. There were more speaking tests today. The rest of the children did pp. 36-37: 1b, a writing/reading exercise. Their homework for Friday 5/12 was to prepare for their listening and reading tests, revising vocabulary from two textbook modules (1 & 2). For Monday 8/12, they have a writing test – they should learn the letter by heart to receive a good mark.

Day 3: Vendredi 5-Dec-2008 S2 class – Dramatic 'Conversation au lycée'

Learning outcome: Understand overall themes and specific details about the French text.

We went through the words after listening once to the dramatic dialogue and they wrote down the words in the blanks, while I translated into English and encouraged them to translate certain words that they knew.

Then they went through the answers to the teacher's questions about the film and attempted to have a full-class discussion in French about it.

1. C'est quel genre de film ? Comédie dramatique

2. Tu as aimé ?

3. Pourquoi ? Parce que...

4. Qu'est-ce que le 'Malabar Princesse' ?

5. Qu'est-ce que tu as préféré ?

6. C'était comment ? génial, amusant, triste.

The teacher reminded them how to use the dictionary: crash (v.) = *s'écraser* (reflexive verb) so it needs the auxiliary verb *être* rather than *avoir*. She showed them a present tense verb conjugation sheet (with -er, -ir, -re verbs). She also reminded them about how 'ne...pas' works in the passé composé (that it forms a sandwich around the helping verb, not the main verb).

Day 3: Vendredi 5-Dec-2008 S1 class – Song 'Le Tourbillon'

Learning outcome: Understand overall themes and specific details about the French text.

We listened to the song once at the start to wish them good luck and cheer them up before the assessment.

The teacher told them that if they were speaking during the French listening test, they would be sent out. No interruptions or questions were allowed during her explanation about how to do the listening test. The pupils became very quiet because of her tense tone of voice and the classroom atmosphere was not relaxed.

For the listening test, they heard the audio recording three times and they had to write down the answer to each question in English (out of 14 possible). One sentence was presented after another, separated by a pause of approximately 1 second (so the listening text was split into short sentences). Also, some of the items have more than one correct possible answer. The teacher read the phrases and another French teacher asked comprehension questions in English. The teacher gave them a few minutes to check their answers and she continued to be very strict about talking. The listening test items were in approximately the same order of presentation in the text for the reading assessment, too.

Dictionaries were allowed for the reading test. One boy got into trouble for asking for a pencil, and was put in one corner to complete the test (and he was almost sent out of the classroom altogether). The pupils appeared to know most of the words in the text, although a few words needed to be looked up. The teacher was frequently reminding the pupils to be quiet as she passed out the test sheets. At the end of class she reminded them to revise for their writing test, which is on Monday.

WEEK 4

Day 4: Lundi 8-Dec-2008 S2 class Dramatic 'Conversation au lycée'

Learning outcome: Obtain and provide information by reading the French text and practise French speaking skills.

I had the children answer a few comprehension questions in French (and English where necessary) about what was happening at various points in the dialogue. Then we did a 'Bonjour' warm-up and I went through translating each line after the girls or boys read each of their lines. They listened to the dialogue again as I collected the Cloze exercise sheets.

They had a quiz on vocabulary from the films unit, with questions 1-10 from français-anglais and items 11-20 from anglais-français. The children's desks were not separated for this vocabulary assessment, and one boy was looking at someone else's sheet. Then they corrected the quiz together and there was lots of discussion about half-points. Afterward they did exercises in the textbook, p. 42-43, 1a: 'vrai ou faux ?' (and had to correct the false sentences, in French) and a listening exercise; Écoute 1c took longer than expected because they were too talkative when the teacher tried to play the recording. Their homework for Friday is p. 11 in their workbooks.

Day 4: Lundi 8-Dec-2008 S1 class – Song 'Le Tourbillon'

Learning outcome: Obtain and provide information by reading the French text and practise French speaking skills.

I started the lesson by playing the song once. We discussed answers to a few listening comprehension questions, as a reminder of what the song was about. Then I tried to go through the missing words on their Cloze sheets with them, but this wasn't possible because we didn't have the computer and SmartBoard to use at the start of class (another teacher was using it for Spanish listening tests). I tried to spell out the words that belonged in each blank, but they couldn't follow. So I stopped that and we did a 'Bonjour' warm-up and tried to sing it together as best we could without having all the missing words filled in. They were singing quite happily, but they wanted to hear the correct words and to know what they meant – the children sang several lines on 'la' rather than singing the very fast French words.

Then they had 30 minutes to do their Level D writing test, which they should have already memorised. They were stressed but also very talkative regarding the exam so their teacher wasn't in a great mood, resulting in a tense classroom atmosphere.

Day 5: Mardi 9-Dec-2008 S2 class – Dramatic 'Conversation au lycée'

Learning outcome: Practise French speaking skills.

We started by listening once to the dramatic dialogue while I distributed the Cloze exercise worksheets (with pupils' help). Then I played the dialogue one line at a time with girls and boys repeating the lines separately (playing the two roles in the conversation). They had decent pronunciation and for the most part they kept quiet while the other group was speaking.

During class they did work in the Métro textbook, pp. 42-43, Rappel et Écoute 3b, cahier p. 23, create a genealogical tree, write about your family. The teacher reminded them about how to say my mother, my sister, my cousin, vs. my brother, my father in French. Then they did Écoute 3b – how old they are vs. 'il y a 5 ans' – 5 years ago. They completed Parler 2, p. 43, grammaire, *le détective* p. 42, *notre* vs. *nos*, both mean 'our' but one is singular and one is plural. A few children read aloud Lire 3a on p. 43. All three children mispronounced *habitent* by pronouncing the '-ent' ending. Their homework is to revise family related vocabulary, p. 11. Then they were meant to do some more advanced writing about their family, but ran out of time.

Day 5: Mercredi 10-Dec-2008 S1 class – Song 'Le Tourbillon'

Learning outcome: Practise French speaking skills.

At the start, there were a few children who arrived 3 minutes late and the others were very talkative, which was not very conducive to the children's concentration because this made the teacher agitated. We went through the words of song in English first, then we tried to sing while the children sat in their seats. They were well behaved when listening to the song and as I went through the words in English. They clearly wanted to know what the words meant. Some of the boys were humming the song afterward as I collected the Cloze worksheets.

While the last few pupils did their French speaking tests with the teacher, the others chose 6 academic subjects, drew a symbol for each one, then wrote the French word. Then they completed pp. 1, 12 in their new booklets and wrote their opinion of 5 academic subjects (e.g., 'j'aime' on p. 12).

The teacher promised they will have Christmas songs and games next week if they finish their homework on Friday. There was a 'Who wants to be a Millionaire' style game at the end of class after they wrote down their homework assignment, with new words (and le/la/les/l' practice) in French. Their homework is to learn subjects and opinions and to complete p. 12 of their workbooks.

Day 6: Vendredi 12-Dec-2008 S2 class – Dramatic 'Conversation au lycée' post-test

Learning outcome: Demonstrate recall of the text practised in class and French pronunciation skills.

We started with the 'petite quiz,' then we did a recording of the spoken dialogue, but unfortunately the microphone's battery was dying. It isn't clear if the microphone picked up enough sound to salvage this version for the DVD; we tried to listen to it, but the sound quality was extremely quiet and no one could hear it. Then they completed the opinion questionnaires.

Several children left the class for a music lesson today. Until the end of class, the rest of the children worked on a genealogical tree exercise and wrote a description about their families.

Day 6: Vendredi 12-Dec-2008 S1 class – Song 'Le Tourbillon' post-test

Learning outcome: Demonstrate recall of the text practised in class and French pronunciation skills.

I ran back to the department during the 15-minute break to find a better microphone, then hurried back to the school. We started with the 'petite quiz' (Cloze post-test and French to English translations) and then did an audio recording. The children sang the words of the song as best they could, but it was a challenge because unfortunately they had only seen all the words in French for one day before completing the post-tests. They completed the opinion questionnaires afterward.

With the French teacher, they did a 'Rappel des matières' on the SmartBoard, with one boy 'erasing' each one to find the word in French which corresponded to each picture, and then the class repeated the word after the teacher said it aloud in French. Then they did a speaking activity about their opinions of different school subjects, 'C'est comment, les maths ?' The teacher and a few children asked each other (as a full class activity) what they thought of different school subjects. Afterward they did an exercise in the textbook, Écoute 3b, items 1-7.

Day 7: Lundi 15-Dec-2008 S2 class – Dramatic 'Conversation au lycée'

At the end of this class session, I re-recorded the dramatic dialogue with the S2 pupils because it didn't work properly last week due to the microphone's battery dying.

Presentation of results to the children and teachers – 27-Feb-2009

I went to St Thomas of Aquin's to give them the DVDs and to present the results (in abbreviated form) of the arts intervention. It was nice to see the children again and they were extremely attentive to the presentation, probably since they were curious about what we were doing all that time in class together. I presented at the end of the S2 class (from 10:35 onward) and at the start of the S1 class.

Some of the children wanted to sing or listen to the song, but there was no way to play the DVD just then, so I said we could try singing it. Some of the S1 kids thought that the song was more fun and that they wanted to learn the words more, and some thought it was easier to find words in the chorus to fill in the blanks for the song. I also talked to them briefly about what a PhD is and about going to university.

The modern foreign language teachers said that the results were interesting. The French teacher agreed with me that the S1 song was too difficult for them, which might help explain the lower performance on the French translation post-test in that group compared to the S2 class.

Jacques Brel's biography for S2 class music lesson 1

Jacques Romain Georges Brel est né le 8 avril 1929 en Belgique. Quand il était jeune, sa famille a parlé français (au lieu de néerlandais). Jacques Brel a commencé à chanter et à composer des chansons sur le piano et sur la guitare. Brel a déménagé à Paris pour grandir sa carrière comme chanteur et musicien. Plus tard, il est devenu acteur et réalisateur de films. Il a fait 16 disques et 12 films, et quelques chansons sont traduites en langues étrangères. Il est mort le 9 octobre 1978 près de Paris, France. En décembre 2005, Jacques Brel est élu au rang de plus grand belge de tous les temps.

Jacques Brel : « Le talent, c'est d'avoir l'envie de faire quelque chose. »

Jeanne Moreau's biography for S1 class music lesson 1

Jeanne Moreau est née le 23 janvier 1928 à Paris. Son père était un restaurateur français et sa mère était une danseuse anglaise. Jeanne Moreau a étudié à la Conservatoire de Paris. Puis, elle a interprété beaucoup de rôles – elle était actrice dans 50 films ! Elle a travaillé avec des réalisateurs de films célèbres, par exemple, Jean-Luc Godard, Orson Welles et Marguerite Duras. Comme chanteuse, elle a fait quelques disques et un concert avec Frank Sinatra à Carnegie Hall.

