Music and Language: Exploring the Acoustic Dimension of ESL Silent Reading Comprehension Through Music Perception, Phonological Awareness and Auditory Working Memory

Tianai Zhang

St Cross College

University of Oxford

Thesis submitted in partial fulfilment of the requirements for the degree of

Master of Science in Applied Linguistics and Second Language Acquisition

Trinity 2024

DECLARATION BY THE CANDIDATE AS AUTHOR OF THE DISSERTATION

1. I understand that I am the owner of this dissertation and that the copyright rests with me unless I specifically transfer it to another person.
2. I allow the Department to deposit on my behalf a copy of this dissertation in the Oxford University Research Archive ('ORA') where it shall be freely available online for use in accordance with ORA's Terms and Conditions of Use [https://ora.ox.ac.uk/terms_of_use].
3. I understand that this dissertation should not contain material that can be used to personally identify individuals or specific groups of individuals (unless permission has been obtained from the individuals) and that such material should be removed before this dissertation is deposited in ORA.
4. I agree to be bound by the terms of the ORA Grant of Non-exclusive Licence [https://ora.ox.ac.uk/deposit_agreements] and I warrant that to the best of my knowledge, making my thesis available on the internet will not infringe copyright or any other rights of any other person or party, nor contain defamatory material.
5. I agree that my dissertation shall be available for download in ORA in accordance with paragraphs 2, 3 and 4 above.

Signed [an electronic signature is sufficient]:	Tianai Zhang
Date:	14th August, 2024

Table of Contents

List of Tables

List of Figures

Acknowledgements

In the words of my favourite writer, Albert Camus, “The struggle itself toward the heights is enough to fill one’s heart.” Reflecting on the journey of conducting research and writing this thesis, I find a deep resonance in this sentiment. The challenges and perseverance required have made this experience profoundly fulfilling.

My deepest gratitude goes out to my supervisor, Dr. Anna-Maria Ramezanzadeh, for her much-appreciated suggestions and patience throughout my thesis. She guided me so positively that I always felt confident in my abilities after holding a conversation with her. I have great admiration for her exceptional logical thinking and problem-solving skills. During this year, I have learned so much from her approach to tackling complex issues, and I feel very fortunate to have Anna-Maria as my supervisor.

Furthermore, I am extremely grateful to Prof. Robert Woore, Dr. Faidra Faitaki, and Cate. They generously provided their knowledge and expertise, which allowed my research to go the extra mile. Robert’s course on reading and phonology sparked my interest in this research area and he provided invaluable advice at the outset of my study. Faidra also offered insightful recommendations for my research tools during the formative stages of my project. Additionally, Cate’s statistics course is exceptionally clear and accessible, and her statistical guidance has been crucial in refining my research methodology.

I thoroughly enjoy this year’s courses and find all of them to be exceptionally beneficial. Prof. Elizabeth Wonnacott and Prof. Hamish Chalmers present the intricate concepts of L1 and bilingual language acquisition with exceptional clarity, significantly enhancing my comprehension of language development. Prof. Heath Rose’s explanations on how language shapes and is shaped by societal contexts, both locally and globally, are truly engaging and enlightening. I cannot list all their names here, but every one of my teachers has enriched my life, both academically and in many other ways.

I also would like to express my thanks to the students who participated in this research and to the teachers who supported me on-site. Their involvement and dedication were crucial to the success of this study. Additionally, I am deeply thankful to Molly and Rob, the native English speakers who recorded the audio for the phonological awareness and auditory working memory tests. Their contribution is instrumental in advancing this research.

To conclude, I would like to thank my parents and friends for their constant love and encouragement. Whenever I faced difficulties, my parents were always there for me via video calls, offering various ways to lift my spirits. My friends also regularly reached out to check on me through messages and sent thoughtful care packages with snacks and essentials.

Looking ahead to my academic career, I am filled with a sense of boundless anticipation and enthusiasm. I am eager to embark on new discoveries, embrace opportunities for growth, and contribute meaningfully to my field. May everything be pleasant to all involved.

Abstract

A large body of research has explored the intersection of music and language but mostly focused on the area of first language (L1) acquisition in early childhood. The few studies on second language (L2) learning and music have mainly concentrated on listening and speaking skills (Jekiel & Malarski, 2021; Talamini et al., 2018). Therefore, this study aims to address this gap by investigating how music perception relates to L2 reading comprehension in terms of the acoustic dimension.

This study involves Chinese adult undergraduates who learn English as a second language (ESL). In a supervised classroom, 139 participants completed the background information questionnaire, music perception test, phonological awareness test, and auditory working memory test on electronic devices and took the English reading comprehension test with pen and paper. They were divided into three groups based on their past and present involvement in musical activities: no training, basic training, and advanced training. Quantitative data were analysed using a correlation matrix, multivariate analysis of variance, and partial least squares (PLS) modelling. Participants with basic and advanced musical training outperformed their untrained counterparts in phonological awareness. Specifically, musical training enhanced their ability to segment sounds in words and blend word parts. The auditory-based PLS model revealed that phonological awareness directly predicted L2 reading comprehension, and that music perception affected reading comprehension directly and indirectly through auditory working memory. Overall, variations in L2 silent reading comprehension are well explained by this model. A rigorous pilot study, sufficient sample size, diverse and reliable research tools, and advanced data analysis enhance the robustness and generalisability of the results.

This study offers valuable evidence to support further investigation into the effectiveness of music-related interventions for L2 reading. It holds considerable potential for both pedagogical and therapeutic applications. For instance, future research can explore the integration of musical elements into reading strategy instruction and the use of background music for treating reading disabilities.

1. Introduction

Humans use both language and music to exchange information and express feelings. Patel (2012) emphasises that, based solely on neuropsychological evidence, one might conclude that music and language are independent cognitive functions. However, neuroimaging has challenged this view by revealing similarities in individuals' abilities regarding certain aspects of music and language processing. The interaction between music and early L1 acquisition has also been well studied in behavioural research (e.g., Brandt et al., 2012; Schön et al., 2008). Notably, Patel's (2011) OPERA hypothesis based on available neurological evidence explains why music plays a key role in positively influencing linguistic abilities. The theory suggests that musical training induces adaptive plasticity in the neural networks involved in speech processing through five key aspects: (1) enhancing brain networks that are overlapped between music and language processing (O); (2) improving brain functions that ensure precise processing (P); (3) activating brain networks associated with strong motivation and emotional engagement (E); (4) modifying brain networks through consistent repetition (R); and (5) involving brain networks responsible for maintaining focused attention (A).

However, Sun et al. (2024) express concern that the OPERA hypothesis may not fully explain the development and acquisition of L2. They highlight that brain activation patterns for L2 can be shaped by the learner's L1 (Cao et al., 2013; Tan et al., 2003) and influenced by the language's orthographic depth (Assche et al., 2012). Therefore, cross-linguistic comparisons of learners with different L2s are needed to understand their performance in music and language learning, and further analyse different mechanisms by which musical training may affect these groups.

The growing emphasis on research on integrating music into children's L2 classrooms within Indo-European language contexts, highlights the lack of studies focusing on adult L2 learners and those whose L1s are not Indo-European. Additionally, the interplay between music and L2 reading remains underexplored, as the shared acoustic resources between L2 listening, speaking, and music have garnered more attention from researchers, leading them to focus more on the visual aspects of L2 reading, such as lexical decoding, while neglecting auditory factors like the potential impact of music on reading skills. These factors together provided a strong impetus for this study.

This thesis is organised into six chapters. The following literature review clarifies the definitions of musicality and L2 reading comprehension, and critically reviews the literature investigating the nature of relationship between musicality and two acoustic precursors of reading. These studies are discussed in detail within different contexts (i.e., L1 and L2 reading) and via methodological designs (i.e., longitudinal and cross-sectional designs). The final section of literature review chapter highlights significant methodological improvements for the field, which this current study aims to address through a partial replication of Foncubierta et al.'s (2020) research, a major study exploring how music perception affects silent reading fluency in 117 Italian adult learners of Spanish.

The methodology chapter outlines the participants' profiles, research design, and instruments. Five instruments were designed and employed in response to three research questions:

1) What are the correlations between L2 silent reading comprehension (RC), music perception (MP), phonological awareness (PA), and auditory working memory (AWM)?
2) What are the relationships between participants' musical training backgrounds (no training, basic training, advanced training) and their performance on MP, RC, PA, and AWM?
3) What are the structural relationships among RC, MP, PA, and AWM?

These research questions explore the relationship between musical perceptive skills and the acoustic dimension of L2 reading. The findings chapter presents the results for each question. The correlation matrix is created for the first question, discovering positive corelations among four variables. The MANOVA is conducted for the second question, finding significant between-group differences in participants' performance on PA. Finally, a partial least squares model is established to explore the individual and combined explanatory power of MP, PA, and AWM on RC, as well as their interactions.

Following the findings chapter is the discussion chapter, which addresses the key results in turn, drawing on important theories and existing empirical evidence. It also discusses potential implications for teaching L2 reading to typically developing adult learners and treating L2 reading disabilities, outlines the study's limitations and proposes corresponding suggestions. The conclusion chapter summarises the study's key findings, highlights its strengths, and points out directions for future research.

2. Literature Review

This chapter first clarifies the definitions of musicality and reading comprehension, with relevant discussion of literature. Detailed critiques of studies exploring the relationship between musicality, music instruction, and reading abilities are then provided, culminating in the rationale for the current research.

2.1 Musicality

2.1.1 Musical Gifts and Training

Musicality is often used interchangeably with many terms in the literature, such as "musicianship", "musical expertise" and "musical ability". At a macro level, an individual's musicality is a consequence of both their musical gifts and acquired exposure/training (Swaminathan & Schellenberg, 2020; Schellenberg et al., 2023). Specifically, musical gifts refer to the innate ability to discriminate features such as melody and rhythm, and the capacity for long-term musical memory retention (Swaminathan & Schellenberg, 2020). Musical exposure/training refers to unconscious daily exposure (e.g. listening to music playing in the mall while shopping), as well as active participation in music-related activities (e.g. playing instruments) (Hargreaves, 2002).

2.1.2 Music Perception and Production

In addition to the above nature-nurture divide, musicality as a skill can be interpreted in terms of perception and production (Li et al., 2022; Politimou et al., 2019). Music perception is concerned with the ability to perceive pitch, melody, rhythm, and speed, whereas music production mainly refers to singing or producing songs (Politimou et al., 2019). In previous studies measuring music perceptive skills, the most frequently used instruments include the Musical Hearing Test (Mandell, 2009), the Seashore Musical Aptitude Test (Seashore et al., 1960), and the Profile of Music Perception Skills (PROMS, Zentner & Strauß, 2017). Musical productive skills, on the other hand, are measured by having participants sing familiar songs and then having professional musicians rate them on rhythm and melody (Christiner & Reiterer, 2013; Christiner, 2022). In this current study, participants' music perceptive skills were determined by their performance on the PROMS test and were thought to be shaped by both inherent music abilities and informal exposure to music.

2.2 Reading

2.2.1 Reading Comprehension & Fluency & Accuracy

Reading is the process by which the reader absorbs information from a text and constructs meaning in their mind (Grabe & Stoller, 2019; Kintsch & Kintsch, 2005; Mullis & Martin, 2019). Reading fluency, generally considered to comprise three elements (i.e., reading rate, accuracy, and prosody), is defined as accurate reading of texts at a conversational rate with appropriate prosody (Armbruster et al., 2001; Hudson et al., 2008). In addition to this well-recognised definition, there are two groups of scholars who propose slightly different definitions. Daane et al. (2005) describe fluency as the accurate understanding of the author's phrasing, syntax, and expression, that is, the combination of accuracy and prosody. One possible explanation for their removal of the rate element is that when the reader is trying to construct meaning very quickly in a short period of time, he tends to overlook things related to prosody. Perhaps for Daane et al. (2005), the ideal reading does not need to be fast. On the other hand, other scholars have equated fluency directly with rate, that is, separate from accuracy (Pilonieta, 2012; Skinner, 1938).

To avoid further confusion of the concepts of fluency in previous literature, silent reading comprehension in this study is defined as rapid and accurate reading (Figure 2.1). There are two main reasons for this definition. First, prosody is not easy to measure for silent reading. According to Holmes and Allison (1986), reading has four modalities: oral reading to an audience, oral reading to oneself, silent reading, and silent reading while listening. Participants in this study are required to complete a time-limited reading test in a quiet environment to assess their ability to achieve high accuracy within a stipulated time. When students read aloud, prosody appears easier to measure and more important since it reflects their expressive ability and deep understanding of the text. Second, it is difficult to objectively evaluate prosody. Prosody is shaped by readers' different reading habits and speaking styles. Specifically, when individuals silently read the same sentence in their minds, the tone and emotion of their inner voice may vary. In this case, defining what constitutes suitable prosody is challenging.

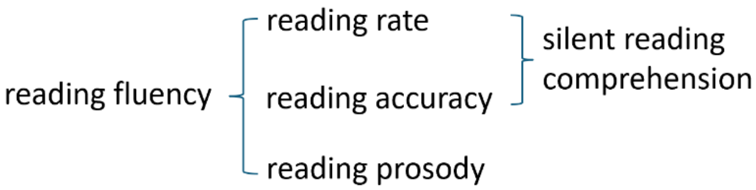


Figure 2.1. Definition of Silent Reading Comprehension

2.2.2 Underlying Knowledge and Abilities Involved in Reading

Having examined the definition of reading comprehension, it is essential to delve into the underlying knowledge and abilities involved. Understanding these foundational elements will provide deeper insights into the core processes that contribute to reading comprehension. Jeon and Yamashita (2014) conducted a groundbreaking meta-analysis on correlates of reading comprehension. Their findings validate the significance of the following ten subcomponents: L2 decoding, L2 vocabulary knowledge, L2 grammar knowledge, L2 phonological awareness, L2 orthographic knowledge, L2 morphological knowledge, L2 listening comprehension, L1 reading comprehension, working memory, and metacognition. Other researchers have identified different subcomponents which constitute the knowledge and skills required for reading. For example, 'rapid automatised naming', 'parsing', and 'mental lexicon' are also often highlighted in the literature (Graesser et al., 1997; Stevens et al., 2019; Ziegler & Goswami, 2005). For the four aforementioned reading modalities (Holmes & Allison, 1986), these underlying abilities and knowledge are interconnected but vary slightly across each modality.