Jeanne Moreau : « Vous ne devez pas séparer votre vie de votre travail. »

C.1.2 Cloze exercises

S1 Class Cloze Activity 1: 'Une conversation au parc'

Rachelle : Bonjour, Sébastien !

Sébastien : Bonjour, Rachelle. (1) ?

Rachelle : Oui, ça va bien. Oh, quel beau (2) ! Quel est son nom ?

Sébastien : Elle (3) Mimi.

Rachelle : Bonjour, Mimi ! Elle est (4). Quel âge a-t-elle ?

Sébastien : Elle a (5) ans.

Rachelle : Ah, elle mange (6) , donc ?

Sébastien : Oui. Mimi s'amuse au (7) , elle mange, et elle dort.

Rachelle : Ah, mais c'est pas mal, comme (8).

Sébastien : Oui, c'est vrai. Mimi est (9), je crois. Elle aime surtout chasser les (10) et les écureuils.

Rachelle : Ah ! Quelle (11) est-il ?

Sébastien : Il est treize heures et (12).

Rachelle : Zut alors, je dois m'en (13) tout de suite ! Je retrouve mon amie Sabine avant d'aller au cours (14). Nous avons un examen (15) prochain. Je n'aime pas les (16) !

Sébastien : Moi non plus !

Rachelle : Bon.

Sébastien : C'était sympa de te (17).

Rachelle : Oui, et ça m'a fait plaisir d'avoir (18) la connaissance de Mimi.

Sébastien : Ah, Rachelle, est-ce que tu vas (19) au restaurant grec (20) avec les copains du lycée ?

Rachelle : Oui.

Sébastien : On se voit ce (21), dans ce cas.

Rachelle : Oui, à bientôt alors !

Sébastien : À (22) !

S2 Class Cloze Activity 1: 'Les Bonbons'

Je vous ai apporté (1) bonbons
Parce que les (2), c'est périssable
Puis les (3), c'est tellement bon
Bien (4) les fleurs soient plus (5)
Surtout quand elles sont (6) boutons
Je vous (7) apporté des bonbons.

J'espère qu'on pourra (8) promener
Que Madame votre (9) ne dira rien
On ira (10) passer les trains
À huit (11), oui, je vous ramènerai
Quel beau (12) pour la saison
Je vous ai apporté des (13).

Si vous (14) ce que je suis fier
De (15) voir pendue à mon bras
Les gens (16) regardent de travers
Y'en a (17) qui rient derrière moi
Le (18) est plein de polissons
Je vous ai (19) des bonbons.

Oh ! Oui ! Germaine est moins (20) que vous
Oh ! Oui ! Germaine, elle est (21) belle
C'est vrai que Germaine a des (22) roux
C'est (23) que Germaine, elle est cruelle
Ça, vous (24) mille fois raison
Je (25) ai apporté des bonbons.

Et nous (26) sur la grande place
Sur le kiosque, on (27) Mozart
Mais dites-moi que (28) par hasard
Qu'il y a là (29) ami Léon
Si vous (30) que je cède la place
J'(31) apporté des bonbons.

Mais (32), Mademoiselle Germaine
Je vous (33) apporté des bonbons
Parce que les fleurs, c'est (34)
Puis les bonbons, c'est (35) bon
Bien que les fleurs (36) plus présentables...

S1 Class Cloze Activity 2: 'Le Tourbillon'

Elle avait des bagues à chaque (1),
Des tas de bracelets autour (2) poignets,
Et puis elle chantait (3) une voix
Qui, sitôt, m'enjôla.

Elle avait des yeux, des (4) d'opale,
Qui me fascinaient, qui me (5).
Y avait (6) de son visage pâle
De (7) fatale qui me fut fatale.
De femme fatale (8) me fut fatale.

On s'est connu, on s'est (9),
On s'est perdu de vue, on s'est reperdu de vue
(10) s'est retrouvé, on s'est réchauffé,
Puis on (11) séparé.

Chacun pour soi est (12),
Dans le tourbillon de la (13)
Je l'ai revue un (14), aïe, aïe, aïe !
Ça fait déjà (15) fameux bail.
Ça fait (16) un fameux bail.

Au son des banjos, (17) l'ai reconnue.
Ce curieux (18) qui m'avait tant plu.
Sa (19) si fatale, son beau visage pâle
M'émurent (20) que jamais.

Je me (21) soûlé en l'écoutant.
L'alcool fait oublier le (22).
Je (23) suis réveillé en sentant
Des baisers sur mon (24) brûlant.
Des (25) sur mon front brûlant.

On s'est connu, on (26) reconnu,
On s'est perdu de (27), on s'est reperdu de vue
On (28) retrouvé, on s'est séparé,
Puis (29) s'est réchauffé.

(30) pour soi est reparti,
Dans le (31) de la vie.
Je l'ai (32) un soir ah, là, là !
Elle est retombée dans (33) bras.
Elle est (34) dans mes bras.

Quand on s'est connu,
(35) on s'est reconnu,
Pourquoi se (36) de vue,
Se reperdre de vue ?

Quand on (37) retrouvé,
Quand on s'est réchauffé,
(38) se séparer ?

Alors tous (39), on est reparti
(40) le tourbillon de la vie
On a continué à (41)
Tous les deux enlacés
Tous les deux (42).

S2 Class Cloze activity 2: 'Un dialogue au lycée'

Olivier : Bonjour, Camille ! Ça va ?

Camille : Salut, Olivier. Oui, (1). Et toi ?

Olivier : Oui, ça va, merci. J'ai (2) dire que vous êtes allé voir le (3) film de James Bond. C'était comment ? Il paraît que tu l'as (4).

Camille : Mais non, je l'ai détesté, (5) assez fatigant. Les autres pensaient que c'était bien, mais je (6) trop ce genre de film.

Olivier : J'y serais (7) allé, mais malheureusement je n'avais pas (8). En plus, j'avais un examen (9).

Camille : Eh bien, c'est mieux d'étudier avant de (10) un examen au lieu de s'amuser au (11).

Olivier : Oui. C'est dommage que le (12) ne t'ait pas plu.

Camille : Oui, mais au moins c'était moins (13) que d'autres films d'action ou à (14).

Olivier : Bon. Est-ce que tu auras le temps de (15) quelque chose ensemble ce soir ?

Camille : Oui, (16), mais malheureusement je n'ai pas d'argent maintenant (17) je suis allée au cinéma hier soir !

Olivier : Moi non plus. J'avais envie (18) au concert de musique (19) ce soir, mais ça cote environ (20) livres, je crois.

Camille : Ah oui, j'avais (21) d'y aller aussi. Ça serait amusant, mais c'est (22) trop cher.

Olivier : Bon, on (23) aller à la piscine avec tes frères ?

Camille : Oui, peut-être...on verra (24). Ma mère vient de me téléphoner, mais je ne (25) ai pas demandé s'ils y vont aujourd'hui.

Olivier : Bon. (26) est-ce que ta mère te retrouve pour (27) aller ?

Camille : À (28) heures, je crois. Ça dure (29) une heure d'habitude.

Olivier : C’est (30), alors je finirai mes devoirs (31).

Camille : Oui, ce sera très bien. Je te (32) après les cours, alors.

Olivier : Oui, à (33) à l’heure !

C.2 French tests

C.2.1 S1 class Translation Pre/Mid/Post-test

Nom : ____________________________ Classe : _____________

Date : _________________________

Instructions: Translate the French phrases into the English equivalent. Do as much as you can and if you aren't sure, take a guess.

1. c'est périssable

____

2. qui rient derrière moi

____

3. j'avais envie d'y aller

____

4. surtout quand elles sont en boutons

____

5. j'avais apporté des bonbons

____

6. tu auras le temps de

____

7. mais non, je l'ai détesté

____

8. malheureusement je n'avais pas d'argent

____

9. on verra bientôt

____

10. si vous voulez que je cède la place

____

C.2.2 S2 class Translation Pre/Mid/Post-test

Nom : ____________________________ Classe : _____________

Date : _________________________

Instructions: Translate the French phrases into the English equivalent. Do as much as you can and if you aren't sure, take a guess.

1. tu vas manger au restaurant grec aujourd'hui

   ____

2. d'avoir fait la connaissance

   ____

3. nous avons un examen lundi prochain

   ____

4. elle est mignonne

   ____

5. autour des poignets

   ____

6. elle avait des bagues à chaque doigt

   ____

7. ce curieux sourire qui m'avait tant plu

   ____

8. surtout chasser les lapins et les écureuils

   ____

9. elle chantait avec une voix

   ____

10. le tourbillon de la vie

   ____

C.2.3 S1 class Cloze test 1: 'Une conversation au parc'

Rachelle : Bonjour, Sébastien !

Sébastien : Bonjour, Rachelle. Ça va ?

Rachelle : (1), ça va bien. Oh, quel beau (2) ! Quel est son nom ?

Sébastien : Elle s'appelle (3).

Rachelle : Bonjour, Mimi ! Elle est mignonne. Quel (4) a-t-elle ?

Sébastien : Elle a trois ans.

Rachelle : Ah, (5) mange beaucoup, donc ?

Sébastien : Oui. Mimi s'amuse (6) parc, elle mange, et elle dort.

Rachelle : Ah, (7) c'est pas mal, comme vie.

Sébastien : Oui, (8) vrai. Mimi est contente, je crois. (9) aime surtout chasser les lapins et (10) écureuils.

Rachelle : Ah ! Quelle heure est-il ?

Sébastien : Il est (11) heures et demie.

Rachelle : Zut alors, je (12) m'en aller tout de suite ! Je (13) mon amie Sabine avant d'aller au (14) d'histoire. Nous avons un examen lundi (15). Je n'aime pas les examens !

Sébastien : Moi (16) plus !

Rachelle : Bon.

Sébastien : C'était sympa de te (17).

Rachelle : Oui, et ça m'a fait plaisir (18) fait la connaissance de Mimi.

Sébastien : Ah, Rachelle, (19) que tu vas manger au restaurant (20) aujourd'hui avec les copains du lycée ?

Rachelle : (21).

Sébastien : On se voit ce (21), dans (22) cas.

Rachelle : Oui, à bientôt alors !

Sébastien : À (23) !