In general, skills and knowledge required for reading can be categorised into visual and auditory domains. In the auditory domain, previous literature suggests that two core acoustic skills are closely related to L2 silent reading: L2 phonological awareness and auditory working memory (Baddeley et al., 1985; Birch & Fulop, 2020; Kato, 2009; Koda, 2007; Melby-Lervåg & Lervåg, 2011; Özata et al., 2016; Strait et al., 2011). Phonological awareness involves an implicit understanding of the target language's phonological system, allowing individuals to recognise the internal sound structure of words at the phoneme and phonology levels (Coumel et al., 2019; Januri et al., 2022). Auditory working memory serves as a temporary storage system that enables the accurate manipulation and integration of both linguistic and musical information (Foncubierta et al., 2020; Strait et al., 2011).

2.3 Role of Working Memory in Connecting Reading and Music

Both musical and reading activities require working memory. Figure 2.2 illustrates two types of working memory (WM) concerning language and music. Like language, music perception can occur through two distinct activities: listening to music and reading music notation. Reading music on a score involves visual WM, which operates differently from the auditory WM system (Baddeley, 1992; Fennell et al., 2021; Shah & Miyake, 1996; Silas et al., 2022). For non-music professionals, reading music notation can be challenging as it requires specialised knowledge and training. Consequently, they may primarily understand and appreciate music through listening, a skill often developed through informal exposure.

Since both music perception and silent reading rely on auditory WM, this current study explores an intriguing issue, that is, whether music perception affects reading comprehension through this cognitive factor. From a macro level, this current study further investigates the intersection between music perception and the acoustic dimension of silent reading.

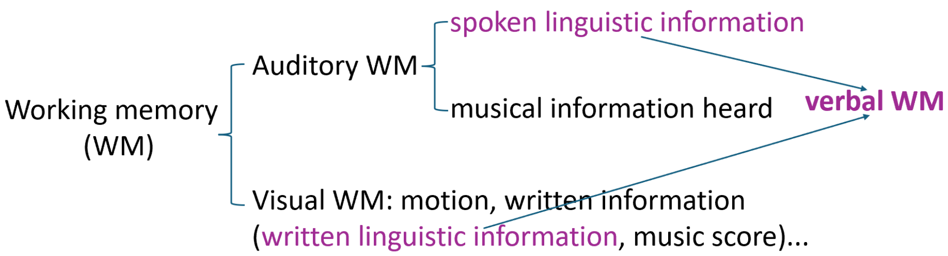


Figure 2.2. Classification of Working Memory

2.4 Musicality and Reading Abilities

2.4.1. Musicality and Verbal Working Memory

As mentioned above, WM can store linguistic information, musical information, and spatial information. However, this study revolves around the interaction between music and reading. Therefore, this section narrows its focus on the literature investigating the relationship between WM elicited by linguistic stimuli (termed verbal WM in this current study) and musicality. It should be noted that all verbal WM tests mentioned in this section were conducted in participants' L1s.

2.4.1.1. Cross-sectional research: Positive correlation between musicality and verbal working memory

Talamini et al. (2017) conducted a meta-analysis encompassing 29 cross-sectional studies published between 1987 and 2017 to explore whether musicians perform better than non-musicians on memory tasks. They found that musicians had significantly better WM, with a greater advantage under musical stimuli, a moderate advantage under spoken linguistic stimuli, and little to no advantage under visual stimuli. The finding on visual stimuli is supported by Okhrei et al. (2016) who also found no differences between the two groups in short-term memory tasks where verbal stimuli were presented visually. Similarly, Talamini et al. (2016) reported that musicians outperformed non-musicians on digit span tasks where numbers were presented orally, but performed worse on visually presented numbers. In response, Talamini et al. (2017) offered two explanations for musicians' superior memory for spoken linguistic stimuli. First, people born with a good memory are more likely to pursue music and become musicians, in comparison to those who struggle with memory. Second, musicians' better memory may benefit from long-term training to play an instrument that improves their ability to recall tonal stimuli. Since musicians are better at processing auditory stimuli than non-musicians (Rammsayer & Altenmüller, 2006; Spiegel et al., 1984; Tervaniemi et al., 2005), they equally perform well on orally presented linguistic stimuli. However, this meta-analysis has two limitations. First, there is no uniform standard for classifying musicians and non-musicians, and the length of musical training received by musicians is not clearly reported in many of the included studies, which may lead to heterogeneity. Second, some studies also did not control for confounding variables that could arise between the two groups. For example, differences found in memory abilities may not be due to musical training but rather to factors such as intelligence and personality.

Following Talamini et al.'s (2017) meta-analysis, more cross-sectional studies have emerged, exploring the relationship between musical experience and verbal WM. Most of these studies involve adult populations (e.g., Arndt et al., 2023; D'Souza et al., 2018; Gagnon & Nicoladis, 2021), and a smaller number focus on children and adolescents (Degé & Schwarzer, 2017; Jain et al., 2019).

Five of the studies used a similar research design where groups of musicians and non-musicians were recruited to complete verbal WM tests and their performance was statistically analysed using either T-tests or ANOVA (Arndt et al., 2023; Gagnon & Nicoladis, 2021; Garcia et al., 2020; Nisha et al., 2021; Taylor & Dewhurst, 2017). Notably, the definition of musicians and non-musicians also varied across these studies. Some of them defined musicians as professionals engaged in music-related work (Arndt et al., 2023), while in others, musicians were those who received a certain amount of musical training (Gagnon & Nicoladis, 2021; Garcia et al., 2020; Nisha et al., 2021; Taylor & Dewhurst, 2017). These five studies demonstrated that adults with musical experience performed better on the verbal WM tasks, supporting a positive relationship between musicality and verbal WM.

Unlike the design of five studies above, an interesting study by D'Souza et al. (2018) investigated the unique contributions of long-term musical training and L2 learning to verbal

WM, based on previous literature suggesting that both are cognition boosters. The researchers divided 153 adults into four groups: monolingual musicians, bilingual musicians, bilingual non-musicians, and monolingual non-musicians. Monolinguals are English native speakers with little L2 training, and bilinguals were proficient in English and at least one additional language. Musicians were defined as those having at least eight years of music-playing experience and maintaining regular practice (6-9 hours weekly). Their findings indicated that musical training was associated with enhanced verbal WM. However, no advantage was observed for bilinguals in verbal WM tasks, nor was any combined benefit of L2 learning and musical training detected. The key strength of this study is the use of multiple tasks (i.e. digit span backward, reading span, and operation span) to measure verbal WM, overcoming the common limitation of previous research that relied on a single task (e.g., Paap & Greenberg, 2013). A diverse set of tasks offers a comprehensive representation of verbal WM, enhances the robustness of their findings, and demonstrates that musical training, rather than L2 learning, positively predicts verbal WM.

Degé and Schwarzer (2017) explored the possible mechanisms that might explain the connection between musicality and verbal WM among younger participants. That is, whether an enhanced articulatory rehearsal mechanism (i.e. a cognitive process within the WM phonological loop where individuals silently repeat verbal information) could explain the better verbal VM performance of music-trained children. After controlling for gender, socioeconomic status, intelligence, motivation, and personality, they found that musically trained children significantly outperformed untrained children under normal conditions. However, this advantage disappeared under the condition of suppressing their articulatory rehearsal mechanism. Therefore, they believe that the articulatory rehearsal mechanism is very likely to play an important role in giving outperformers advantage. Due to the cross-sectional rather than longitudinal design, they cannot determine whether musical training predicts an enhanced articulatory rehearsal mechanism and better verbal WM. Despite this, Degé and Schwarzer's (2017) dual-condition design, which included both normal and articulatory suppression conditions, was particularly innovative and allowed for a nuanced examination of the underlying mechanisms at play. Additionally, their careful control of confounding variables enhanced the robustness of their findings. For example, after deliberating whether to use non-verbal or verbal intelligence tests as control variables, they ultimately decided on non-verbal intelligence tests. This decision was reasonable, given the rationale that verbal intelligence tests might be influenced by verbal WM, which musical training is hypothesised to enhance. Using verbal intelligence tests in this context could potentially confound the dependent variable (verbal WM) with the control variable (intelligence), thereby leading to inaccurate results.

2.4.1.2. Longitudinal research: Music instruction enhances verbal working memory

In contrast to these cross-sectional studies, three longitudinal studies investigated the impact of music intervention on verbal WM among children and adolescents (Nie et al., 2022; Saarikivi et al., 2019; Shen et al., 2019), with fewer studies focusing on adults (Yeşil & Ünal, 2017). Overall, these longitudinal studies support that musical training predicts improvements in verbal WM. Participants in Shen et al.'s (2019) study were preschool children with an average age of 4 years, younger than those in the other two studies, who were between 7 and 17 years old.

Furthermore, although all three studies are longitudinal, their experimental designs differ. Apart from the intervention group that received musical training, Shen et al. (2019) had the only control group participate in typical classroom activities, whereas Nie et al. (2022) set up two control groups: a L2 training group and a passive control group. Nie et al.'s (2022) intervention lasted one year, whereas Shen et al.'s (2019) intervention spanned only three months. However, Shen et al. (2019) included an additional delayed post-test three months after the intervention ended, which Nie et al. (2022) did not. Additionally, Shen et al. (2019) measured verbal WM solely using the digit span backward task. In contrast, Nie et al. (2022) employed a more comprehensive approach by using the digit span forward (recalling numbers in order) and backward (recalling numbers in reverse order) tasks, along with the vocabulary and block tests from the Wechsler Intelligence Scale for Children IV. A major strength of Shen et al.'s (2019) study is that they tracked the subsequent effects of musical training and provided a convincing explanation. In follow-up interviews, they found that children enjoyed music and therefore continued to practice at home. This ongoing spontaneous practice could explain their enhanced verbal WM.

Saarikivi et al. (2019) measured the participants' verbal WM every two years for six years, enrolling a new group of seven-year-olds each time. This means that their study included both new participants and repeated measurements of previously recruited participants. To be more specific, in 2011, a group of 7-year-olds is recruited and measured. In 2013, the 2011 group (now 9 years old) is measured again, and a new group of 7-year-olds is recruited. In 2016, the 2011 group (now 11 years old) and the 2013 group (now 9 years old) are measured again, and still a new group of 7-year-olds is recruited and measured. Their measurements of verbal WM include digit span forward and backward tasks. They found that musically trained children performed better on the digit span forward test compared to their non-trained peers, while there were no group differences in the backward test. Interestingly, this finding contrasts with Nie et al. (2022), who found that the music intervention group outperformed the untrained group only in the backward test, with no significant between-group differences in the forward test.

Saarikivi et al.'s (2019) work was supported by Lee et al. (2007) and Hansen et al. (2013), while Nie et al.'s (2022) research also received backing from Guo et al. (2018). It is noteworthy that in Lee et al.'s (2007) study, musically trained adults outperformed their untrained counterparts in the forward test, while musically trained 12-year-old children outperformed their untrained peers in both the forward and backward tests. Therefore, age may be an important moderating factor. In any case, there is evidence to suggest forward and backward tests involve different verbal WM processes. As early as 1997, Reynolds discovered through factor analysis that forward and backward tasks engage different cognitive functions. This has also been acknowledged by Saarikivi et al. (2019) and Nie et al. (2022).

Specifically, Saarikivi et al. (2019) believe that forward test requires the active maintenance of rules in short-term WM, while the backward test necessitates the manipulation and updating of information. Thus, they concluded that in late childhood and early adolescence, the relationship between musical training and increased WM capacity is stronger than that with updating information in WM. However, they merely emphasised their study's large statistical power, making the negative results in the backward test group differences highly credible. In other words, they believe that even if there were unobserved differences, these differences would be extremely subtle and of no practical research value. Unfortunately, this explanation is not tenable since interpreting results solely based on statistical forces ignores many important theoretical and practical factors.

By contrast, Nie et al. (2022) offered a more intriguing and plausible explanation. In their study, participants speak Mandarin, recognised as a tonal language. In this case, regardless of whether they had musical training, digit sequences sounded melodic to Mandarin-speaking children. In other words, daily exposure to speaking and hearing sentences with melodies may potentially enhance auditory memory in these children compared to those who do not speak tonal languages. Consequently, in the forward task, the additional effects of musical training were overshadowed by the tonal language effect. However, in the backward task, the advantage of tonal language melody diminishes because this task requires more manipulation and updating of information. Musical training may gain the upper hand in this scenario by enhancing the central executive system of verbal WM. In summary, the results regarding whether musical training is associated with improvements in verbal WM measured by forward or backward tests, or both, are mixed. Hence, it is crucial to further explore whether there are significant between-group differences on these two memory tests in different contexts and among other populations.

2.4.2. Musicality and L1 Phonological Awareness

Phonological awareness (PA) is a crucial acoustic skill for silent reading, and its relationship with musicality in the context of L1 has been extensively studied. Most of the research were conducted in English-speaking contexts, with some studies involving PA in other languages, such as German and Portuguese. This section reviews key studies in this field.

2.4.2.1. Cross-sectional research: Positive correlation between musicality and L1 phonological awareness

There are numerous cross-sectional studies examining whether musicality can predict L1 PA, but there is currently a lack of robust meta-analysis to integrate them. To this end, Table 2.1 is offered to provide a comprehensive and clear overview of the evidence in this field.

Table 2.1 Overview of Cross-sectional Research on the Relationship Between L1 PA and Musicality

Study, Year	Language	Participants (mean age/ age range in years, sample size)	Operationalisation of Musicality	Operationalisation of PA
Lucas & Gromko, 2007	English	6-7 (n=27)	• tonal discrimination • rhythm discrimination	• sound segmentation • nonsense word production
Cardillo, 2008	English	1-5 (n=23)	• tonal/ melodic discrimination • rhythm discrimination	• rhyming discrimination • rhyming production • sentence segmentation • syllable segmentation
Loui et al., 2011	English	7.6 (n=32)	pitch perception and production	• sound categorisation • compound deletion

Degé et al., 2015	Germany	6.3 (n=55)	• music perception: melody, pitch, rhythm, tone length, meter • music production: singing a song, rhythm production, meter execution	• rhymes • word segmentation • phoneme synthesis • phoneme recognition
Culp, M. E., 2017	English	7.5 (n=17)	• tonal discrimination • rhythm discrimination	• rhyming • segmentation • isolation • deletion • substitution • blending
Sun et al., 2017	English	18-24 amusia (n=20) normal (n=20)	• pitch discrimination • rhythm discrimination	elision
Steinbrink et al., 2019	Germany	5.7 (n=54) 8.7 (n=96)	• pitch perception • contour perception • tempo perception • rhythm perception	• syllabic vowel onset detection • rhyming • consonantal onset detection • phoneme synthesis • phoneme analysis • phoneme identification • phoneme elision • phoneme substitution[1]
Bonacina et al., 2020	English	6.4 (n=55)	• drumming to an isochronous beat • tapping rhythmic patterns	• elision • blending words
Degé et al., 2020	Germany	9-12 (n=45)	• singing	• segmentation of pseudo words • determination of vowel length • words backwards repetition
Sousa et al., 2022	European Portuguese	7.3 (n=74)	• melodic discrimination • rhythm discrimination	recognition and production of shared syllables in words
Rimmer et al., 2024	English	8.9 (n=21, autistic)	beat perception	• word segment recognition • sound/syllable manipulation

Overall, all studies presented in the Table 2.1 acknowledge that musicality positively correlates with L1 PA. Most studies share similar designs and control for variables such as socioeconomic status and intelligence. Sun et al. (2017)'s study is slightly unique in that they compared 20 typically developing adults and 20 adults with congenital amusia, who had significant deficits in pitch and rhythm perception, and found similar performance in PA tests between the two groups. They further analysed data from pitch discrimination tasks, revealing eight people with

[1] The two groups of participants, preschoolers (marked in green) and school children (marked in purple), were given different tasks to measure PA due to their differing levels of formal reading instruction and cognitive development stages.

severe pitch impairment had significantly worse PA compared to all other participants. Lastly, their hierarchical regression analysis indicated that within the musical tests, pitch discrimination predicted PA more strongly than rhythm discrimination. While interesting, this finding should not be generalised hastily, as conclusions drawn from atypical populations may not hold true among typically developing individuals.

Notably, in addition to demonstrating that both rhythm and melody perception predict PA, Sousa et al (2022) revealed that PA mediates the impact of rhythm perception on reading ability in first-grade children. Interestingly, melody perception showed no significant correlation with reading ability. However, this finding is contrary to Steinbrink et al.'s (2019) result that although rhythm and pitch perception were significant predictors of PA in stepwise regression analysis, only rhythm perception contributed to literacy skills when PA was controlled for. This could be due to differences in measurement tools and participants' different L1s.

Taken together, since existing cross-sectional studies have primarily focused on L1 PA, one of research gaps addressed by this current study is the lack of empirical evidence concerning the relationship between musicality and L2 PA.

2.4.2.2. Longitudinal research: Music instruction enhances L1 phonological awareness

After discussing cross-sectional studies, the focus now shifts to longitudinal studies, for which meta-analyses and systematic reviews have been conducted. Gordon et al. (2015) conducted two meta-analyses to separately explore the impact of musical training on children's PA and reading fluency. Their inclusion criteria required that the studies implemented interventions with control groups (excluding within-group interventions, observational, and correlational studies). Their results indicated that musical training enhanced PA but did not improve reading fluency.

Even better, Gordon et al. (2015) divided the results on PA into two categories: rhyming and other phonological outcomes, incorporating many moderating factors such as the duration of musical training. They found a strong correlation between rhythmic skills and PA, with the number of training hours influencing the positive effect of music interventions on rhyming outcomes. However, the additional meta-analysis on other phonological outcomes revealed a small effect size ($d = 0.2$), with no significant moderating factors (i.e. age or training duration). This might be due to the inclusion of various types of phonological tasks (e.g., initial phoneme oddity, alliteration) and participants' diverse L1s. Therefore, a plausible inference is that rhythmic skills gained from musical training positively transfer to rhyming awareness, whereas other aspects of PA require further, more targeted research.

On the other hand, the meta-analysis evaluating the impact of musical training on reading fluency did not yield a significant effect size, and only five studies were included, making moderator analysis unfeasible. Regarding why musical training predicts PA rather than reading fluency, it is commonly believed that musical skills and PA share an auditory foundation (Anvari et al., 2002; Lathroum, 2011). Interestingly, Bus and van IJzendoorn (1999) offer another possible explanation: PA is a necessary but not sufficient condition for reading since they found in their meta-analysis that short-term PA training had a moderate impact on reading fluency, but long-term studies only produced a small impact.

Although the results supporting positive effects of musical training on PA are based on Gordon et al.'s (2015) strict inclusion criteria and consideration of moderating factors, the heterogeneity and limited number of included studies prevent the conclusions from being generalisable. Specifically, the activities conducted by control groups in the original studies varied widely, and many studies did not confirm group equivalence, failing to report which confounding variables were controlled. Furthermore, most studies adopted the quasi-experimental design rather than randomised control trial to create treatment and control groups. These issues can either diminish or exaggerate the benefits of musical training. Most importantly, only 13 studies were included, resulting in insufficient power for moderator analyses, making it challenging to distinguish different subcategories and understand how various study features might influence outcome trends.

Apart from these two insightful meta-analyses, the systematic review by Eccles et al. (2021b) further compared the influence of mainstream music education methods–Orff, Kodály, Suzuki, and Dalcroze–on children's early PA. In short, Orff combines singing, movement, and percussion instruments; Kodály involves solfeggio and hand signs; Suzuki emphasises learning by ear and parent participation; Dalcroze focuses on eurhythmics and improvisation. Eccles et al. (2021b) only included studies involving typically developing children aged 5-8 years, and using one or more of the four aforementioned methods, which resulted in very few studies (only five) that met the requirements. Based on the results of included studies, music education positively impacts PA in typically developing children aged 5 to 8. Compared to Gordon et al.'s (2015) quantitative analysis, Eccles et al.'s (2021b) narrative synthesis lacks the statistical precision, making it more difficult to generalise the findings. This is not entirely their fault, as the insufficient number of included studies reflects an overall lack of evidence on this topic. Additionally, they aimed to compare the effects of specific types of music interventions, but the lack of standardized measurement tools in the original studies affected the reliability of their conclusions to some extent. Despite these limitations, they keenly observed all five studies were conducted in high- to middle-income countries and highlighted a lack of resources in less developed countries for using music education as a means of early PA intervention.

Some other studies are worth of discussion as supplements to the aforementioned meta-analysis or systematic reviews. Kempert et al. (2016) focused on children with inherently weak PA. Using a quasi-experimental design, they investigated the effects of combining two years of musical training with traditional phonological training on these children's PA. Children with weak PA were divided into two experimental groups: Group 1 (G1) received a combination of music and phonological training, while Group 2 (G2) received only phonological training. Group 3 (the control group) consisted of normal children who continued with their regular curriculum. There were four training time points as the researcher has drawn in Figure 2.3. During the period from T1 to T2, G2 also served as a control group, as these children did not receive any specific training before T3. For children with weaker PA, phonological training proved to be a stronger predictor of improvement, while the additional musical training did not achieve a significant effect. This result is constrained by certain limitations. For instance, although the overall sample size was sufficient, the sample size for children with weak PA was inadequate, which may have led to the difficulty in reliably detecting significant effects of musical training within this population.

Figure 2.3 Kempert et al.'s (2016) Research Design

In contrast, four more recent studies focusing on typically developing children support the positive impact of musical training on children's L1 PA (Table 2.2). Overall, they adopt the pretest-intervention-posttest design but differ in the specific implementation details.

Table 2.2 Overview of Longitudinal Research on the Impact of Music Instruction on L1 PA

Study, Year	Language	Participants (mean age/ age range in years, sample size)	Intervention	Operationalisation of PA
Vidal et al., 2020	European Portuguese	3-4 (n=44)	30 sessions (one 45-minute session per week) • music group (n=23): sing; play instruments, etc. • visual arts group (n=21): discuss famous artists and their techniques; paint, model, and draw using various materials, etc.	Conf-IRA–Phonological Awareness Screening and Assessment Instrument (Castro et al., 2018)
Bolduc et al., 2021	French	5-6 (n=160)	19 weeks • music group (n=50): rhythm-based musical training • motor group (n=52): guided play-based physical activities • control group (n=58): other creative activities	Phonological processing subtest of the Developmental NEuroPSYchological Assessment (Korkman et al., 2007)
Eccles et al., 2021a	English	6.3 (n=42, all male)	38 weeks • low-exposure music group (n=21): school-based music instruction • high-exposure music group (n=21): school-based and private music instruction (e.g., individual music instrument lessons)	• phonological processing: rhyming, segmentation, isolation, deletion, phoneme substitution and blending • phoneme-grapheme correspondence • phonemic decoding
Skubic et al., 2021	Slovakia	4-7 (n=70)	one hour per week for six months • music group (n=35): music instruction (e.g. listening songs, rhythmic education) • control group (n=35)	PA test (Ažman, 2011)

Vidal et al. (2020) found that the music group exhibited significantly greater progress. Their selection of the PA test is commendable since it is exceptionally comprehensive, encompassing 18 tasks that cover all syllabic structures present in European Portuguese. This extensive scope ensures a thorough assessment of PA across multiple dimensions. Besides, the ongoing validation and standardisation of this test further enhances its reliability, underscoring Vidal et al.'s (2020) commitment to using the best possible tool for their sample. However, there are two limitations, which could affect the robustness and generalisability of the study. First, they did not control for confounding variables, particularly by not measuring the motivation of two groups of children. If there were significant differences in motivation or engagement between the two groups, the superior performance of the music group over the visual arts group might not be due to the music curriculum itself but rather to differences in participation or motivation levels. Second, they did not include a control group continuing with traditional curriculum. The absence of a baseline control makes it difficult to ascertain the unique impact of music or visual arts education compared to regular education. In other words, it is unclear whether any observed effects are specifically attributable to music or arts education or to any form of extracurricular activities.

In Skubic et al.'s (2021) study, the post-test results revealed a significant advantage in PA for the music group. However, they do not specify what training the control group received in the intervention period. This lack of information about the control group's activities makes it difficult to rule out other factors that might influence the results. In spite of this, Skubic et al. (2021) incorporated covariate analysis into their study. This methodological rigor is beneficial as it accounts for initial disparities in PA levels between two groups and differences due to age. By including these two covariates, the analysis more accurately isolates the effect of the music intervention on PA, thus offering a more reliable interpretation that the music introduction significantly influenced PA, even when controlling for age and initial PA.

Unlike the other three studies comparing the music group to other groups, Eccles et al. (2021a) examined the differences in PA improvement among children receiving varying levels of music instruction. Although the differences between the high-exposure and low-exposure groups were not statistically significant, within-group comparisons revealed that the high-exposure group showed greater PA improvements at the end of the intervention compared to the low-exposure group. The sample size may be too small to detect significant differences between groups. Furthermore, their study was conducted less than four months, which may not be sufficient to capture long-term effects of music instruction on PA. Other external factors, such as home environment and additional extracurricular activities, were not controlled for and could have influenced PA. These limitations suggest that further research with more rigorous controls, longer intervention periods, and larger sample size, is needed to improve the robustness of the findings.

2.4.3. Music and L2 Reading

2.4.3.1. Longitudinal research: Music-based pedagogy improves L2 reading comprehension

Music-based language pedagogy gained popularity in ESL classrooms and has also been well studied in experimental settings. A large body of research has focused on the benefits of music-based pedagogy for oral production (e.g. Ashtiani & Zafarghandi, 2015; Heikkola & Alisaari, 2019; Jekiel & Malarski, 2021), and listening skills (e.g. Talamini et al., 2018; Tasnim, 2022). At present, there are only two studies on teachers using music to enhance L2 reading (Augustine, 2015; Fonseca-Mora et al., 2015).

These two studies were conducted in different contexts, but both adopted the pretest-intervention-posttest quasi-experimental design. In Augustine's (2015) study, 40 five-year-old preschoolers with Malay as L1 were divided into an experimental group receiving music-based instruction and a control group continuing with regular curriculum. He measured L2 reading through print knowledge, definitional vocabulary, and phonological awareness. In contrast, Fonseca-Mora et al. (2015) set up three groups among 63 Spanish native speakers aged seven to eight: one using music materials (e.g. songs) for PA enhancement, one using non-music materials (e.g. posters, audio books), and a control group receiving traditional instruction. Furthermore, Fonseca-Mora et al. (2015) operationalised reading into letter name knowledge, initial sound identification, oral reading fluency, and verbal WM. In general, despite the varying degrees of improvement in these subcomponents, both studies acknowledged that the groups receiving music-related training demonstrated better overall reading performance in the post-test compared to control groups.