C.2.4 S2 Class Cloze Test 1: 'Les Bonbons'

Je vous ai apporté des bonbons
(1) que les fleurs, c'est périssable
Puis (2) bonbons, c'est tellement bon
Bien que (3) fleurs soient plus présentables
Surtout quand (4) sont en boutons
Je vous ai (5) des bonbons.

J'espère qu'on pourra se (6)
Que Madame votre mère ne dira (7)
On ira voir passer les trains
(8) huit heures, oui, je vous ramènerai
(9) beau dimanche pour la saison
Je (10) ai apporté des bonbons.

Si vous (11) ce que je suis fier
De (12) voir pendue à mon bras
Les (13) me regardent de travers
Y'en a (14) qui rient derrière moi
Le monde (15) plein de polissons
Je vous ai (16) des bonbons.

Oh ! Oui ! Germaine est (17) bien que vous
Oh ! Oui ! Germaine, (18) est moins belle
C'est vrai que (19) a des cheveux roux
C'est vrai (20) Germaine, elle est cruelle
Ça, vous (21) mille fois raison
Je vous ai (22) des bonbons.

Et nous voilà sur (23) grande place
Sur le kiosque, on (24) Mozart
Mais dites-moi que c'est (25) hasard
Qu'il y a là (26) ami Léon
Si vous voulez que (27) cède la place
J'avais apporté (28) bonbons.

Mais bonjour, Mademoiselle Germaine
Je (29) ai apporté des bonbons
Parce que (30) fleurs, c'est périssable
Puis les bonbons, (31) tellement bon
Bien que les fleurs (32) plus présentables...

C.2.5 S1 Class Cloze Test 2: 'Le Tourbillon'

Elle avait des bagues à chaque (1),
Des tas de bracelets autour des (2),
Et puis elle chantait avec une (3)
Qui, sitôt, m'enjôla.

Elle avait des (4), des yeux d'opale,
Qui me fascinaient, (5) me fascinaient.
Y avait l'ovale de (6) visage pâle
De femme fatale qui (7) fut fatale.
De femme fatale qui (8) fut fatale.

On s'est connu, on (9) reconnu,
On s'est perdu de vue, (10) s'est reperdu de vue
On s'est (11), on s'est réchauffé,
Puis on s'est (12).

Chacun pour soi est reparti,
Dans (13) tourbillon de la vie.
Je l'ai (14) un soir, aïe, aïe, aïe !
Ça (15) déjà un fameux bail.
Ça fait (16) un fameux bail.

Au son des (17), je l'ai reconnue.
Ce curieux sourire (18) m'avait tant plu.
Sa voix si (19), son beau visage pâle
M'émurent plus (20) jamais.

Je me suis soûlé en (21).
L'alcool fait oublier le temps.
Je (22) suis réveillé en sentant
Des baisers (23) mon front brûlant.
Des baisers sur (24) front brûlant.

On s'est connu, on (25) reconnu,
On s'est perdu de vue, (26) s'est reperdu de vue
On s'est (27), on s'est séparé,
Puis on s'est (28).

Chacun pour soi est reparti,
Dans (29) tourbillon de la vie.
Je l'ai (30) un soir ah, là, là !
Elle (31) retombée dans mes bras.
Elle est (32) dans mes bras.

Quand on s'est (33),
Quand on s'est reconnu,
Pourquoi se (34) de vue,
Se reperdre de vue ?

(35) on s'est retrouvé,
Quand on s'est (36),
Pourquoi se séparer ?

Alors tous deux, (37) est reparti
Dans le tourbillon de (38) vie
On a continué à tourner
(39) les deux enlacés.
Tous les deux (40).

C.2.6 S2 class Cloze Test 2: 'Un dialogue au lycee'

Olivier : Bonjour, Camille ! Ça va ?

Camille : Salut, Olivier. (1), ça va bien. Et toi ?

Olivier : Oui, (2) va, merci. J'ai entendu dire que (3) êtes allé voir le nouveau film (4) James Bond. C'était comment ? Il paraît (5) tu l'as aimé.

Camille : Mais non, je (6) détesté, c'était assez fatigant. Les autres (7) que c'était bien, mais je n'aime pas (8) ce genre de film.

Olivier : J'y serais (9) allé, mais malheureusement je n'avais pas (10). En plus, j'avais un examen aujourd'hui.

Camille : (11), c'est mieux d'étudier avant de passer (12) examen au lieu de s'amuser au (13).

Olivier : Oui. C'est dommage que le film (14) t'ait pas plu.

Camille : Oui, mais au (15) c'était moins effrayant que d'autres films (16) ou à suspense.

Olivier : Bon. Est-ce que (17) auras le temps de faire quelque (18) ensemble ce soir ?

Camille : Oui, peut-être, mais (19) je n'ai pas d'argent maintenant parce (20) je suis allée au cinéma hier (21) !

Olivier : Moi non plus. J'avais envie d'aller (22) concert de musique rock ce soir, (23) ça cote environ vingt livres, je (24).

Camille : Ah oui, j'avais envie d'y aller (25). Ça serait amusant, mais c'est beaucoup (26) cher.

Olivier : Bon, on pourrait aller à (27) piscine avec tes frères ?

Camille : Oui, peut-être ... (28) verra bientôt. Ma mère vient de (29) téléphoner, mais je ne lui ai (30) demandé s'ils y vont aujourd'hui.

Olivier : Bon. (31) est-ce que ta mère te retrouve (32) y aller ?

Camille : À huit heures, je (33). Ça dure environ une heure d'habitude.

Olivier : (34) parfait, alors je finirai mes devoirs (35).

Camille : Oui, ce sera très bien. Je (36) retrouve après les cours, alors.

Olivier : Oui, (37) tout à l'heure !

C.3 French Study Questionnaires

Language Learning Experience Questionnaire. Below are some statements relating to your language experience and background. There are no 'correct' answers. Carefully read each question and then indicate how true it is for you by circling the response which best corresponds to your past experience.

1. I find it easy to learn French.	Very untrue	Somewhat untrue	Equally true and untrue; unsure	Somewhat true	Very true
2. I enjoy speaking in French.	Very untrue	Somewhat untrue	Equally true and untrue; unsure	Somewhat true	Very true
3. I find it easy to remember words in French that I have heard someone say.	Very untrue	Somewhat untrue	Equally true and untrue; unsure	Somewhat true	Very true
4. I enjoy learning French.	Very untrue	Somewhat untrue	Equally true and untrue; unsure	Somewhat true	Very true
5. I prefer to read and write in French rather than speak.	Very untrue	Somewhat untrue	Equally true and untrue; unsure	Somewhat true	Very true
6. I know a language other than English to a very high standard:	Very untrue	Somewhat untrue	Equally true and untrue; unsure	Somewhat true	Very true
7. I regularly use a language other than English:	Very untrue	Somewhat untrue	Equally true and untrue; unsure	Somewhat true	Very true
8. It is important to have good pronunciation and conversational skills in French.	Very untrue	Somewhat untrue	Equally true and untrue; unsure	Somewhat true	Very true
9. Knowing French is important to me.	Very untrue	Somewhat untrue	Equally true and untrue; unsure	Somewhat true	Very true
10. I began learning French or a language other than my native language during primary school.	Very untrue	Somewhat untrue	Equally true and untrue; unsure	Somewhat true	Very true

Language Learning Experience Questionnaire.

11. About how much time altogether have you spent in a country where a foreign language is spoken?	< 1 month	1-6 months	6-12 years	1-2 years	2+ years
12. How many foreign countries have you visited where English was not the language spoken?	1 country	2-3 countries	4-6 countries	7+ countries	
13. Did you learn the language(s) before visiting?	Yes	No			
14. If yes, for how long did you learn the language before visiting, on average?	< 1 month	1-6 months	6-12 years	1-2 years	2+ years

Scoring for the Language Learning Experience Questionnaire.

For the first 10 Likert-style items, responses were coded with values between 1 (‘Very untrue’) and 5 (‘Very true’). If the child wrote that they knew or regularly used another language, a note was made. For the second section of the LLEQ, the following scoring procedure was used to calculate a total score.

Question 11	0 points = None or blank
	1 = less than 1 month
	2 = 1-6 months
	3 = 6-12 months
	4 = 1-2 years
	5 = 2+ years
Question 12	0 = None or blank
	1 = 1 country
	2 = 2-3 countries
	3 = 4-6 countries
	4 = 7+ countries
Question 13	0 = No or blank
	1 = Yes a little bit, or some
	2 = Yes
Question 14	0 = None or blank
	1 = less than 1 month
	2 = 1-6 months
	3 = 6-12 months
	4 = 1-2 years
	5 = 2+ years

Artistic Experience Questionnaire. Below are some statements relating to your artistic experience and background. There are no 'correct' answers. Carefully read each question and then indicate how true it is for you by circling the response which best corresponds to your past experience.

1. I enjoy listening to recorded music (CDs, MP3s, and/or radio).	Very untrue	Somewhat untrue	Equally true and untrue; unsure	Somewhat true	Very true
2. I enjoy participating in a drama group or performing in theatre productions.	Very untrue	Somewhat untrue	Equally true and untrue; unsure	Somewhat true	Very true
3. I enjoy drawing, painting, or other creative, artistic activities.	Very untrue	Somewhat untrue	Equally true and untrue; unsure	Somewhat true	Very true
4. I often sing or hum along as I listen to music for fun.	Very untrue	Somewhat untrue	Equally true and untrue; unsure	Somewhat true	Very true
5. I often pretend to be someone else or act in different ways for fun.	Very untrue	Somewhat untrue	Equally true and untrue; unsure	Somewhat true	Very true
6. I often draw, paint, or create artwork for fun.	Very untrue	Somewhat untrue	Equally true and untrue; unsure	Somewhat true	Very true
7. I enjoy attending concerts or live music events.	Very untrue	Somewhat untrue	Equally true and untrue; unsure	Somewhat true	Very true
8. I enjoy watching films at the cinema and/or plays at the theatre.	Very untrue	Somewhat untrue	Equally true and untrue; unsure	Somewhat true	Very true
9. I enjoy participating in a choir or musical ensemble.	Very untrue	Somewhat untrue	Equally true and untrue; unsure	Somewhat true	Very true
10. I enjoy going to art museums or art galleries.	Very untrue	Somewhat untrue	Equally true and untrue; unsure	Somewhat true	Very true

Artistic Experience Questionnaire.