Notably, Fonseca-Mora et al. (2015) found no significant differences between the music-phonological group and the non-music-phonological group. Thus, they cautiously suggested that the additional support music could provide might be limited. However, their concern may be somewhat unfounded, as the pre-training musicality test showed that children in the non-music-phonological group performed significantly better than those in the music-phonological and control groups. According to their background information survey, the children in the three groups came from different regions of Spain. Fonseca-Mora et al. (2015) estimate that the superior performance of the non-music-phonological group in the musicality test might be related to their backgrounds; many of these children were from Spanish Gypsy families, who may have more frequent exposure to music, potentially leading to higher musical abilities compared to children from other regions of Spain. In this case, it is, in itself, a very promising result that children with average musical ability, after receiving musical training, can achieve reading performance surpassing the control group, and on par with children who initially demonstrated high musical aptitude.

2.4.3.2. Longitudinal research: Musical training enhances L2 phonological awareness

Research exploring how musical training affects PA in L2 learners is rare, with only two studies identified (Herrera et al., 2011; Patscheke et al., 2016). Herrera et al. (2011) assessed the impact of a two-year musical training programme on PA in preschool children who spoke either Spanish or Tamazigh. They found that children who received music-enhanced phonological training or phonological training alone performed significantly better in Spanish PA tasks than the control group, regardless of whether they were Spanish native speakers or L2 learners. Particularly, phonological training that included music activities was more effective in enhancing PA. However, Herrera et al. (2011) incorporated isolated words into songs rather than adopting a full musical training programme. Thus, the extent to which music itself affects L2 PA remains unclear.

In response, Patscheke et al. (2016) randomly assigned preschool children from immigrant families to musical training, traditional phonological training, and a motor programme group. Sessions were held three times per week for 20 minutes each over 14 weeks. They found that both the musical training and phonological groups showed significant L2 PA improvements, while the motor programme group did not. This suggests that in addition to traditional phonological programmes, musical training may serve as an additional option to enhance L2 PA in immigrant children.

Patscheke et al. (2016) carefully controlled several important variables during the pretest phase, including age, gender, intelligence level, socioeconomic status, language background, and music experience. This control ensured that the baseline levels were similar across groups before the intervention, thereby enhancing the reliability of results. The only minor drawback is the lack of control over external influences during the training phase, such as parents potentially practicing L2 PA with their children at home. A suggestion for future research is to conduct follow-up questionnaires with parents during the training phase to better control for these factors.

2.4.3.3. Cross-sectional research: Positive relationship between musicality and L2 reading precursors

Likewise, the previous discussions on the relationship between musicality and both verbal WM and PA have been primarily within the context of L1 reading. Only two cross-sectional behavioural studies investigated their relationship in the L2 field (Foncubierta et al., 2020; Fonseca-Mora et al., 2021) and only one neuroimaging study observed the automatic brain responses to English reading in L2 learners with extensive musical experience compared to those without (Wang et al., 2020). Foncubierta et al. (2020) ultimately developed a statistical-causal model to determine L2 silent reading fluency based on L1 and FL segmentation, PA, auditory WM, and musical aptitude (Figure 2.4). Building upon a partial replication of Foncubierta et al.'s (2020) research, this current study clarifies some ambiguous definitions and addresses some of their methodological limitations.

Figure 2.4 Foncubierta et al.'s (2020) Structural Equation Model (p. 7)

To begin with, Foncubierta et al. (2020) focused on reading fluency, which they argue acts as a prerequisite for reading comprehension, and were operationalised into PA, auditory WM, and L1 and L2 segmentation. However, this current research centres on reading comprehension rather than fluency. This shift of focus is due to the understanding that reading involves transforming the meaning on the page into meaning in the mind, as highlighted by Kintsch and

Kintsch (2005). The present study holds that when looking at how music relates to reading, it is more reasonable to pay attention to how readers understand and interpret text rather than just how smoothly they read it.

The second modification is that, in the original study, PA was measured by the Perin's Spoonerism task where participants were required to listen to 18 pairs of first and last names of famous Spanish individuals, then swap the initial consonants of the first name with those of the last name, and orally produce a modified name. They evaluated participants' music perception, but measured PA using oral production, which is not modally equivalent. Instead, this study adopted a perceptive PA test, which is more accurate and closely aligned.

Finally, Foncubierta et al.'s (2020) study included segmentation tests for Spanish and Italian unspaced texts. However, these tests were excluded from this study because Chinese characters are independent, making segmentation irrelevant. Moreover, English segmentation depends more on morphological awareness and automatic recognition, with less influence from acoustic factors (Zhang, 2021).

2.5 Research Gap and Rationale

The positive relationship between verbal WM, PA, and musicality has been established in the literature on L1 reading. However, research examining the interplay between musicality and L2 reading remains notably scarce. While some longitudinal studies have encouragingly demonstrated that music-based pedagogies or additional musical training can bolster students' L2 PA and reading proficiency, only Foncubierta et al. (2020) established a model to explore the interaction between L2 learners' musical ability and their L2 silent reading fluency. Nevertheless, their study has several issues necessitating improvement, and there is a need to extend their efforts beyond the Spanish context into a brand-new research environment. Therefore, this current study incorporates statistical modelling to investigate the relationship among music perception, L2 PA, auditory WM, and L2 silent reading comprehension, representing the first cross-sectional behavioural research in an Asian context.

3. Methodology

This study is underpinned by the positivist research philosophy, which argues that knowledge created should be visible and measurable (Bell et al., 2022). Adopting a quantitative approach, this study aims to establish a model for the acoustic dimension of L2 silent reading, explore the interplay of the cognitive and linguistic factors, and further examine possible mediating and moderating mechanisms revolving around music perception. Data was collected from the background information questionnaire, L2 silent reading comprehension test, music perception test, phonological awareness test, and auditory working memory test.

3.1 Research Questions

In order to explore this topic, the following research questions were devised:

1) What are the correlations among L2 silent reading comprehension (RC), music perception (MP), phonological awareness (PA), and auditory working memory (AWM)?
2) What are the relationships between participants' musical training backgrounds (no training, basic training, advanced training) and their performance on MP, RC, PA, and AWM?
3) What are the structural relationships among RC, MP, PA, and AWM?

3.2 Research Design

Based on Foncubierta et al.'s (2020) research design described in the section 2.4.3.3, the current study made three main modifications: shifting the focus to reading comprehension, using a perceptive phonological awareness test, and excluding segmentation tasks due to the nature of Chinese characters. The following diagram illustrates the potential relationships between variables and the tests used to measure them in this current study (Figure 3.1).

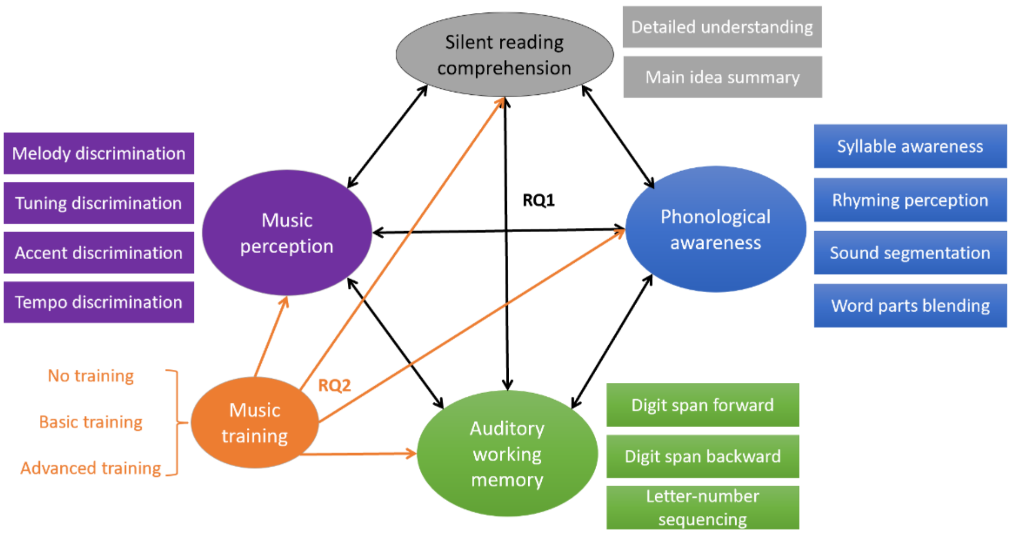


Figure 3.1 Diagram of the Interaction Among Variables

3.3 Participants

The determination of the sample size was based on Foncubierta et al.'s (2020) study that recruited 117 participants. 147 third-year students majoring in English teaching at Nantong University participated in this study. They are Mandarin native speakers with English as L2. Despite initially having 147 students participate in the study, six withdrew for personal reasons, and their data were completely removed. Additionally, due to technical issues, scores for multiple tests from two participants were not recorded and thus excluded from the dataset. The final effective sample size was 139.

Nantong University is in a well-developed coastal city. The English department of this university is highly regarded and ranked among the top 100 in China. In addition, approximately 93% of English teaching majors at this university pass the Test for English Majors Band Four (TEM-4) annually. TEM-4 is a nationwide examination in China aimed at assessing the English proficiency of undergraduate English majors. Despite certain limitations, such as uneven distribution of scores across different language skills, the TEM-4 has no superior alternatives and demonstrates an English proficiency level approximately equivalent to CEFR B1+ (Liu, 2012).

3.4 Instruments

The instruments used are the background information questionnaire and four tests. The reading comprehension and music perception tests were directly sourced from well-established instruments. A pilot study was conducted for the reading comprehension test because it had previously been used to evaluate the English reading abilities of non-native speakers with study abroad requirements and not for research purposes. Thus, through piloting, the reliability and effectiveness of the test can be validated in this study. The phonological awareness and auditory working memory tests were fine tuned and refined based on existing assessments.

3.4.1 Background Information Questionnaire

Participants first completed a background information questionnaire (see Appendix A). Some questionnaire items drew on participant information reported by Foncubierta et al. (2020), including age, gender, and participation in music-related activities. This study additionally required participants to self-assess their English proficiency, report their TEM-4 passing status, and detail their experiences with other languages to fully grasp their L2 proficiency and multilingual capabilities. Importantly, collecting background information facilitates the interpretation of potential outliers in subsequent data analysis. The participant profiles are presented in Table 3.1.

Table 3.1 Participant Profiles (Total sample: N = 139)

		N	Total=139 (100%)
Gender	Male	26	19%
	Female	113	81%
Age	20	1	1%
	21	81	58%
	22	48	34%
	23	8	6%
	26	1	1%
Self-reported English proficiency	Beginner	12	9%
	Intermediate	113	81%
	Advanced	14	10%
	Native-like	0	0%
Passed TEM-4	Yes	128	92%
	No	11	8%
Over six months of learning/ speaking other languages than Mandarin and English	None	12	8%
	Cantonese	1	1%
	Japanese	58	42%
	French	68	49%
Past involvement in formal music-related activities (before university enrolment)	Never	53	38%
	Only compulsory music lessons at school	52	37%
	Compulsory music lessons & over-one-year participation in after-class activities	34	25%
Current weekly time commitment for formal music-related activities (since university enrolment until now)	None	74	53%
	Less than 1h	42	30%
	2-5h	17	12%
	5-10h	5	4%
	More than 10h	1	1%
Weekly time spent on listening to music	Less than 1h	27	20%
	2-5h	70	50%
	5-10h	24	17%
	More than 10h	18	13%

3.4.2 Pilot Study for Reading Comprehension Test

The research was piloted with 20 Chinese third-year students majoring in Business English at Nantong University. Both pilot and formal participants are from the English department but enrolled in different programmes. They have similar age, language backgrounds, English proficiency, and exposure to English through regular lessons.

As the music perception, phonological awareness, and auditory working memory tests have been well-established and widely used in previous research, there is confidence in their reliability and validity for the current study's purposes. Hence, the resources and time for the pilot study were directed towards the reading comprehension test. It includes two subtests with questions drawn from the Cambridge English C1 Advanced (CAE) Reading tests (Cambridge University Press, n.d.).

Participants were required to read a magazine article about an African film festival and four news reports about an abandoned baby. They completed 10 multiple-choice questions, each worth one point. Six questions assessed their understanding of details in the magazine article, and four questions examined their ability to compare the main ideas and attitudes presented in four reports.

The piloting aims to ensure whether the difficulty of the reading comprehension test is suitable for third-year English majors and whether the questions and options are unambiguous and comprehensible. Participants in the pilot study were asked to complete the test while keeping track of the actual completion time. Following the test, they were invited to provide feedback regarding anything they could think of, such as the time length and difficulty of the test, topics of the passages, and the comprehensibility of the questions.

All pilot participants reported that the test questions and options were clear and easy to understand, and that the topics of the articles and news reports were engaging. Their data were visually examined using the histogram and Q-Q plot (see Appendix B.1 and B.2). Additionally, z-scores for skewness and kurtosis were calculated to minimise subjective bias in visual inspection (see Appendix B.3). The accuracy rates following a normal distribution is a positive indicator, suggesting that the difficulty of questions is likely balanced, with none being overly easy or difficult.

To further ensure the test difficulty level was appropriate for the formal study, pilot participants evaluated the test using a five-point Likert scale (1=very easy/5=very difficult). The five-point scale with fewer response options, compared to 7-point and 9-point scales, makes it easier for participants to understand and use. The five options provide a moderate balance, avoiding oversimplification while preventing respondents from feeling overwhelmed by too many choices. Pilot participants' assessments of test difficulty are presented in Appendix B.4. The mean score was 3.75 and the standard deviation was .716. This indicates that most participants perceive the test as moderately challenging, with consistent views. This stability in perceptions of test difficulty enhances the reliability of subsequent data analysis.

Despite the lack of variety in question types, the CAE reading comprehension test was ultimately deemed suitable for Chinese ESL students due to three careful considerations. First, Cambridge language assessments are relevant to higher education settings and extensively practiced by similar student groups across the world, thereby having high credibility and reliability. Furthermore, the normal distribution of accuracy and participants' evaluations of difficulty in the pilot study indicate that the test presents an ideal moderate challenge. This facilitates differentiation among participants across various reading comprehension levels. If the test is too easy, scores may be similar across participants, while an overly difficult test may result in participants being unable to complete all items, both impairing effective differentiation. Finally, some of the target participants are known to be familiar with some assessments widely used in existing literature such as IELTS and TOEFL reading tests. Therefore, selecting the CAE test, which students have not practised before, may reduce the possible bias on the results. This test with answers is presented in Appendix C.

The actual time it took participants to complete the reading test, as reported by themselves, is shown in Appendix B.5. Seven participants finished within 20 minutes, seven between 20 and 25 minutes, and six between 25 and 30 minutes. The time they suggested the test would take is presented in Appendix B.6. Out of 20 participants, only 5 recommended that the test time should be less than 20 minutes. The remaining 15 suggested either 25 or 30 minutes, with seven recommending 25 minutes and eight recommending 30 minutes. Therefore, to accommodate most pilot study participants' suggestions and align with their actual completion time, the final test duration set for the formal study is 25 minutes. This decision helps ensure that the test strikes a balance, neither being overly stressful nor relaxed for participants.

3.4.3 Music Perception Test

Participants completed a 15-minute Mini-PROMS, the shorter version of the Profile of Music Perception Skills (PROMS) test (Zentner & Strauß, 2017). This is currently among the most reliable and widely-used music perception tests by psychologists, musicians, and linguists. This quick version was chosen to reduce the task load for participants and save time. Each question in the Mini-PROMS consists of an audio clip that participants listen to twice and a comparison played once. Participants were asked to answer whether the reference and the comparison were the same or different in terms of melody, tuning, accent, and tempo.

The melody subtest assesses the ability to distinguish individual musical notes within a sequence where the intervals between notes remain consistent. The tuning subtest evaluates the ability to identify different notes within a chord where notes are played together. The accent subtest evaluates the ability to recognise rhythmic patterns within a sequence of 5 to 12 beats which may be strong or weak. In the tempo subtest, each beat or group of beats varies in duration, as well as the intervals between beats. The key points of these four subtests are presented in Table 3.2.

Table 3.2 Key Points of Four Mini-PROMS Subtests

	Reference	**Comparison**
Melody	do→so→*mi*→la→re	do→so→*fa*→la→re
Tuning	do + re + mi + la + *so*	do + re + mi + la + *fa*
Accent	**da** dada **da** dadada **da** **da**	**da** **da** dadada **da** dada **da**
Tempo	da~[1s]dadada~[1s]dada~[1s]da~	da~~[0.5s]dadada~~[0.5s]dada~~[0.5s]da~~

The differences might be large or small. As shown in Figure 3.2, participants were given five options (i.e. “definitely same”, “definitely different”, “probably same”, “probably different”, and “I don’t know”). They were asked to choose “definitely same/different” if they had clearly heard no difference/a difference and choose “probably same/different” if they had a certain intuition. If uncertain about whether detecting a difference or not, they could choose “I don’t know”. One point is for each correct answer with full certainty (“definitely same/different”) and 0.5 points for each correct answer with low certainty (“probably same/different”). The melody and accent subtests each consist of 10 items, while the tuning and tempo subtests each contain 8 items, totalling 36 items. After completing the test, participants received their total scores and scores for each of the four subtests.

The figure originally presented here cannot be made freely available via ORA because of copyright.
The figure was sourced at Zentner, M., & Strauß, H. (2017). The Profile of Music Perception Skills (PROMS): Manual. University of Innsbruck.

Figure 3.2 An Example of Music Perception Test

3.4.4 Phonological Awareness Test

The phonological awareness test includes four subtests (see Appendix D). Every subtest includes eight questions, each with one point for the correct answer and zero points for incorrect responses. The stimuli provided to participants were auditory and not visually represented on their screen in written form. The subtests were selected from overlapping sections of the Quick Phonological Awareness Screening Test (QPAS) (HPEDSB, 2013) and Assessment of Phonological Awareness (Phillips & Kelly, 2018). The overlapping sections are typically considered the most reliable and stable parts of the two tests, representing their shared focus. Receiving repeated confirmation across different tests, the reliability, stability, validity, and comprehensiveness of the assessment are enhanced.

There are two justifications for why other subtests from these two readily available tests were not included in this study. First, they involve participants' oral output and require one-on-one recording by researchers, which imposes significant demands on space and resources. With participants exceeding 130, online testing is more feasible for this study. Second, This study focuses on the relationship between music and reading in the acoustic dimension. Since music emphasises auditory perception, PA assessments of oral production may not be suitable for research purposes.

Detailed information on the four subtests is provided in Appendix D, with key points summarised below. First, during the rhyming perception test, participants heard words in pairs and triads, after which they determined whether these stimuli rhyme. Second, the syllable awareness test required them to answer the number of syllables in the words they heard. Third, during the sound segmentation test, participants answered how many sounds they heard in the words. The final is word parts blending test, where participants listened to parts of high-frequency words, pieced them together, and typed them down. Participants listened to each stimulus twice, once pronounced by a male and once by a female native speaker, with either British or American accents.

The two speakers were recruited from Italki (i.e. an online foreign language teaching and learning platform) to record audio of the words for the phonological awareness test. The female speaker comes from London with a British accent, and the other male speaker from Mississippi has a standard American accent. To ensure the intelligibility of speakers' pronunciation, their audio was assessed using a five-point Likert scale (1=very clear/5=very unclear) by the participants in the pilot study. The primary reason for not using 7-point or 9-point scales is that a simpler scale suffices for determining the intelligibility of a speaker's pronunciation without requiring nuanced distinctions in listeners' perception. The descriptive statistics for participants' ratings are presented in Appendix B.7, with a mean score of 1.3 for the British speaker and 1.4 for the American speaker. This suggests that the pronunciation of the two native speakers is highly intelligible.

The British speaker holds a professional Teaching English as a Foreign Language (TEFL) certificate and has six years of experience in teaching English to non-native speakers. The American speaker holds a bachelor's degree in international and public communication and has over four years of English teaching experience. The correct answers to the phonological awareness test were jointly determined by these two native speakers and the researcher, reducing individual biases and increasing the objectivity and accuracy of the judgments.

3.4.5 Auditory Working Memory Test

Following the recommendations of Wechsler et al. (2008), this study measured auditory working memory using digit span forward and backward, as well as letter-number sequencing tests. There was no time limit on all tests, and each test began with an explanation for clarity.

The digit span tests used in this study were adapted from the test developed by Dean (2020). The original test has been widely used on Gorilla. However, considering that all participants are tested in the same classroom, the original version is less practical as it requires recording with a microphone, which might lead to noise and interference. Therefore, participants were asked to type their answers into the box to better suit the research environment. Participants heard a sequence of digits and then were immediately prompted to type out the sequence, either as it was heard (forward), or in reverse order (backward). The digit span tests consist of lists ranging from two to ten numbers for the forward test and nine numbers for the backward test, with two trials at each length, totalling 34 trials.

In the letter-number sequencing test, participants heard a sequence of numbers and letters. They needed to recall numbers in ascending and letters in alphabetical order, with the numbers preceding letters. To minimise the stress on participants and maintain a similar format to the digit span tests, the total length of letters and numbers ranges from four to nine, with two trials at each list length and 12 in total. For example, when participants hear "T - 1 - A - 7 - 2 - Z - 8 - Y - 4", the correct response is "1, 2, 4, 7, 8, A, T, Y, Z".

3.5 Data Collection

Website links of the background information questionnaire, phonological awareness test, and auditory working memory test were compiled and generated on the experiment builder Gorilla. Music perception test was completed on the website developed by the Music Emotion Assessment and Personality Lab, University of Innsbruck. All the study links were tested several times to ensure that they work on different devices. Following this, the links were shared with the researcher's contacts outside the university to ensure that the wording of questions was clear and understandable.

This research took place in a classroom at Nantong University. All participants were asked to arrive at the classroom with their pens, electronic devices (such as laptops, phones, tablets) and headphones during the study. The tasks (1.25 hours in total) were conducted in two sessions with a 20-minute break in between.

During the first session, participants filled out an online questionnaire about their background information (5 mins). Then, they completed a phonological awareness test (15 mins) and an auditory working memory test (15 mins) on their devices with headphones. Participants were forbidden to use paper and pen during these two tests and only permitted to use electronic devices for typing. During audio presentation, the typing box remains hidden. It only becomes visible after the audio finishes, allowing participants to respond. These measures aim to prevent participants from relying on visual or memory aids that could influence test results and to ensure the authentic assessment of their auditory working memory and phonological awareness.

During the next session, the participants took a reading comprehension task (25 mins) and a music perception test (15 mins). The reading comprehension test was distributed in printed format to participants for ease of paging through and making annotations. Participants were required to handwrite their answers using a pen during the reading test but were not permitted to use electronic devices or dictionaries. Throughout the data collection stage, the researcher was present to offer appropriate task instructions and supervise participants' completion along with a teacher.

3.6 Data Processing

Participants' responses to the paper-based reading comprehension tests were collected on-site upon completion. Subsequently, the researcher manually graded these tests based on the correct answers provided by Cambridge University Press, and then recorded the results in an Excel spreadsheet. Other online tests will be automatically graded by the system. Participants' responses to each question and their total scores can be directly downloaded in CSV format from accounts provided by Gorilla and University of Innsbruck. Ultimately, the data was imported into SPSS (version 29) for further analysis.

3.7 Data Analysis

Descriptive statistics were generated to test normality and identify outliers, which informed the selection of appropriate statistical methods. A summary of the data analysis methods is provided in Table 3.3. Regarding the first research question, if all variables conform to normal distribution and exhibit linear relationships, a Pearson correlation matrix can be conducted. For the second question, considering the sole independent variable as the categorical variable representing levels of musical training, and multiple dependent variables, a MANOVA is employed. Ultimately, MP, RC, PA, and AWM are seen as latent variables, and a statistical model is developed for their interaction. Multiple linear regression is not adopted because it fails to account for measurement error, whereas structural equation modelling (SEM) incorporates it into the analysis.

SEM can be categorised into two types: covariance-based SEM, the most widely used and considered mainstream, and variance-based SEM, such as the partial least squares (PLS) employed in this study. Hair et al. (2011) analysed the distinctions between two types of SEM, identifying three main differences: first, covariance-based SEM aims to optimise parameter estimation, whereas variance-based SEM focuses on maximising predictive power. Second, covariance-based SEM is limited to using reflective indicators, while variance-based SEM can incorporate both reflective and formative indicators. Finally, covariance-based SEM necessitates a strong theoretical foundation to support confirmatory research, whereas variance-based SEM is mostly used in exploratory research, without requiring an extensive theoretical basis.

The intersection of reading and music skills is a novel research area with an underdeveloped theoretical foundation. Thus, this exploratory study aims to develop rather than test theory, and to maximise the predictive power of the model rather than assess its overall fit. Additionally, the study includes both reflective and formative indicators. Based on these considerations, PLS is deemed the optimal choice.

Table 3.3 Summary of Data Analysis Methods

RQs	Analyses	Statistical Methods	Software
What are the correlations between RC, MP, PA, and AWM?	Correlation matrix	• Pearson correlation (normality, linear relationships) • Spearman's rank correlation (non-normality, non-linear relationships)	SPSS
What is the relationship between participants' musical training backgrounds and their performance on MP, RC, PA, and AWM?	Multivariate analysis of variance (MANOVA)	Multivariate test Tests of between-subjects effects Post-hoc test	SPSS
What are the structural relationships among RC, MP, PA, and AWM?	Partial least squares regression (PLS)	Structural equation modelling	SmartPLS

3.8 Ethical Considerations

Participants were recruited to participate in this study after the researcher received ethical approval (see Appendix E). At the beginning, the information sheet and consent form (see Appendix F) were sent through teachers at Nantong University to each potential participant. The information sheet outlined the rationale for this research and detailed the procedures participants would experience if they decided to take part. The consent form explained how participants could withdraw themselves from the study without any negative consequences, how their deidentified and confidential data would be securely stored, and how they could contact the researcher if they had questions or concerns. The student willing to participate in this study signed the consent form and personally returned it to the researcher on site.

Each participant was given a unique participant code without using their true names to take part in this research. All data collected was non-identifiable through fully integrated security management platforms of Gorilla and https://musemap.org/resources/proms, preventing any disclosure of participants' private information. Data, including consent records, was stored in two formats: paper records, secured in a safe and then securely destroyed upon digitisation to ensure participants' privacy and data integrity, and electronic records, stored in the researcher's university Nexus 365 OneDrive for Business. After the researcher leaves the University at the end of the project, the supervisor will be responsible for the data stored in Nexus 365 OneDrive.

4. Findings

This chapter begins by testing the normality of the collected data and identifying any outliers. Following this, all three research questions are addressed in detail, providing robust statistical analysis and visual presentations.