11. Have you trained on a musical instrument or received singing lessons?	Yes	No
Which instrument(s) did you play?		
For how long?		
If you did grades, what was the highest level you reached?	Grade:	
12. Is there a piano or keyboard at your house?	Yes	No
If you have a piano or keyboard at home, does (or did) anyone play it?	Yes	No
Who?		
For how long?		
13. About how many hours do you listen to music each week?		
14. About how many concerts or live music events do you attend each month?		
15. What is your favourite band or style of music?		
16. Have you had any acting experience or taken acting lessons?	Yes	No
For how long?		
17. Has anyone in your family had acting experience or training?	Yes	No
Who?		
For how long?		
18. About how many hours do you watch television and films each week?		
19. About how many films or plays do you attend each month?		
20. Who is your favourite actor or actress?		
21. What is your favourite film or television show?		
22. Have you had any artistic training or taken art lessons?	Yes	No
What art form(s) did you practise?		
For how long?		
23. About how many art museums or galleries do you visit each month?		
24. Who is your favourite artist, painter, or sculptor?		
25. What is your favourite kind of art?		

Scoring for the Artistic Experience Questionnaire.

For the first 10 Likert-style items, responses were coded with values between 1 (‘Very untrue’) and 5 (‘Very true’) and a total AEQ score was calculated. Pupils’ responses to questions 1, 4, 7 and 9 were used to calculate the Music sub-score; responses to questions 2, 5, and 8 comprised the Drama sub-score, and responses to questions 3, 6, and 10 were used to calculate the Visual Art sub-score. For the second section of AEQ questions, responses were scored as shown in Tables C.1 to C.4 below and total AEQ and sub-scores were calculated for this questionnaire. The raw scores on each sub- section were transformed into percentages because there were unequal numbers of points possible for the different sub-sections.

Table C.1: Musical experience calculation (1)

Question 11	0 points = None or blank
	1 = Yes
Instrument:	0 = None or blank
	1 = 1 instrument
	2 = 2 instruments
	3 = 3 instruments
	4 = 4 instruments
	5 = 5 or more instruments
How long:	0 = None or blank
	1 = less than 1 year
	2 = 1 year to less than 3 years
	3 = 3 to less than 5 years
	4 = 5 to less than 8 years
	5 = more than 8 years
Grades:	0 = None or blank
	1 = Grades 1-2
	2 = Grades 3-4
	3 = Grade 5
	4 = Grade 6-7
	5 = Grade 8
Question 12	0 points = None or blank
	1 = Yes
Who?	0 = None, blank or self
	1 = 1 non-nuclear family member
	2 = 1 nuclear family member
	3 = 2 family members
	4 = 3 family members
	5 = more than 4 family members
How long:	(see above)

Table C.2: Musical experience calculation (2)

Question 13	0 = None, blank, or not sure
	1 = 1 hour or less
	2 = more than 1 hour up to 5 hours
	3 = more than 5 hours up to 10 hours
	4 = more than 10 hours up to 20 hours
	5 = more than 20 hours
Question 14	0 points = None, blank, or not sure
	1 = less than 1
	2 = 1-2 concerts
	3 = 2-3 concerts
	4 = 3-4 concerts
	5 = 5+ concerts
Question 15	0 = None, blank, or negative
	1 = One artist or style (well known)
	2 = One artist or style (less well known)
	3 = Two artists or styles
	4 = Three artists or styles
	5 = Four or more artists or styles

Table C.3: Dramatic experience calculation

Question 16	0 points = None or blank
	1 = Yes
How long:	(see Music Question 11)
Question 17	0 points = None or blank
	1 = Yes
Who?	(see Music Question 12)
How long:	(see Music Question 11)
Question 18	(see Music Question 13)
Question 19	0 = None, blank, or not sure
	1 = 1 film or less
	2 = 2 films
	3 = 3 films
	4 = 4-5 films
	5 = more than 5 films
Question 20	0 = None, blank, or unsure
	1 = One actor (well known)
	2 = One actor (not well known)
	3 = Two actors
	4 = Three actors
	5 = Four or more actors
Question 21	0 = None, blank or negative
	1 = 1 film/show
	2 = 2 films/shows
	3 = 3 films/shows
	4 = 4 films/shows
	5 = 5 or more films/shows

Table C.4: Visual Art experience calculation

Question 22	0 points = None or blank
	1 = Yes
What kind:	0 = None, blank or negative
	1 = 1 artform
	2 = 2 artforms
	3 = 3 artforms
	4 = 4 artforms
	5 = 5 or more artforms
How long:	(see Music Question 11)
Question 23	0 = None or blank
	1 = less than 1
	2 = 1-2 museums
	3 = 2-3 museums
	4 = 3-4 museums
	5 = 4-5 museums
Question 24	0 = None, blank, or unsure
	1 = One artist (well known)
	2 = One artist (not well known)
	3 = Two artists
	4 = Three artists
	5 = Four or more artists
Question 25	0 = 0 = None, blank or negative
	1 = 1 artform
	2 = 2 artforms
	3 = 3 artforms
	4 = 4 artforms
	5 = 5 or more artforms

Table C.5: AEQ: Music sub-scores for participants in the S1 and S2 classes

Music Total Percent	*N*	*M*	*SD*	Range	Possible
S1 Class	19	49.4%	15.2%	24.1 - 76.8%	0 - 100%
S1 Class: Male	8	45.9%	19.4%	24.1 - 76.8%	0 - 100%
S1 Class: Female	11	52.0%	11.6%	27.0 - 74.5%	0 - 100%
S2 Class	23	48.7%	11.6%	31.1 - 73.6%	0 - 100%
S2 Class: Male	10	44.3%	10.3%	31.1 - 61.6%	0 - 100%
S2 Class: Female	13	52.2%	11.6%	33.4 - 73.6%	0 - 100%
Music sub-section 1	*N*	*M*	*SD*	Range	Possible
S1 Class	19	3.89	0.97	1.5 - 5.0	1 - 5
S1 Class: Male	8	3.52	1.16	1.5 - 4.7	1 - 5
S1 Class: Female	11	4.16	0.76	2.3 - 5.0	1 - 5
S2 Class	23	3.86	0.71	2.8 - 5.0	1 - 5
S2 Class: Male	10	3.33	0.57	2.8 - 4.3	1 - 5
S2 Class: Female	13	4.27	0.53	3.3 - 5.0	1 - 5
Music sub-section 2	*N*	*M*	*SD*	Range	Possible
S1 Class	19	11.6	9.08	1 - 35	0 - 55
S1 Class: Male	8	11.8	12.4	1 - 25	0 - 55
S1 Class: Female	11	11.5	6.42	5 - 27	0 - 55
S2 Class	23	11.2	7.82	1 - 26	0 - 55
S2 Class: Male	10	12.1	7.17	4 - 21	0 - 55
S2 Class: Female	13	10.5	8.50	1 - 26	0 - 55

Table C.6: AEQ: Drama sub-scores for participants in the S1 and S2 classes

Drama Total Percent	*N*	*M*	*SD*	Range	Possible
S1 Class	19	47.0%	13.1%	26.7 - 78.9%	0 - 100%
S1 Class: Male	8	49.6%	15.4%	26.7 - 78.9%	0 - 100%
S1 Class: Female	11	45.1%	11.7%	32.2 - 68.9%	0 - 100%
S2 Class	23	44.3%	11.9%	26.7 - 66.7%	0 - 100%
S2 Class: Male	10	36.8%	5.8%	27.8 - 45.6%	0 - 100%
S2 Class: Female	13	50.1%	12.3%	26.7 - 66.7%	0 - 100%
Drama sub-section 1	*N*	*M*	*SD*	Range	Possible
S1 Class	19	3.74	0.81	2.67 - 5.0	1 - 5
S1 Class: Male	8	3.92	0.87	2.67 - 5.0	1 - 5
S1 Class: Female	11	3.61	0.77	2.67 - 5.0	1 - 5
S2 Class	23	3.58	0.96	2.00 - 5.0	1 - 5
S2 Class: Male	10	2.80	0.42	2.0 - 3.67	1 - 5
S2 Class: Female	13	4.18	0.81	2.67 - 5.0	1 - 5
Drama sub-section 2	*N*	*M*	*SD*	Range	Possible
S1 Class	19	8.63	6.58	0 - 26	0 - 45
S1 Class: Male	8	9.38	8.03	0 - 26	0 - 45
S1 Class: Female	11	8.09	5.65	2 - 20	0 - 45
S2 Class	23	7.65	4.60	0 - 18	0 - 45
S2 Class: Male	10	7.90	3.93	4 - 16	0 - 45
S2 Class: Female	13	7.46	5.21	0 - 18	0 - 45

Table C.7: AEQ: Visual Art sub-scores for participants in the S1 and S2 classes

Visual Art Total Percent	*N*	*M*	*SD*	Range	Possible
S1 Class	19	38.2%	12.1%	15.5 - 60.0%	0 - 100%
S1 Class: Male	8	37.6%	15.5%	15.5 - 60.0%	0 - 100%
S1 Class: Female	11	38.6%	9.8%	23.3 - 54.4%	0 - 100%
S2 Class	23	37.8%	12.1%	16.7 - 58.9%	0 - 100%
S2 Class: Male	10	28.3%	10.0%	16.7 - 47.8%	0 - 100%
S2 Class: Female	13	45.0%	8.0%	33.3 - 58.9%	0 - 100%
Visual Art sub-section 1	*N*	*M*	*SD*	Range	Possible
S1 Class	19	3.55	1.04	1.3 - 5.0	1 - 5
S1 Class: Male	8	3.44	1.36	1.3 - 5.0	1 - 5
S1 Class: Female	11	3.64	0.81	2.3 - 5.0	1 - 5
S2 Class	23	3.55	1.09	1.7 - 5.0	1 - 5
S2 Class: Male	10	2.70	0.96	1.7 - 4.7	1 - 5
S2 Class: Female	13	4.21	0.63	3.0 - 5.0	1 - 5
Visual Art sub-section 2	*N*	*M*	*SD*	Range	Possible
S1 Class	19	2.37	3.37	0 - 12	0 - 30
S1 Class: Male	8	2.88	4.19	0 - 12	0 - 30
S1 Class: Female	11	2.00	2.79	0 - 10	0 - 30
S2 Class	23	2.04	2.50	0 - 10	0 - 30
S2 Class: Male	10	1.20	1.40	0 - 4	0 - 30
S2 Class: Female	13	2.69	2.98	0 - 10	0 - 30

Opinion Questionnaire (Mid-Point and Post-Intervention)

Learning Foreign Languages through the Arts

Name:

Please take a moment to reflect on the activities we have done together in French class.