4.1. Normality and Outliers

The participant profiles were detailed in the methodology chapter (see Table 3.1). In each of the 139 valid participant entries, there were no missing data for either the sub-scores or total scores of each test item. Checks for normality and outliers were carried out in SPSS for all four variables: music perception (MP), L2 silent reading comprehension (RC), phonological awareness (PA), and auditory working memory (AWM). First, the data was examined visually with Q-Q plots and histograms (Appendix G). These visualisations were scrutinised to assess the distribution shape and the positioning of the best-fit curve. In general, all variables almost matched the anticipated normal distribution criteria. Further, the normality of the item scores was rigorously assessed using z-scores for skewness and kurtosis. Each variable was found to follow a normal distribution, as z-scores within the range of ±1.96 typically indicate normality (Filed, 2013). The descriptive statistics of variables are presented in Table 4.1.

Table 4.1 Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Music perception	139	8.0	32.0	19.036	4.6441	.091	.206	-.065	.408
Reading comprehension	139	1.00	9.00	4.6259	1.69515	.360	.206	.007	.408
Phonological awareness	139	7	29	18.77	4.661	-.185	.206	-.699	.408
Auditory working memory	139	14	42	29.01	5.650	-.134	.206	-.242	.408
Valid N (listwise)	139								

Based on boxplots, MP had one outlier, RC had seven, whereas PA and AWM had none (Appendix H). There is no universal standard for handling outliers. Researchers' needs determine appropriate methods based on the specific research context, data characteristics, and analysis objectives. This current study decided to retain all outliers without additional processing for the following two reasons. First, the tests were conducted under strict supervision by the researcher and a teacher, ensuring a well-regulated environment where participants completed tests as instructed within the allotted time. This minimises the likelihood of participants not taking the tests seriously. Therefore, the researcher believes that these genuine outliers reflecting individual differences in ability, rather than errors or invalid responses. In other words, these outliers may capture real-world extreme cases, and retaining them helps to reflect the underlying reality of the data, avoiding excessive data manipulation and distortion. Second, with outliers included, the descriptive statistics reported earlier indicate that the data still follow a normal distribution and do not exhibit significant skewness. Thus, the researcher believes that these outliers do not substantially impact the results.

4.2. Correlations Among Reading Comprehension (RC), Phonological Awareness (PA), Auditory Working Memory (AWM), and Music Perception (MP)

In this section, RC, PA, AWM, and MP were treated as a whole by aggregating the scores of their respective subtests into a total sum. The first question, examining the correlations among these four variables, was addressed by creating a correlation matrix. The four variables were known continuous and normally distributed. Therefore, the final crucial prerequisite before constructing a correlation matrix is to verify whether there exists a linear relationship among them. The verification typically involves the use of scatter plots for visual inspection. By examining the scatter plots for all pairs of variables, some patterns such as a straight line or another recognisable trends can be observed. Figure 4.1 is the scatterplot matrix where each cell in the grid represents the scatterplot between two variables. The fit line revealed that there was a positive linear trend between each pair of the four variables.

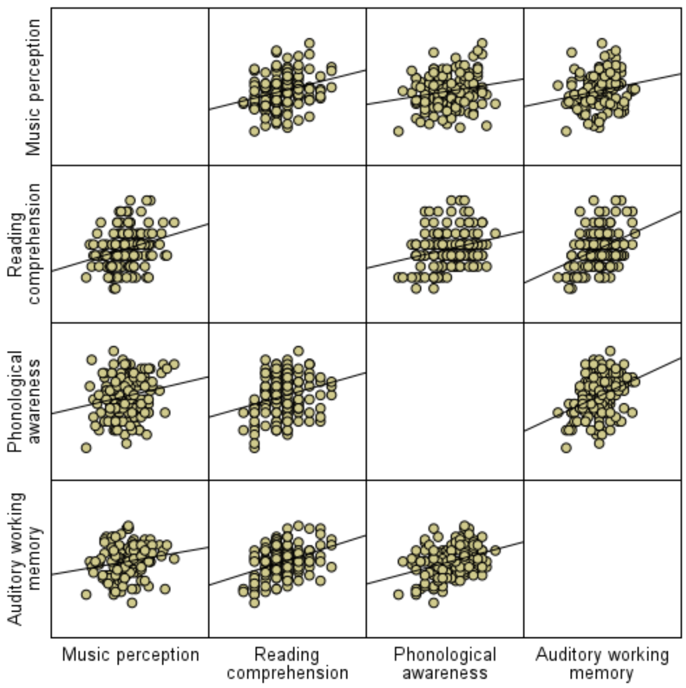


Figure 4.1 Scatterplot Matrix of MP, RC, PA, and AWM

The linear relationships observed in the scatter plots allow for the calculation of Pearson correlation coefficients to quantify the strength and direction of these relationships. Table 4.2 displays Pearson correlation coefficients between MP, PA, RC, and AWM. The matrix revealed a modest correlation between MP and RC ($r = .275$, $p < .01$), approaching the bound of a moderate correlation, and indicating that higher levels of MP are associated with better RC. MP also exhibited a significant positive relationship with PA ($r= .195$, $p < .05$) and AWM ($r = .190$, $p < .05$), though the strength of these correlations was relatively modest. RC was significantly correlated with PA ($r = .258$, $p < .01$) and AWM ($r= .381$, $p < .01$) in this current study, which supports the view in the literature that they are precursors to RC.

Table 4.2 Peason Correlation Matrix

	Music perception	Reading comprehension	Phonological awareness	Auditory working memory
Music perception	1	.275**	.195*	.190*
Reading comprehension		1	.258**	.381**
Phonological awareness			1	.354**
Auditory working memory				1

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

As mentioned in the previous chapter, the MP, RC, PA, and AWM tests are composed of multiple subtests, which were also analysed in a 13*13 correlation matrix (Appendix I). Four MP subtests showed significant correlations among themselves, and similarly, the AWM subtests were correlated with each other. However, among the PA subtests, word part blending was significantly correlated with rhyming perception, syllable awareness, and sound segmentation. Syllable awareness was also associated with rhyming perception and sound segmentation, while rhyming perception was not related to sound segmentation. By analysing the correlations between subtests, a better understanding of the different aspects measured can be achieved, and potential redundancies can be identified. This analysis improves the overall validity of each test and guides subsequent improvements in the test content and structure.

4.3. Relationship Between Musical training and RC, PA, AWM, and MP

Background information collected from participants shows that there are differences in the music activities and training they have participated in, both in the past and currently. Therefore, the second research question examines whether there are significant differences in the performance of participants with no training, basic training, and advanced musical training in MP, RC, PA, and AWM tests.

According to their responses to the item regarding past involvement in formal music-related activities, participants are categorised into three groups: those never participated in any formal music activities, those only took compulsory music lessons at primary/secondary school, and those participated in compulsory school music lessons along with over one year of engagement in extracurricular music-related activities. The item concerning current weekly time commitment for formal music-related activities offers five response options: none, less than 1 hour, 2-5 hours, 5-10 hours, and more than 10 hours. Combining both past and current participation in music activities, participants' music backgrounds can be divided into the following three groups: no training, basic training, and advanced training (Table 4.3).

Table 4.3 Classification of Musical training Backgrounds

	Past involvement in formal music-related activities	**Current weekly time commitment for formal music-related activities**
No training (n=48)	never	none / less than 1h
Basic training (n=42)	only compulsory music lessons at primary/secondary school	none / less than 1h
Advanced training (n=49)	never	2-5h / 5-10h / more than 10h
	only compulsory music lessons at primary/secondary school	2-5h / 5-10h / more than 10h
	compulsory music lessons at school & over-one-year participation in extracurricular music-related activities	none / less than 1h / 2-5h / 5-10 h/ more than 10h

MANOVA is adopted since the only independent variable is categorical and four dependent variables are involved. MANOVA allows for simultaneous handling of multiple dependent variables, assessing their joint effects while also enabling separate analysis of each variable. Compared to conducting ANOVA for several times, MANOVA enhances statistical efficiency by considering the correlations among variables, thereby reducing the risk of Type I error.

Before using a MANOVA to test group differences in performance on MP, RC, PA, and AWM tests, two assumptions need to be check: multivariate normality and homogeneity of covariance matrices (Field, 2013). Detailed descriptive statistics of the four variables across three groups, along with Box's test of equality of covariance matrices, confirm that both assumptions have been satisfied (Appendix J). After the MANOVA was performed, multivariate tests that report four test statistics - Pillai's Trace, Wilks's Lambda, Hotelling's Trace, and Roy's Largest Root- were generated (Table 4.4). Olson (1974) points out that for moderate sample sizes (such as n=139 in this current study), the four statistics exhibit negligible differences in statistical power, thus it is possible to choose any one of them. The table indicates that the p-values for three of the four statistics are .028, and one is .013, all of which are below the .05 threshold. Therefore, the differences in MP, RC, PA, and AWM attributable to varying musical training backgrounds are statistically significant overall.

Table 4.4. Multivariate Tests

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.976	1378.074[b]	4.000	133.000	<.001	.976
	Wilks' Lambda	.024	1378.074[b]	4.000	133.000	<.001	.976
	Hotelling's Trace	41.446	1378.074[b]	4.000	133.000	<.001	.976
	Roy's Largest Root	41.446	1378.074[b]	4.000	133.000	<.001	.976
Training	Pillai's Trace	.123	2.200	8.000	268.000	.028	.062
	Wilks' Lambda	.880	2.200[b]	8.000	266.000	.028	.062
	Hotelling's Trace	.133	2.199	8.000	264.000	.028	.062
	Roy's Largest Root	.099	3.310[c]	4.000	134.000	.013	.090

a. Design: Intercept + Training

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

However, a significant MANOVA usually reflects a significant difference for one rather than all the dependent variables. Thus, the next step involves a detailed investigation of individual differences for each of the four dependent variables. This requires examining the tests of between-subjects effects (Table 4.5), which essentially represent the outcomes of conducting separate ANOVAs for each dependent variable. Among the three groups with varying degrees of musical training, significant differences were observed only in PA ($F = 4.353$, $p = .015 < .05$, partial $\eta^2 = .06$). A partial eta-squared value of 0.06 is considered as a small effect size.

Table 4.5 Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Music perception	106.834[a]	2	53.417	2.532	.083	.036
	Reading comprehension	5.418[b]	2	2.709	.942	.392	.014
	Phonological awareness	180.393[c]	2	90.197	4.353	.015	.060
	Auditory working memory	76.078[d]	2	38.039	1.195	.306	.017
Intercept	Music perception	50011.375	1	50011.375	2370.301	<.001	.946
	Reading comprehension	2966.534	1	2966.534	1031.498	<.001	.884
	Phonological awareness	48926.893	1	48926.893	2361.069	<.001	.946
	Auditory working memory	116804.473	1	116804.473	3669.605	<.001	.964
Training	Music perception	106.834	2	53.417	2.532	.083	.036
	Reading comprehension	5.418	2	2.709	.942	.392	.014
	Phonological awareness	180.393	2	90.197	4.353	.015	.060
	Auditory working memory	76.078	2	38.039	1.195	.306	.017
Error	Music perception	2869.486	136	21.099			
	Reading comprehension	391.129	136	2.876			
	Phonological awareness	2818.240	136	20.722			
	Auditory working memory	4328.915	136	31.830			
Total	Music perception	53345.500	139				
	Reading comprehension	3371.000	139				
	Phonological awareness	51969.000	139				
	Auditory working memory	121362.000	139				
Corrected Total	Music perception	2976.320	138				
	Reading comprehension	396.547	138				
	Phonological awareness	2998.633	138				
	Auditory working memory	4404.993	138				

a. R Squared = .036 (Adjusted R Squared = .022)

b. R Squared = .014 (Adjusted R Squared = -.001)

c. R Squared = .060 (Adjusted R Squared = .046)

d. R Squared = .017 (Adjusted R Squared = .003)

Despite this, MANOVA tends to safeguard the dependent variable where genuine group differences exist. Therefore, applying a Bonferroni correction to these ANOVAs to protect against Type I error is necessary (Bray & Maxwell, 1985; Harris, 2001). Results of Bonferroni post-hoc test are shown in Table 4.6. The untrained group scored significantly lower in PA compared to the basic training group ($p = .029 < .05$) and advanced training group ($p = .047 < .05$). However, no significant differences were found between the basic and advanced training groups in PA. Regarding MP, RC, and AWM, no significant differences were observed among the three groups. A follow-up MANOVA was conducted on the PA subtests to pinpoint the specific sources of significant between-group differences (Appendix K). The analysis revealed significant differences were in the sound segmentation and word part blending subtests.

Table 4.6 Multiple Comparisons

Bonferroni

Dependent Variable	(I) Training	(J) Training	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Music perception	advanced	basic	1.527	.9659	.349	-.814	3.868
		untrained	2.017	.9328	.097	-.244	4.278
	basic	advanced	-1.527	.9659	.349	-3.868	.814
		untrained	.490	.9705	1.000	-1.863	2.842
	untrained	advanced	-2.017	.9328	.097	-4.278	.244
		basic	-.490	.9705	1.000	-2.842	1.863
Reading comprehension	advanced	basic	.014	.3566	1.000	-.851	.878
		untrained	.421	.3444	.670	-.413	1.256
	basic	advanced	-.014	.3566	1.000	-.878	.851
		untrained	.408	.3583	.771	-.461	1.276
	untrained	advanced	-.421	.3444	.670	-1.256	.413
		basic	-.408	.3583	.771	-1.276	.461
Phonological awareness	advanced	basic	-.27	.957	1.000	-2.59	2.05
		untrained	2.26*	.924	.047	.02	4.50
	basic	advanced	.27	.957	1.000	-2.05	2.59
		untrained	2.53*	.962	.029	.20	4.86
	untrained	advanced	-2.26*	.924	.047	-4.50	-.02
		basic	-2.53*	.962	.029	-4.86	-.20
Auditory working memory	advanced	basic	-1.82	1.186	.380	-4.70	1.05
		untrained	-.67	1.146	1.000	-3.45	2.11
	basic	advanced	1.82	1.186	.380	-1.05	4.70
		untrained	1.15	1.192	1.000	-1.74	4.04
	untrained	advanced	.67	1.146	1.000	-2.11	3.45
		basic	-1.15	1.192	1.000	-4.04	1.74

Based on observed means.