1. What was your favourite activity (or activities)? How was it fun?
2. How did you feel challenged by that activity (or those activities)?
3. How did you feel that activity (or those activities) helped you?
4. What was your least favourite activity (or activities)? What didn't you like about it?
5. Please give an example of an activity in which you felt you were a successful learner, a confident individual, a responsible citizen, or an effective contributor. Explain why you feel this way.
6. Did the words we were listening to ever repeat in your head (other than during the listening activities in French class)? If so, can you say when this repetition happened?[1]
7. Can you suggest any ways of improving the activities or the learning experience in future?
8. Did you prefer listening to the song or the dramatic dialogue together in class?[2]
9. Do you have any other comments or suggestions?

[1]This question was only asked on the post-intervention questionnaire.
[2]This question was only asked on the post-intervention questionnaire.

Appendix D

Supplemental Materials for the Knowledge Exchange Study

D.1 Transcript of pilot workshop discussion with Masters students, 22 May 2009

[Audio recording file 2 - 00:53:03 (after a short break)]

Karen: Okay, everyone. Hello again and welcome back. I hope that was refreshing and now we have got some cooler air coming in again, which is a very positive development because it was getting a bit warm.

So, has anyone had any thoughts about these questions and wants to share?

F1: I can go first.

Karen: Okay.

F1: I think the major problem of adapting music in the classroom is the resources you can get.

Karen: Okay.

F1: Like me, we have to use textbooks that are assigned by the school. So we have to use the textbook on one hand. On the other hand, we have to find the music on our own. That will be difficult, that will be real difficult. And in the meanwhile, I teach young learners, so most of the English songs, the lyrics are too difficult for young learners. So I have to look through the particular songs for young learners. And even so, most of them are still very difficult. But there can be very good background music, like it is very cheerful and cheering, but the thing is, the student cannot sing along with such lyrics because it is too difficult. Even though it is made for children, but it is made for native speakers instead of the foreign language learners. So that is the main difficulty that I have. So, although I try to use music and songs, but there is a limitation there. And the CDs are very expensive as well.

Karen: Yes.

F1: Yes, so to get the database or building up the music lessons in teaching language is a little bit pricey.

F2: My difficulty would not be having the resources, but finding the proper songs with the proper message in them, because currently there are so many music – the songs out there, they are just so current and very pop, and the words that they use are sometimes not suitable for teaching. You know, there are sort of taboo words out there.

Karen: Yes.

F2: So to select the right music which tries to suit everyone is going to be quite challenging. Some people like ballads, some people like rock, some people like heavy metal and it's, you have to find that centre to ensure that everyone is engaged in the music. That is my difficulty, for me.

And as well as the lyrics, I guess as you know, songs are not proper sentences, they are just chunks of languages, so you have to ensure that the kids know that these are not how you speak normally, communicative – in a communicative sort of environment. So to me, it's going to be – have to find a balance or provide the support for kids to know that songs are songs, you can use the words in the music, but you need to know the place and the time for using those sort of words as well.

Karen: Yes.

F2: So I would find that sort of support, to try to answer this difficulty for me. So it is just getting the right song for the right time.

F3: I agree with [F1]. Usually when teachers try to use songs for teaching, it is usually supplementary materials, it is not main class. So usually teachers have to depend on themselves, they have to find by themselves. So it takes a long time and as you know, teachers are very busy all the time, so they have to put extra efforts for that. And there are tons of songs, but if they want to teach some – certain grammar, and she tried to find some lyrics that has that grammar, that they have to go through every song and try to find, oh, this grammar is here, so maybe I can use these songs. They have to – it takes lots of time.

Karen: Yes.

F3: So I think that's one of the biggest challenges to using songs.

Karen: Okay.

M1: Well, the first thing I've got down here is shy. Some people are shy, so they just won't – well, some people? I think everybody is shy.

All: [laughter]

F1: [The teachers, the teachers.

M1: Hands up whoever is not shy.

All: [laughter]

M1: Yes, even the teacher as well, eh? So that would be one challenge, just to sort of introduce this new methodology slowly and bit by bit.

Karen: Build up confidence, yes.

M1: To build up confidence, not to shock them too much. And another difficulty might be that there are some people who are amusical, who just don't quite understand music. I don't know if, like, musicologists or theorists –

Karen: There are very few, actually. I mean, fewer than you might think.

M1: Yes, it depends what you define as amusical.

Karen: Yes.

M1: I remember one of my best friends, he earned a living playing the bagpipes on street corners all over the world. And so he must have played 'Scotland the Brave' half a dozen times each day for ten years. How many hundreds of times is that? If you asked him to hum it, he couldnt do it. He didn't know how to sing. In fact, when he was growing up, one of the – sharing a conversation between people about childhood memories, what TV programmes do you remember? Top of the Pops, did you like Top of the Pops? He hated Top of the Pops because he just found it to be annoying. He was annoyed. He couldn't distinguish one song from another. So maybe he was an extreme case, but what I'm saying is that there must be a spectrum of –

Karen: [Yes.

M1: [People who do like music, which is most people I think. Because I do remember being told it is a silly question to ask in sort of interview situations as a teacher, 'Do you like music?'

Karen: Right.

M1: It is a silly question, because who is going to say no? Apart from [proper name], that is. So most people do like music, but there are that – there will be that band of people...

Karen: I think that is true. Yes.

M1: Who don't appreciate it. So there are a couple of difficulties.

Karen: Yes, that is true.

M1: What you said there with, that it is time-consuming with choosing the songs, I think, yes, just turn it on its head and get the students to choose the songs. Because that would be ultimately much more motivating for them, because they're not going to want to listen to David Bowie I don't think.

Karen: Yes.

M1: Or Talking Heads. I don't know who they would want to listen to these days, but –

Karen: Yes, that is a really good idea actually, to have the students suggest songs and then of course you can review the lyrics before you actually play it for the whole class. But yes, that's a great idea.

M1: And then you could even get peer – what do you call it – peer evaluation of choices of music. So you know, I want to listen to, I don't know, X, Y, Z, and you put it on, and then yes, you know, this is great. This is my choice of music and at the end of it, they suddenly realise, well, actually, he only says three words. So what have we learned there? And there might be sort of peer evaluation that goes on. For the next choice, I want to actually make a better choice the next time. They might actually analyse the lyrics themselves –

Karen: [Yes.

M1: [To make sure their next choice is a better choice. Perhaps.

F4: Yes. I agree with what [M1] had said. And according – it depends on the student's personality. I just remember one of my students who was from German – a guy, he is not a child, but he said he doesn't like singing. And whatever effort I make, I just couldn't get him singing. So I guess it is maybe not a good way for him. But most of the students love learning songs.

And another problem is time control, and when we learn songs in class, the students quite love it and I enjoy it, too. And sometimes I found, oh, we spent too much time and we couldn't finish with the plan. So that is a problem.

And also, in choosing material, I think I can't let students to decide what song to learn. I have an – a goal what I want them to learn from the songs, so probably I would choose the song for them. But the problem is, how can I make sure this is a good song for them? That the terms, the vocabulary, or the proper repetitions of the new words? Something like that. So it is what I – my goal for the song.

Karen: Okay. Well, I will just – one of your first points, what did you say? The first or second thing you said.

F4: Second?

Karen: The first or second thing that you said.

F4: Personality?

Karen: Ah, yes, about personality. There is no, absolutely no reason that anyone should be forced to sing if they don't want to sing a song. Let them be – they are not in music class at that point. And I guess what you can do in a situation like that, when somebody just – they aren't shy, but they're just not interested in singing, they just don't want to. I mean, thirteen-year-old boys – yes, I know that feeling. But basically what you can do is – well, there's a few things you can do.

You can say to them, to the whole class, that this is why I want you guys to try to sing along. It's because I think it's going to help your pronunciation and your speaking abilities and it's not because I want to hear how beautiful you sound. It's because I want you to learn these particular skills through this practice. So that's one thing you can do, is help them see why – your motivation for using it.

And then another thing is, even just listening to it and mouthing the words – if they don't even sing it out, it's okay as long as – if their mouth is moving around in time with the words. I think that's still almost just as good as actually singing it. If they just think they don't have a nice singing voice, then you can still – they can still practise along, just not loudly.

M1: Is that called sub-vocalisation?

Karen: Yes, sub-vocalisation, exactly.

M1: Who wrote about that?

Karen: There was someone named Karen Gfeller. She wrote about that a bit. It is a 1983 paper I think, where she found that children's spelling was improved by either singing along or sub-vocalising to the words of songs, whether they were – in the native language – even if they had learning disabilities, actually, it was helping them. That was really good.

M1: That reminds me of – what's her name? Karen Gfeller?

Karen: Gfeller. G-F-E-L-L-E-R.

M1: G-F?

Karen: G-F-E-L-L-E-R. Yes.

M1: G-F-E-L-L-E-R.

Karen: Yes.

M1: 1983?

Karen: 1983, yes.

M1: That reminds me of somebody, it was in Japan, he was teaching pronunciation and this sort of trick question to his students was, where does good pronunciation start? And was wanting answers like, tongue, lips, [throat –

Karen: [Mouth, yes.

M1: Lungs, whatever, and he goes, no. Good pronunciation starts here in the mind, and you've got to think about how the sounds are made before you actually try to make them. And so he said more or less something like, if this is the teacher and this is the student, the old method would be, 'Listen and repeat. Today.' 'Today.' New method, the teacher and the student would be, 'Listen and don't repeat. Today.' () Very good.

All: [laughter]

M1: And what the student is doing is, they're not moving their lips or anything, they're imagining their own voice inside their own heads, saying what the teacher has modelled. One stage on from that would be, I think, sub-vocalising.

Karen: Sub-vocalisation, yes.

M1: So these sort of little techniques, they do a little [indiscernible] with music as well.

Karen: Yes. [F5], do you want to – I know you're not teaching a foreign language, but still, do you have any comments?

F5: Well, I must admit, my interest has been to turn this around and to try and support some local children going to the secondary.

M1: What nationalities are they?

F5: It's just a local primary school. All sorts. Polish and –

M1: [Right.

F5: [Somali, Indian, Chinese –

M1: Polish, Spanish.

F5: Spanish. [Yes.

M1: [Yes, right.

F5: And the idea – and this is based on Caroline's experience. Actually I am just talking about it at the moment – these are all just up in the air projects, but you never know – is that you, the idea is to – the other side affecting music is to [indiscernible] confidence. It takes down some barriers. Now obviously, I am – and my colleagues here – are musically trained to some extent, so we have more confidence. And the idea for someone like me would be to – as a support person. So you were talking about finding the song as well. The idea behind a community musician or a music specialist going in would be to work under the authority of the current language teacher or the English teacher, in my case, to help them resource all this and to support that. The other side to the story is – and the idea behind that has been worrying, is the first job is to search for funding from somewhere. Sometimes the education authorities have some, but sometimes this sort of thing allows you to look at other places like the Arts Council.

Karen: Yes.

F5: Which traditionally don't necessarily go out to do education. So you're talking – it's all politics. You're trying to get some money out of somewhere in order to support resourcing this sort of work. And you employ a freelancer person or someone like that to do it or to support it. And they kind of go in – not every week, not every time, but just every so often to facilitate that.

Karen: Yes, to help out.

F5: And the basic process, this is where they – is to, you take their songs, so you're basing it on local folk tradition. And this is aimed at primary, so their tastes in music – I read somewhere that taste in music hasn't developed quite so strongly until secondary level. So you're taking their local folk tradition and Scottish local folk tradition, and you start with songs from their cultures, the whole class learn them, you translate them into English, and then you learn some Scottish songs in both languages if possible. And it just brings people together, working together, builds confidence. And that's – you're talking possibly, I mean, if you're talking initially immigrants and stuff, they could be quite traumatised by the transfer and everything. And music all is new, new this, new that. And if you can start with something that might at least be somewhat familiar to them –

Karen: Yes.

F5: And the idea is to encourage communication, because music is a non-specific communication tool. That's kind of the point. So those are my thoughts, really.

Karen: Great, that's very interesting. [F6], do you want to say anything?

F6: The challenges I have with teaching Mandarin to Scottish children is – well, it is a few things. First of all, it is the resources. I actually couldn't find any Chinese folk songs around here. I needed to actually look for friends who are from China.

Karen: Right.

F6: Actually, my main resources came from really, well, a Chinese singer who is based in Glasgow. I don't if you have heard of her, Fong Liu? But she has done a lot of singing for the Ricefield Centre – Chinese Centre [Ricefield Chinese Arts and Cultural Centre] and she has been doing a lot of things for the BBC in Scotland.

Karen: Cool.

F6: And MTV and things like that. But the thing is, I find when teaching Mandarin songs to Scottish children is that it's – speaking it, learning it is different to singing it.

Karen: Okay.

F6: So is there any point of teaching them the tones and everything, and then – or should I just teach them the music itself. And that's my problem because in Mandarin there's four tones and after teaching them the tones and actually saying it, the words and stuff in Pinyin, and then you try to add in the melody, and then they think okay, now I'm going to have to apply the tones together with the melody line. And then that was their problem. And then I have to tell the teacher, no, when we sing, we just have to cancel out the tones and it's always like, so what is the point of learning the tones then? You might as well just learn the song itself with the Pinyin. But when they were singing it with the Pinyin in Mandarin, it was really difficult because they never sung Chinese folk songs, so they are trying to understand how to sing it and at the same time trying to say the – pronounce them properly.

Karen: Yes.

F6: And that was the problem.

F4: But I think when you learn Chinese songs, you can just forget about tones.

F6: Yes, me too, forget about tones.

F4: Ignore the tones.

Karen: One thing you can do – I don't know, you guys probably know many better strategies than I do – but when you're first trying to teach a song in this context of foreign language learning at least, it can be helpful to present it – make sure they hear it several times all the way through. Don't just instantly say, 'Okay, we're going to sing a song today, and here it is and go ahead, here are the words.' Although I'm sure you wouldn't do that, but I could imagine someone trying that and it would not work. You need to make sure that the melody has been learned as well as the sound of the words, before you ever try to start saying it out loud or singing it out loud. And then even when you do start singing it, especially with a real song, even with folk songs, they have a lot of words in them and sometimes they're quite fast, maybe not as fast as speech, but still quite fast.

And so you do need to be careful that you don't try to bite off too much at one time. Don't say, 'Okay, now we are going to sing a song, starting at verse one and going to the end.' Maybe just start with the chorus first or learn one verse at a time or before you even start singing the verses or the chorus, sing the melody of the chorus. Maybe play the song if you have got a recording of it, but just sing it with 'la', you know? Just 'la, la, la, la.' They can do that. They don't have to think about their mouth as much then, and then they can sort of get the melody and the rhythm down and then you can introduce the words more gradually. It's a bit less pressure than sort of doing so many new things all at once.

F6: That is something that I was wanting to ask, something that I didn't actually think about. After teaching one of the folk songs from my Chinese workshop, the music teacher for that school actually said to me, if it's okay I could find a shorter song which is literally like one line or something. And for me, it was really hard to find a Chinese folk song that has only one line.

Karen: Yes.

F6: Yes, I think what –

Karen: Why did she want that?

F6: Sorry?

Karen: Why did they want that?

F6: I think first of all, when it comes to learning culture, when they talk about multicultural music, arts here in Scotland, the first thing that will come through their mind is African music.

Karen: Oh, I see.

F6: So a lot of schools always just do African music because it's easier to adapt and obviously there is very different between the Western culture and Africans, and so when they pick up – a lot of teachers prefer to learn African music because in African songs, there is only one or two or three different words and that is repeated again and again, again. So it's a [sings an African song] You know, it just repeats again, again.

Karen: Right.

F6: You just move around. And so when I went in, they wanted something similar.

Karen: Wow.

F6: So when I – when he asked me for a short song, I was like, this is the shortest song I could find. And they want me to look for something like a few words, and then sing that line again. So it's just really hard. So I just –

Karen: Well, you should tell the teachers that that's not part of the cultural tradition.

F6: Yes, so I don't know if, I mean, if you're teaching language maybe that's a good way to start. Find a short song with words that repeats again, again, again to kind of get the words out.

Karen: Yes. Yes.

F7: The thing that is coming up for me is, well, I think we need to get away from songs and what we think of as songs. And if it's about learning language, then it's about being really creative with what we've got. So if you've got – if you're not confident singing, then there is always a CD player, and there is always a – you can always get backing tracks for songs as well. So having backing tracks and getting the children singing. But if it's – well, for you it's about teaching about culture, so maybe that's different. But I think the work that I did in India, I mean, I know it probably sounds kind of daunting if you're not a musician, but we just wrote simple songs. So, [singing] 'Hello, hello. The sun is in the sky. The sun is in the sky.' So really simple little tunes where the words were just repeated and repeated, because they're just not available – I mean, childrens songs are really complicated. So then having an idea about what you want to teach, whether it's general conversation – 'Hello, hello / How are you? How are you? / Goodbye, goodbye,' and it was done by call and response.

So I think that we can move away from thinking how can we get what we want out of the songs that are out there. And the other really obvious thing which I think works really well is tunes that you already know. So even 'Happy birthday to you' – adding words to the tune that you know the children already know, that you want to teach them. So I mean, they'll automatically want to join in with that, I think, as well, because the tune is already learnt. So I'd think it is like manipulating the music, because actually in a way it's the least important thing. Although it's what I think is motivating and catchy and gets people in, it's – if that bit is assumed, then you get all the motivational benefits from it and their concentration can be on the vocabulary.

F5: Actually, I was talking to other teachers – this was at Storytelling [the Scottish Storytelling Centre] – and they just, yes, exactly that. They took 'London is burning' and they must have made up about five different songs to 'London is burning' for various different purposes for some things they did.

Karen: Cool.

F5: You know, just to encourage –

Karen: [F6] needs to go. Thanks very much for coming.

F5: Yes, because especially nursery rhymes and stuff, and now I'm back to this Orff training I did and [F6] did as well, which they – the other thing, there is short versions of traditional – again this is probably primary, but traditional clapping games from other cultures. Some of them are knotted, right, but they still have the intonation of [that language.

Karen: [The rhythm. Yes.

F5: If that makes sense. But I mean, and some of the clapping games could teach you English because you're doing something. It's all about just waking people up, I think, isn't it?

Karen: Yes, staying engaged [in the activity.

M1: [Yes.

F5: And building confidence. You can – if, and again, this is primary level again, but early secondary as well. I mean, I used to teach. Design technology was my original subject and I – we were encouraged, when I was doing stuff, to play games. Now imagine playing games with drills and hacksaws, yeah, right! But you know, health and safety. But yes, it was just about waking people up and getting people going, and music and rhyme and poetry and rhythm easily does that. I mean, you can use it for maths, for goodness sake. It's not necessarily just about languages.

Karen: Yes.

M1: Getting them creating the lyrics as well I think is important.

Karen: Yes.

M1: Because that therefore makes the language theirs. They have their – ownership is theirs of the language, and that's very motivating, isn't it?

F7: I think one of the things I'm going to do when I'm in Albania is I'm splitting them – each area into different blocks. One of the blocks is asking them what they want to learn, what they want to be able to say. So they can then tell me, actually the things that we need are, you know, whatever it's going to be. And it might be that they're thinking there are some really nice English girls coming on holiday here and we want to be able to speak. There are probably things that I can't – haven't even considered. So if you took [those words –

Karen: [How to chat someone up?

F7: Yes.

All: [laughter]

F7: I probably would if I was 14 and – I don't know, anyway, I'm trying to think what a 14-year-old might be thinking. There is a whole age range of children and if they come up with something I haven't thought of, because I'm thinking, they want to bring tourists to the area, so they're going to need to know things like 'Hello, how are you?' You know, dull things that possibly European people would appreciate. But if they got the lyrics from me that were useful and you put them to a song, to a tune they already knew, you're a winner.

M1: Yes.

F7: They could sing their love for these English girls or these American girls who turn up.

Karen: Yes, absolutely.

F7: Yes.

Karen: So, I think – I don't know, we didn't really talk that much about how we would adapt the songs to our own classrooms or some of the benefits, but...well, we kind of did, a tiny bit, as we went along here. But I think we should probably move on. I was going to sort of demonstrate just a few of the activities that I was thinking about, that might be useful for other people that perhaps you may have used before or you may have done something similar, or maybe you haven't. It's just kind of a taste of what I've got in the actual workbooks and I would be very happy to hear your feedback. But first, you are all going to be my students and I will teach you an English song.

[END OF DISCUSSION SECTION]

D.2 Slides created for the afternoon workshop on 12 June 2009

INSTITUTE FOR MUSIC IN HUMAN AND SOCIAL DEVELOPMENT [IMHSD]

Teaching foreign languages through songs

Karen M. Ludke
PhD student, University of Edinburgh
K.Ludke@sms.ed.ac.uk
BA, English and French, University of Michigan, USA
MSc by Research, IMHSD, Music, University of Edinburgh, Scotland

Agenda

- Welcome and outline of the afternoon (5 minutes)
- Group introductions (20 minutes)
- Warm-up activity (5 minutes)
- Presentation of research findings (40 minutes)
- Small-group discussion (10 minutes)
- Group discussion (10 minutes)
- Tea and coffee break (30 minutes)
- Demonstration of workbook activities (20 minutes)
- Discussion of the activities (10 minutes)
- Workshop feedback form (10 minutes)

Introductions

- Name
- Which language(s) do you teach?
- How long have you been teaching?
- What do you hope to gain from today's workshop?
- Briefly, what is your past experience with using songs to teach foreign languages?