The error term is Mean Square (Error) = 31.830.

*. The mean difference is significant at the .05 level.

4.4. Structural Relationships Among RC, PA, AWM, and MP

This section builds a partial least squares regression (PLS) model to explore the structural relationships among RC, PA, AWM, and MP. From a rigorous perspective, these four variables are latent constructs that are not directly observable or measurable. Their manifestation requires indirect reflection through the measurement of observed indicators that represent the actual data collected. This rationale underpins the use of PLS instead of multiple linear regression to understand these relationships.

4.4.1. Reflective and Formative indicators

As mentioned in the methodology chapter, indicators have two types. Reflective indicators are considered interchangeable, meaning that removing an indicator does not alter the fundamental nature of the latent construct. In contrast, for formative indicators, removing an indicator means omitting a part of the construct itself (Diamantopoulos & Winklhofer, 2001). In this study, therefore, the formative indicators for MP include melody discrimination, tuning discrimination, accent discrimination, and tempo discrimination. For RC, the formative indicators are detailed understanding and main idea summary. PA consist of four formative indicators, syllable awareness, rhyming perception, sound segmentation, and word parts blending. Finally, AWM is represented by three reflective indicators, digit span forward, digit span backward, and letter-number sequencing.

MP and PA subtests are indispensable; for example, tuning discrimination and word part blending are integral to the overall ability. Similarly, in reading comprehension, grasping details and main ideas are equally important, making them formative indicators. In contrast, AWM can be evaluated using various tests, such as the digit span or letter-number sequencing test, which are not interdependent. Omitting one of these methods does not affect the overall evaluation. The arrows point from formative indicators to latent variables, but from latent variables to reflective indicators. The preliminary model is shown in Figure 4.2.

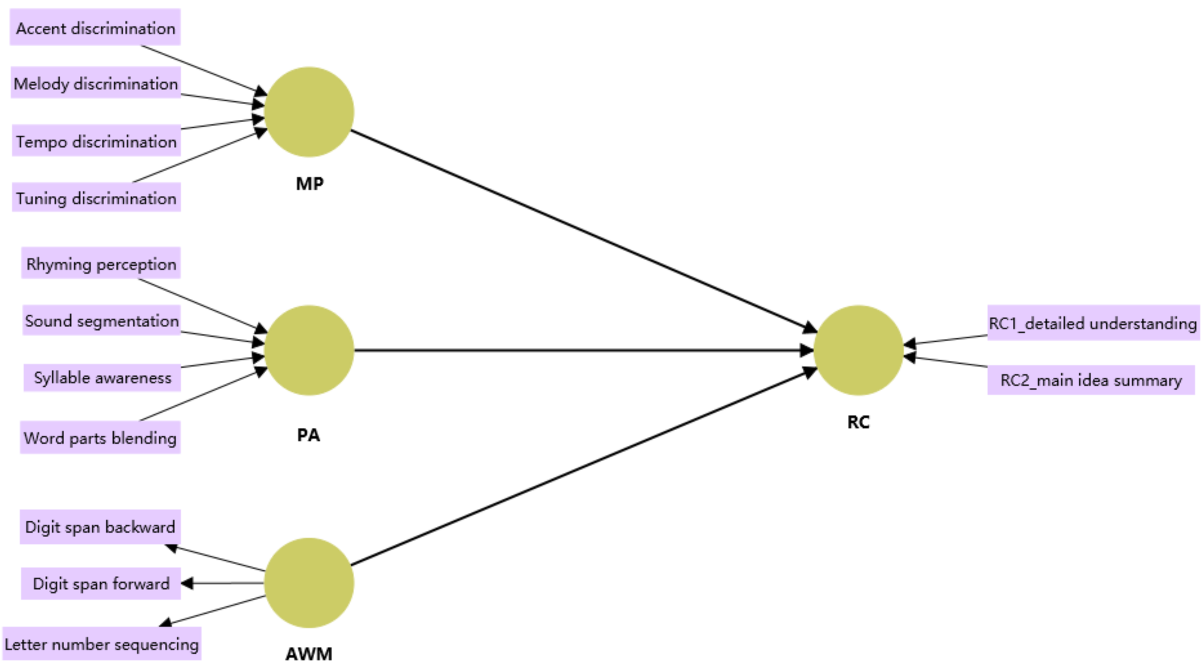


Figure 4.2 Preliminary PLS Model

4.4.2. Validity Check of PLS Measurement Model

After establishing the preliminary model, it is crucial to evaluate the validity of each latent variable. For the variable composed of reflective indicators, the following five PLS algorithm requirements need to be satisfied: 1) the values of factor loadings should be higher than .700; 2) the value of Cronbach's alpha should be higher than .700; 3) the value of composite reliability should be higher than .700; 4) the value of Average Variance Extracted (AVE) should be higher than .500; 5) the values of factor loadings should be higher than those of cross loadings (Chin, 1998; Fornell & Larcker, 1981; Nunnally & Bernstein, 1994).

On the other hand, for the latent variable consisting of formative indicators, there are two different requirements: 1) indicators should be significantly correlated with the construct and factor weights should exceed .100; 2) indicators and the construct should have variance inflation factor (VIF) values less than 3.3 (Chin, 1998; Diamantopoulos & Siguaw, 2006).

Analysis of the preliminary model revealed that two MP indicators (i.e. accent and melody discrimination) and two PA indicators (i.e. rhyming perception and syllable awareness) were not significantly correlated with MP and PA respectively, and their factor weights were less than .100 (accent: .056, melody: .000, rhyming: .068, syllable: .054). This violates the first requirement for formative indicators. Although these four indicators are theoretically relevant, they might be not difficult enough to discriminate the participants in this study. For these third-year Chinese English majors, the differences they show in discriminating accent and melody as well as perceiving rhyming and the number of syllables in a word are not significant, and thus the four indicators contribute little to MP and PA in this context. In other words, if applied to a different context and population, their importance and weights may vary.

After careful consideration, these four indicators were removed from the preliminary model. Figure 4.3 shows the factor weights and p-values of formative indicators in the revised model. The indicators are significantly correlated with the latent variables, with all factor weights exceeding .100. Table 4.7 presents VIF values of RC, PA, and MP indicators, all less than 3.3. Thus, the three latent variables, composed of six formative indicators in total, meet all the PLS algorithm requirements.

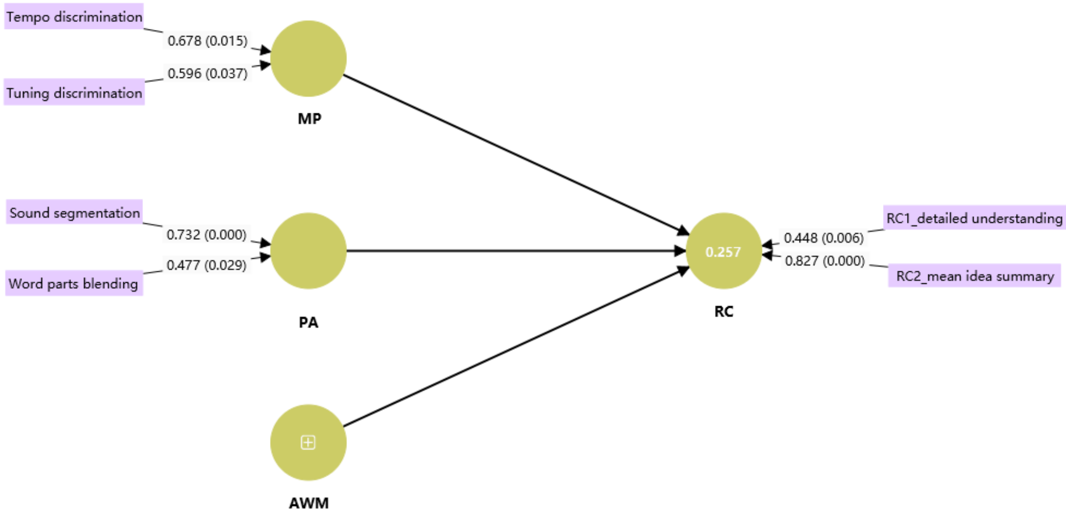


Figure 4.3 Factor Weights and P-values of Formative Indicators

Table 4.7 VIF Values of Formative Indicators

	VIF
RC1_detailed understanding	1.025
RC2_mean idea summary	1.025
Sound segmentation	1.128
Word parts blending	1.128
Tempo discrimination	1.056
Tuning discrimination	1.056

Furthermore, Cronbach's alpha, composite reliability and AVE values of AWM are reported in Table 4.8. Both Cronbach's alpha value and composite reliability value were .723 (> .700). AVE value was .643 (> .500). Table 4.9 presents factor loadings and cross loadings of the three AWM reflective indicators. The values of factor loadings were all higher than .700. Factor loading value of digit span backward was .783, higher than cross loadings with MP (. 115), PA (.323), and RC (.323). The same goes for digit span forward and letter-number sequencing. Therefore, AWM, which consists of three reflective indicators, satisfy all the PLS algorithm requirements. Taken together, the measurement model has been thoroughly reviewed and in accordance with expectations.

Table 4.8 Cronbach's Alpha, Composite Reliability and AVE Values of AWM

	Cronbach's alpha	Composite reliability (rho_a)	Average variance extracted (AVE)
AWM	.723	.723	.643

Table 4.9 Factor Loadings and Cross Loadings of AWM Reflective Indicators

	AWM	MP	PA	RC
Digit span backward	.783	.115	.323	.323
Digit span forward	.827	.112	.324	.325
Letter-number sequencing	.794	.204	.189	.358

4.4.3. Evaluation of PLS Inner Model

After checking the validity of the measurement model, the next step is to examine the VIF values between independent and dependent variables in terms of the inner model, which should be less than 3.3 to rigorously indicate the absence of collinearity (Kock & Lynn, 2012). VIF values between independent and dependent variables is presented in Table 4.10. None of the values exceeded 3.3, suggesting there was no collinearity present.

Table 4.10 VIF Values Between Independent and Dependent Variables

	AWM	MP	PA	RC
AWM				1.157
MP	1.000			1.043
PA				1.139

Then the mediation and moderation analysis revolving around MP were conducted. Mediation analysis explored whether the link between MP and RC (partially) pass through PA/AWM whereas moderation analysis investigated whether MP influence the link between PA/AWM and RC. The final determination of a PLS model usually relies on a series of indices, including the significance of path coefficients, effect size (f^2), R^2, and predictive relevance (Q^2). Notably, f^2 quantifies the unique variance in the dependent variable explained by a predictor in regression analysis, distinguishing its specific contribution from that of other predictors in the model (Cohen, 1988).

The final model is shown in Figure 4.4 where a significant mediation effect, rather than a moderation effect, was observed. MP affected RC both directly and indirectly via AWM. PA also had a direct effect on RC. Table 4.11 presents path coefficients, corresponding p-values and 95% confidence intervals. All p-values below .05 indicated that these paths were statistically significant. For example, there was a 95% probability that the path coefficient from AWM to RC lay within the interval [.163, .449].

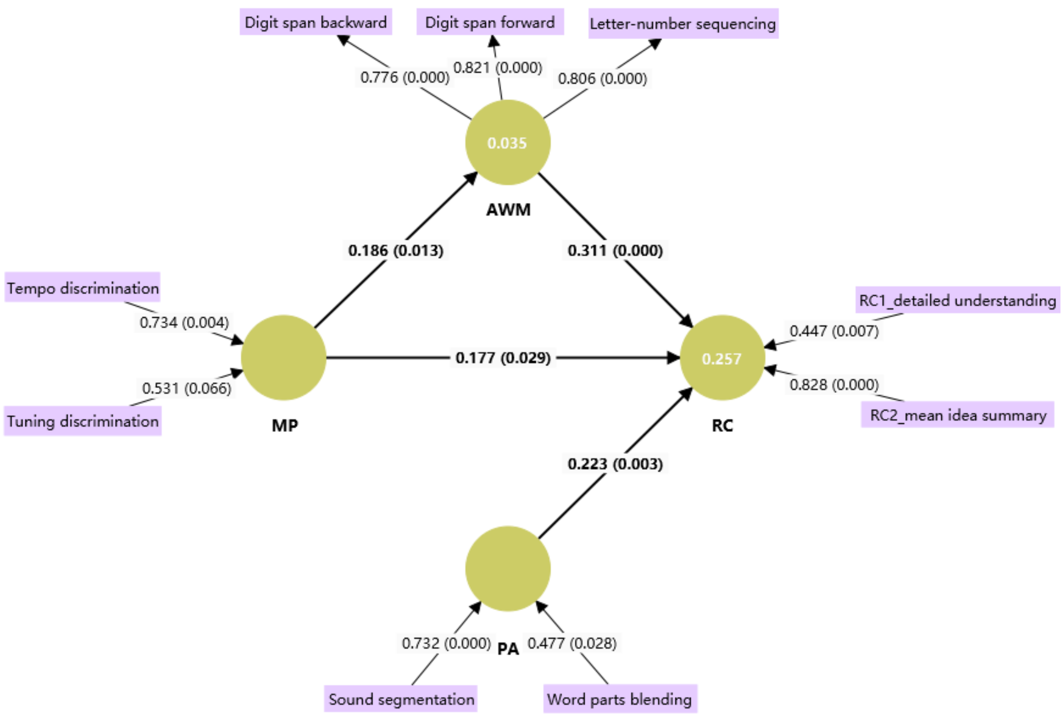


Figure 4.4 Final PLS Model of Acoustic Dimension in ESL Silent Reading Comprehension[2]

Table 4.11 Statistics of Mediation Analysis

	Original sample (O)	Standard deviation (STDEV)	T statistics (|O/STDEV|)	P values	2.50%	97.50%
Direct effects						
MP -> AWM	.186	.075	2.482	.013	.163	.449
AWM -> RC	.311	.074	4.194	.000	.046	.341
MP -> RC	.177	.081	2.18	.029	.018	.333
PA -> RC	.223	.075	2.969	.003	.087	.384
(Total) indirect effects						
MP -> AWM -> RC	.058	.028	2.083	.037	.012	.121
Total effects						
MP -> AWM	.186	.075	2.482	.013	.084	.389
AWM -> RC	.311	.074	4.194	.000	.087	.384
MP -> RC	.235	.078	3.001	.003		
PA -> RC	.223	.075	2.969	.003		

R^2, f^2, and Q^2 are presented in Table 4.12. According to Chine (1998), R^2 values of 0.67, 0.33, and 0.19 are considered substantial, moderate, and small. Therefore, the R^2 value of .257

[2] After AWM became a mediator, the previously significant MP-tuning discrimination relationship (p = .037) became non-significant (p = .066). This is likely due to partial correlation effects but not a concern, as it does not impact the model's accuracy (Baron & Kenny, 1986; Hair et al., 2011).

indicated this auditory-based model could explain 25.7% of the variance in English silent reading comprehension for Chinese students, which, while considered small to moderate, still demonstrated a promising explanatory power. With respect to f^2, AWM made the largest contribution to RC among the three variables. The Q^2 values of .015 and .114, both above zero, suggest that this model is well-structured and possesses predictive relevance.