Motivation for this Research

- Incorporating the arts into the curriculum can make learning enjoyable and fun, engage students' interest and attention, and increase interpersonal skills, creativity, motivation, and self-expression (Armistead, 2007; Deasy, 2002).
- Positive emotions during learning is beneficial for encoding new language material (Schön *et al.*, 2008).
- Music, and folk songs in particular, may provide an effective tool for improving learners' listening, speaking, and pronunciation skills in modern foreign languages (Spicher & Sweeney, 2007).
- Better memory encoding may result in better recall and retrieval of information, leading to improvements in L2 skills that can generalise to future situations.

Theories of Learning and Approaches for Teaching

- **Audio-lingual method** (use of repetitive, memorised scripts in the target language)
- **Lozanov** (communicative, accelerated learning which uses music to set a receptive mood for L2 learning)
- **Krashen** (idea of 'din,' also called 'involuntary mental rehearsal,' after listening to or reading L2 material)
- **Vygotsky** (social constructivism, relating material to students' own lives and experiences)
- **Kodály method** (teaching music through folk songs in the native language)

Teachers report that music in the foreign language classroom can:

1. Lower anxiety, engage learners' attention, and increase motivation to learn (Murphey, 1990)
2. Guide lesson planning around particular language structures or skills (Medina, 1993)
3. Improve learners' listening comprehension, speaking skills, pronunciation, flow of speech, intonation and rhythm (Fomina, 2000)
4. Enhance cultural awareness and sensitivity (Fomina, 2006; Murphey, 1990)
5. Contribute to the development of the whole person through the aesthetic appeal of music (Medina, 1993)

Songs can also support memory

- Studies have found memory benefits for songs and other musical presentations of verbal material in the native language (Tillmann & Dowling, 2007; Wallace, 1994; Yalch, 1991), including word lists (Thaut *et al.*, 2008).
- Memory for verbal material lasts longer and is relearned more quickly if it was originally learned through a song (Rainey & Larsen, 2002).
- Most studies suggest that the benefits for music are greatest for verbatim verbal memory tasks, perhaps because rhythm and the structure of music can improve the efficiency of neural firing patterns (Thaut *et al.*, 2008).

Wallace (1994)

- Conducted a series of experiments to compare memory for folk ballads learned through song (unaccompanied singing) or speech (presented as poetry with a rhythmic beat in the background).
- Stimuli were presented 5 times and participants were tested 3 times on a written, verbatim recall task.
- Results showed a facilitative effect for the sung presentation, under certain conditions:
  - Simple, easy-to-remember ('catchy') tunes
  - Symmetrically ascending and descending melody
  - Repeated enough times for the melody to be learned

Dowling & Tillmann (2007)

- Participants read or heard either prose or poetry and memory was tested, after different delays, with a repetition of the original phrase, a paraphrase with the same meaning (active vs. passive voice), or a phrase with a different meaning.
- Results showed that memory for surface-level details is better for poetry than for prose.
- Findings for poetry are similar to memory for musical excerpts (Dowling et al., 2001).
- Temporal organisation and rhythm may provide a structure that supports memory for both surface-level details and meaning in verbal material.

Schön *et al.* (2008)

- Statistical learning experiments using Saffran's (1996) paradigm to investigate nonsense word-boundary learning with spoken vs. sung stimuli.
- Three conditions:
  - Spoken, monotone syllables (no pitch differences)
  - Sung, inconsistent syllable-pitch match
  - Sung, consistent syllable-pitch match
- Results showed a strong learning facilitation for both singing conditions, compared to spoken.

Experimental Study

- Undergraduate students completed a one-hour psychology experiment, listening to 20 short Hungarian phrases paired with the English phrase.
- Three ‘listen-and-repeat’ learning conditions:
  - Speaking
  - Rhythmic speaking
  - Singing
- Results showed a benefit for the people who learned the material through the singing method, compared to the rhythmically spoken or spoken methods. In particular, the people who had learned through singing were better able to say the phrases in Hungarian, both immediately and after a 20-minute delay (Ludke, in preparation).

Arts Intervention Study 1

- With the help of the Scottish Executive, we located a secondary school in Scotland with two French classes that were well matched for age, gender, IQ, and previous language experience. The pupils and the French teachers were happy to participate.
- We worked with the pupils for a total of 8 weeks.
- Week 1 and Week 8 were set aside for French language tasks and individual interviews with pupils.
- The arts intervention lasted 1.5 hours per week with each French class, over a 6 week-period.
- The rest of the in-class time (1.5 hours) was taught by the French teachers.

Lesson Design

- Lessons developed for the two classes were designed to be as similar as possible.
- In the music group, we taught and practised French by listening to songs, doing cloze exercises, singing, raps, and the creation of new songs, culminating in a classroom production of songs for a musical theatre piece about a school trip to Paris.
- In the visual art and drama group, we taught and practised French with art games, cloze exercises, drawing, vocabulary bingo, photographs, cartoons, and dramatic dialogues, culminating with the pupils creating a script for a film about a student exchange visit to Paris.

Sample Lessons

Music

- Pupils complete two activities in the textbook to introduce new Paris vocabulary
- Teacher explains song plot and context
- Pupils listen once to the song (by Joe Dassin) for enjoyment
- Pupils listen and fill in missing words (Cloze exercise)
- Teacher defines some of the important new words
- Teacher models spoken lyrics line by line and the pupils repeat them
- Class listens again to the song and sings along to the words of the chorus

Art/Drama

- Pupils complete two activities in the textbook to introduce new Paris vocabulary
- Teacher explains background about Scooby Doo's popularity in France and the cartoon's plot
- Pupils watch the cartoon and write a list of vocabulary words they have learned that they saw in the cartoon (Paris monuments, kinds of food, body parts, 'boy,' 'dog,' etc.)
- Whoever lists the most words receives a Scooby Doo prize

Measures

1. Brief interview about pupils' experience of 8-week arts intervention programme
2. Individual differences (Age, Gender, and IQ as measured by CAT scores)
3. Brief drawing test (artistic skills)
4. Brief singing test (musical skills)
5. Battery of French tests, specifically created for this study, to measure a variety of foreign language skills

Pupils' Comments

– MUSIC CLASS: 'It's been fun and different…if we were to learn new words when we sang, we got the song stuck in our heads and it helped us learn the words. They were really good songs and we got to make up some of our own songs as well.'

– VISUAL ART/DRAMA CLASS: 'It was good, especially when we were watching Scooby Doo. I liked the drawing because it was better than just writing everything down and it doesn't stick in your mind. Drawing makes it easier to remember.'

Improvements in French Skills

- In both classes, pupils' overall performance on French language tasks improved significantly over the 8-week intervention period ($p < .001$)
- The mean score for all French activities in the singing class increased 7.1% (from 55.1% to 62.2%). The mean for the visual art/drama class increased 5.6% (from 52.5% to 58.1%). Gender did not show an impact on overall improvement.
- MANOVA revealed that the main effect of method of instruction on French language skills was significant ($F(1,43) = 4.08$, $p = 0.050$), with the greatest improvements made in the singing group.

Arts Intervention Study 2

- A secondary school in Scotland had two pre-existing French classes that were happy to participate.
- This study was set up as a crossed design for a total of 4 weeks, with 2 weeks spent with a dramatic dialogue or a song as the listening activity and then 2 weeks for the other art form, which was counter-balanced.
- The project consisted of 10-15 minutes at the start of each class period spent listening to and working with a dramatic dialogue or a song, with related activities.
- The arts intervention lasted ~40 minutes per week with each class over a 4-week period. The rest of the in-class instruction time (~2 hours per week) was taught by their regular French teacher, who was a native speaker of French.

Data Collection

- Information collected:
  - Questionnaires about learning French and training and preferences for different art forms
  - Pre/mid/post-test scores on a French-to-English translation task
  - Performance on two Cloze (fill-in-the-blank) tests on the French texts for each 2-week period
  - Two questionnaires about the activities and the Curriculum for Excellence
- Complete results for 43 of 59 pupils, 19 in the S1 class and 24 in the S2 class

Questionnaire Results

- Overall, most of the children in the two classes (92.4%) reported that they enjoyed trying the new artistic materials included in their French lessons over the 4-week intervention period because the activities were both fun and challenging.
- Many children wrote that they felt the activities had helped improve their French skills, particularly in terms of listening and speaking skills. They also felt that the listening, singing, and dramatic speaking activities made them more confident to speak in French.
- More pupils reported that the words of the song repeated in their heads after class than the words of the dialogue did (over 50% in both classes for the songs, compared to 9.4% overall for the dramatic dialogues).
- More of the children preferred listening to the song (over 50% in both classes) compared to the dramatic dialogue (16.7% overall).

Curriculum for Excellence

- Children reported that the activities had helped them to achieve all four of the aims of the Curriculum for Excellence.
- *Successful Learner*: The activity most often cited as an example was learning new vocabulary words.
- *Confident Individual*: Many children reported that speaking or singing in French had helped them achieve this goal.
- *Responsible Citizen*: Active participation in the activities.
- *Effective Contributor*: The most common examples were that they had answered comprehension questions, and speaking or singing out loud as part of the group.

Children's Comments

- 'I enjoyed trying to guess the words in the empty gaps. I enjoyed this because it was quite hard to hear all the words, this was challenging!'
- 'My favourite activity was in a day reading the dialogue we have been learning. It was fun because we got to learn a lot of new words in French.'
- 'Singing all together. It was a really fun song and even more fun to sing it.'
- 'I think I was a successful learner because I widened my vocabulary using the song in just a few days.'
- 'Singing because I felt confident to sing loud and I liked the feeling of confidence.'
- 'My favourite activity was when we said the part of Rachelle in the story [dramatic dialogue] because it made me more confident in reading in French out loud.'
- 'An effective contributor in the recordings because I was loud and I tried to pronounce everything properly.'
- 'When we were recorded because it's minted listening to it.'

French Test Results

- Pre/mid/post-test results on a French-to-English translation task and two Cloze post-tests on the French texts the children had learned in class over each 2-week period were collected.
- Overall grammar and vocabulary translation scores improved significantly from the pre-test to the post-test in both classes, $p < .01$.
- On the Cloze tests, children in the S2 class did not attempt to fill in as many blank lines for the dramatic dialogue as they did for the song, whereas pupils in the S1 class filled in approximately the same number of blanks for both of the Cloze tests (first dramatic dialogue and then song).

Summary

- Overall, the results of these arts intervention studies support the idea that presenting new foreign language material using different presentation methods, including musical and dramatic activities, can provide an enjoyable and challenging addition to French class for beginning-level pupils.
- Artistic activities can also enable pupils to achieve the goals for the Curriculum for Excellence in the modern language classroom.

Overall Conclusions

- Using songs in the modern language classroom may provide positive cognitive and emotional benefits for the learning process.
- Songs in the target language can facilitate the enjoyable repetition and practice of new material, even outwith the L2 classroom.
- Presenting material through songs might also result in improved long-term memory encoding, recall, and retrieval of information in the new language.
- Introducing songs into the modern language curriculum over a 6-week period was particularly helpful for supporting pupils' French language skills, compared to using visual art and drama activities.
- Over a 4-week arts intervention period, pupils reported that they enjoyed the addition of fun and challenging singing and dramatic dialogue activities in their French class. They also believed that the activities had improved their French skills and had made them feel more confident when listening and speaking in French.

Discussion Questions

1. What do you think the value of using songs in the modern language classroom might be for you?
2. What are some challenges you associate with - or have encountered when - using songs in the modern language classroom?
3. Can you think of any ways to resolve any of those challenges?
4. Can you think of any song activities that you could use with your students?

Selected References

1. Armistead, M.E. (2007). Kaleidoscope: How a Creative Arts Enrichment Program Prepares Children for Kindergarten. *Young Children*, 62 (6): 86-93.
2. Chan, A.S., Y.-C. Ho, & M.-C. Cheung. (1998). Music training improves verbal memory. *Nature*, 396 (6707): 128.
3. Deasy, R.J. (2002). *Critical Links: Learning in the Arts and Student Academic and Social Development.* Department of Education, Washington, DC.
4. Fomina, A. (2000). Song melody influence on speech intonation memorization. ICMPC conference proceedings.
5. Fomina, A. S. (2006). Using modern popular songs to teach a foreign language and culture: Enhancement of emotional perception when developing socio-cultural awareness. In 9th ICMPC proceedings. Bologna, Italy: SEMPRE.
6. Gfeller, K. (1983). Musical mnemonics as an aid to retention with normal and learning disabled students. *Journal of Music Therapy*, 20(4), 179-189.
7. Smith Salcedo, C. (2002).*The effects of songs in the foreign language classroom on text recall and involuntary mental rehearsal*. PhD dissertation available at: http://etd.lsu.edu/docs/available/etd-1111102-204823/.
8. Medina, S.L. (1993). The effects of music upon second language vocabulary acquisition. *National Network for Early Language Learning*, 6(3).
9. Murphey, T. (1990). The song stuck in my head phenomenon: A melodic DIN in the LAD?
10. Schön, D. et al. (2007). Songs as an aid for language acquisition. *Cognition.*
11. Spicher, L., & Sweeney, F. (2007). Folk music in the L2 classroom: Development of native-like pronunciation through prosodic engagement strategies. *Connections*, 1: 35–48.
12. Sposet, B.A. (2008). *The role of music in second language acquisition: A bibliographical review of seventy years of research, 1937-2007*. New York: Edwin Mellen Press.
13. Tillmann, B., & W.J. Dowling. (2007). Memory decreases for prose, but not for poetry. *Memory and Cognition*, 35(4): 628-639.
14. Thaut, M.H., D.A. Peterson, K.M. Sena, & G.C. McIntosh. (2008). Musical structure facilitates verbal learning in multiple sclerosis. *Music Perception*, 25(4): 325-330.
15. Wallace, W.T. (1994). Memory for music: effect of melody on recall of text. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*
16. Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, 10(4): 420–422.

Mad Libs®

1. un nom masculin : tapis
2. une préposition : derrière
3. un nom féminin : gomme
4. un adjectif : bleu
5. un verbe à la 2e personne du pluriel : donnez
6. un nom propre : Jacques

Les Champs-Elysees - Joe Dassin

Je m'baladais sur l'avenue, le coeur ouvert à l'inconnu

J'avais envie de dire bonjour à n'importe qui

N'importe qui, et ce fut toi. Je t'ai dit n'importe quoi

Il suffisait de te parler, pour t'apprivoiser.

Aux Champs-Élysées, aux Champs-Élysées,

Au tapis, derrière la gomme, à midi ou à minuit,

Il y a bleu que vous donnez

aux Jacques !

Tu m’as dit « J’ai rendez-vous, dans un sous-sol avec des fous
Qui vivent la guitare à la main, du soir au matin. »
Alors je t’ai accompagnée, on a chanté, on a dansé,
Et l’on n’a même pas pensé à s’embrasser.

Aux Champs-Élysées, aux Champs-Élysées,
Au soleil, sous la pluie, à midi ou à minuit,
Il y a tout ce que vous voulez
aux Champs-Élysées !

Hier soir deux inconnus, et ce matin sur l’avenue,
Deux amoureux tout étourdis par la longue nuit.
Et de l’Étoile à la Concorde, un orchestre à mille cordes
Tous les oiseaux du point du jour chantent l’amour.

D.3 Questionnaires administered for the afternoon workshop, 12 June 2009

Figure D.1: Questionnaire for modern language teachers at the end of the afternoon workshop (page 1)

Teaching foreign languages through songs *Edinburgh, 12 June 2009*

POST-WORKSHOP QUESTIONNAIRE

Name (optional) ___________________________

To help us assess the effectiveness of this workshop, please complete and return this questionnaire to the workshop organiser **before you leave today**. Thank you very much.

Overall

1. How satisfied are you with the workshop content?	**Very Satisfied**	**Satisfied**	**Indifferent or Unsure**	**Dissatisfied**	**Very dissatisfied**
2. How satisfied are you with the method of combining the presentation with participatory discussion/activities?	**Very Satisfied**	**Satisfied**	**Indifferent or Unsure**	**Dissatisfied**	**Very dissatisfied**
3. How satisfied are you with the workshop materials?	**Very Satisfied**	**Satisfied**	**Indifferent or Unsure**	**Dissatisfied**	**Very dissatisfied**
4. How do you perceive the information and assistance received prior to the workshop?	**Very Satisfied**	**Satisfied**	**Indifferent or Unsure**	**Dissatisfied**	**Very dissatisfied**

5. Which portions of the workshop session did you find most useful and effective? Please tick all that are applicable.

_____ Research findings (start of session, presented by the workshop organiser)
_____ Small-group brainstorming of ideas, previous experiences with, and challenges related to the use of songs in the classroom (on large sheets posted around the room)
_____ Group discussion of results of the brainstorming session
_____ Sample activities (workshop organiser demonstrated some of the materials in workbooks)
_____ Group discussion of activities and how to adapt them to other classroom environments
_____ Printed workbook materials

Outcomes

6. How familiar are you now with:					
Academic research in this area	**Fully**	**To a good extent**	**Somewhat**	**Very little**	**Not at all**
Practical tips and lesson ideas	**Fully**	**To a good extent**	**Somewhat**	**Very little**	**Not at all**
Challenges and how to overcome them	**Fully**	**To a good extent**	**Somewhat**	**Very little**	**Not at all**
7. How much has this workshop helped increase your practical knowledge of how to teach foreign languages using music and songs?	**Fully**	**To a good extent**	**Somewhat**	**Very little**	**Not at all**
8. How much do you think you can apply what you learned from the workshop to your teaching?	**Fully**	**To a good extent**	**Somewhat**	**Very little**	**Not at all**
9. To what extent will you be able to teach your colleagues about the topic?	**Fully**	**To a good extent**	**Somewhat**	**Very little**	**Not at all**

Figure D.2: Questionnaire for modern language teachers at the end of the afternoon workshop (page 2)

10. How do you think you can apply what you have learned during this workshop in your classroom?

____

____

____

11. What was one of the greatest benefits to you from this workshop?

____

____

____

12. Are there any topics that you wish had been covered in more depth?

____

____

____

13. Are there any additional topics do you wish that the workshop had covered?

____

____

____

14. Do you have any other comments or suggestions?

____

____

____

____

____

Figure D.3: Email questionnaire sent to modern language teachers two weeks after the afternoon workshop

INSTITUTE FOR MUSIC IN HUMAN AND SOCIAL DEVELOPMENT [IMHSD]

Karen M. Ludke
University of Edinburgh
IMHSD, Music
12 Nicolson Square
Edinburgh EH8 9DF

2-WEEKS POST-WORKSHOP QUESTIONNAIRE

Name (optional) ___________________________

To help us assess the value of the workbook distributed at the 'Teaching foreign languages with songs' workshop on 12 June 2009, it would be very helpful if you could please complete and return this questionnaire at your earliest convenience to the workshop organiser via email (K.Ludke@sms.ed.ac.uk) or post (at the address above). Thank you very much.

Overall

1. How satisfied are you with the workbook content?	**Very Satisfied**	**Satisfied**	**Indifferent or Unsure**	**Dissatisfied**	**Very dissatisfied**
2. How satisfied are you with the workbook layout? (e.g., is it easy to use and to find activities?)	**Very Satisfied**	**Satisfied**	**Indifferent or Unsure**	**Dissatisfied**	**Very dissatisfied**
3. How much did the workshop help increase your practical knowledge of how to teach foreign languages using music and songs?	**Fully**	**To a good extent**	**Somewhat**	**Very little**	**Not at all**
4. Do you think the workshop format was a helpful way to learn about this topic and to share ideas with other modern language teachers?	**Fully**	**To a good extent**	**Somewhat**	**Very little**	**Not at all**

5. Which parts of the workshop have you found to be most useful in your own teaching practice? Please tick all that are applicable.

_____ Research findings (start of session, presented by the workshop organiser)
_____ Small-group brainstorming of ideas, previous experiences with, and challenges related to the use of songs in the classroom (on large sheets posted around the room)
_____ Group discussion of results of the brainstorming session
_____ Sample activities (workshop organiser demonstrated some of the materials in workbooks)
_____ Group discussion of activities and how to adapt them to other classroom environments
_____ Printed workbook materials

6. How have you applied what you learned from the workbook in your classroom? If you have not tried any of the activities yet, can you tell us why?

____

____

7. Do you have any other comments or suggestions?

____

____

____

Appendix E

DVD with French and English subtitles created for children at the end of the arts intervention study