Table 4.12 f^2, R^2, and Q^2

	f^2	R^2	Q^2
MP -> AWM	.036	.035	.015
AWM -> RC	.112	.257	.114
MP -> RC	.04		
PA -> RC	.059		

This section examined the interplay among MP, RC, PA, and AWM. The key finding was that MP influenced RC both directly and indirectly through AWM, while PA directly affected RC.

4.4.4. Multi-group Analysis

Building upon findings in the section 4.3, significant differences in PA were observed among three groups with different musical training backgrounds. Therefore, it is necessary to investigate whether these differences extend to the path coefficients across groups in the final PLS model, particularly focusing on the path from PA to RC.

Due to inherent limitations of SmartPLS software, analyses are restricted to comparisons between only two groups within multi-group settings. Consequently, researchers typically contrast the two extreme groups to examine differences in paths. In this study, analysis was conducted between untrained group and advanced-training group. It is worth pointing out that prior to conducting multi-group analysis, measurement invariance should be established to ensure that any observed model differences are not attributed to varying interpretations of the measurement model across groups. Table 4.13 presents results of three-step measurement invariance testing. Despite two values (-.456 and .656) slightly deviating from the 95% confidence interval in the third step, the partial to nearly full measurement invariance was established. Some scholars argue that the third step imposes a more stringent requirement and is not absolutely necessary. Achieving the expectations in the first and second steps is sufficient to indicate measurement invariance (e.g., Saad et al., 2020).

Table 4.13 Results of Three-Step Measurement Invariance Testing Using Permutation

	Step 1	Step 2			Step 3(a)				Step3(b)				
	configural invariance	C=1	5% quantile of Cu	compositional invariance	Original difference	2.50%	97.50%	mean equal	Original difference	2.50%	97.50%	variance equal	measure invariance
AWM	yes	.983	.954	yes	.111	-.407	.386	yes	.656	-.604	.55	no	partial
MP	yes	.744	.21	yes	-.26	-.446	.414	yes	-.515	-.526	.538	yes	full
PA	yes	.951	.546	yes	-.456	-.418	.407	no	-.071	-.382	.418	yes	partial
RC	yes	.988	.8	yes	-.299	-.41	.384	yes	.146	-.616	.554	yes	full

Establishing measurement invariance enables the multi-group analysis, with the results detailed in Table 4.14. None of them reached statistical significance ($p > .05$). Therefore, the path coefficients were not specific to any group, suggesting that different musical training backgrounds did not significantly affect the relationship between these variables.

Table 4.14 Path Coefficients in Multi-group Analysis

	Difference (Group_1 untrained - Group_3 advanced training)	1-tailed (Group_1 untrained vs Group_3 advanced training) p value	2-tailed (Group_1 untrained vs Group_3 advanced training) p value
AWM -> RC	.194	.189	.378
MP -> AWM	-.110	.726	.548
MP -> RC	-.155	.752	.496
PA -> RC	-.079	.660	.680

4.4.5. Importance-Performance Map Analysis

This part pertains to an importance-performance analysis of the final model, aimed at aiding stakeholders in decision-making regarding resource allocation. Specifically, importance typically occupies the horizontal axis of the chart, representing the degree of impact each variable has on the dependent variable. Performance, situated on the vertical axis, represents the operational feasibility of implementing independent variables under current conditions.

According to Figure 4.5, the performance of all three variables is ideal, hovering around 50%. Concerning their significance to RC, AWM emerges as the most pivotal, followed by PA, with MP being the least influential. Nevertheless, the importance scores of both MP and PA exceeded .17, suggesting their acceptable relevance, though relatively lower than AWM. Consequently, for enhancing L2 reading comprehension, the recommendation is to prioritise AWM while still giving appropriate attention to PA and MP.

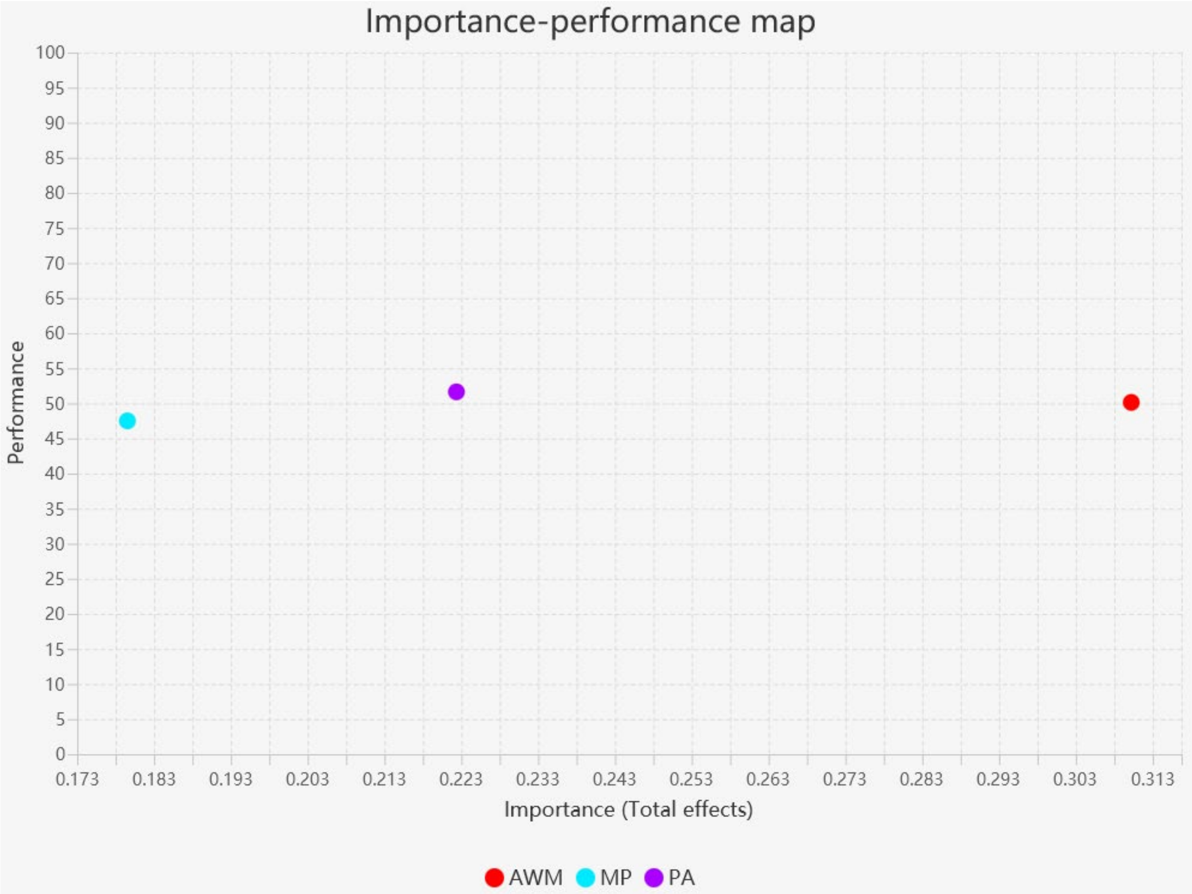


Figure 4.5 Importance-Performance Map

4.4.6. Robustness Checks

Conducting robustness checks for the PLS model aims to ensure its stability and reliability under varying conditions, thereby enhancing the model's explanatory power and the credibility of findings. These checks typically include heterogeneity analysis, nonlinearity assessment, and endogeneity testing (Sarstedt et al., 2020).

4.4.6.1. Heterogeneity analysis

Heterogeneity arises from participants' different perceptions of latent variables, which may result in variations in the outcomes. Observable heterogeneity, including gender, education level, age, and nationality, was absent in this study. However, unobservable heterogeneity implies that differences between groups are not due to previously identified characteristics but rather cause significant variations in path coefficients. This issue is the primary focus of the subsequent examination using FIMIX-PLS. Although a total of 11 fit indices are provided, Sarstedt et al. (2020) note that only AIC3, CAIC, and BIC exhibit stable performance, while the others are less reliable. Consequently, these three indices were retained in Table 4.15, where segment one showed the lowest values across all three indices. This suggested that dividing the data into two, three, or four groups was entirely unnecessary, and that there were no unobservable heterogeneity issues.

Table 4.15 Model Selection Criteria

	Segment 1	Segment 2	Segment 3	Segment 4
AIC3 (modified AIC with Factor 3)	760.806	774.51	780.295	780.836
BIC (Bayesian information criterion)	772.413	799.659	818.984	833.067
CAIC (consistent AIC)	778.413	812.659	838.984	860.067
EN (normed entropy statistic)	0	0.316	0.638	0.749

4.4.6.2. Nonlinearity assessment

Scatter plots were previously used to visually assess the linearity of relationships between latent variables. However, a more rigorous examination is needed because nonlinear relationships between dependent and independent variables could potentially influence the statistical significance of estimates (Hair et al., 2011). In PLS modelling, it is customary to examine the most common form of nonlinearity, that is quadratic relationships. Quadratic testing results showed that all P values were higher than .05 (Table 4.16). Non-significant results suggested that the relationships between constructs were linear.

Table 4.16 Quadratic Testing

	Original sample (O)	Sample mean (M)	Standard deviation (STDEV)	T statistics (|O/STDEV|)	P values
QE (MP) -> AWM	.059	.052	.068	.867	.386
QE (MP) -> RC	-.009	-.003	.066	.135	.893
QE (PA) -> RC	-.001	.000	.072	.015	.988
QE (AWM) -> RC	-.038	-.034	.052	.726	.468

4.4.6.3. Endogeneity testing

Endogeneity arises when independent variables correlate with error terms in a statistical model. This problem stem from several factors, such as omitted variable bias, reverse causality, and measurement error (Antonakis et al., 2014; Hill et al., 2021). In PLS modelling, Gaussian copula approach is usually used to identify endogeneity (Hair et al., 2021). The Gaussian copula test results, displayed in Table 4.17, indicated that all p-values exceeded .05. Non-significant results suggested that the PLS model established in this study was not affected by endogeneity issues.

Table 4.17 Gaussian Copula Test

	Original sample (O)	Sample mean (M)	Standard deviation (STDEV)	T statistics (|O/STDEV|)	P values
GC (MP) -> AWM	.234	.534	.868	.270	.787
GC (MP) -> RC	-.229	-.222	.889	1.383	.167
GC (PA) -> RC	.230	.162	.491	.468	.640
GC (AWM) -> RC	-.275	-.129	.941	.292	.770

4.5. Summary of Findings

Taken together, main findings are as follows. First, music perception, phonological awareness, auditory working memory, and silent reading comprehension were significantly correlated with each other. Second, participants with basic and advanced musical training perform better in overall phonological awareness than those without training. Specifically, such between-group differences were found in sound segmentation and word parts blending subtests. Finally, the auditory-based PLS model identified the mediating role of AWM in the MP-RC relationship and demonstrated an encouraging degree of explanatory power for variations in RC.

5. Discussion

In this chapter, main findings are discussed with reference to previous empirical evidence and key theories. Additionally, pedagogical and therapeutic implications are elaborated upon. Finally, limitations of this study and areas for improvement and expansion in future research are identified.

5.1 Auditory Working Memory: A Mediator Between Music Perception and L2 Silent Reading Comprehension

In the literature review, the researcher posed the following question: given that both music and reading rely on auditory working memory (AWM), does music perception (MP) influence silent reading comprehension (RC) through verbal AWM? The PLS model now indicates that in typically developing adult L2 English readers, MP influences RC both directly and indirectly via verbal AWM, thereby offering an answer to this question. Further, this observed mediation effect invites a closer look at the underlying mechanisms positively linking musical AWM (elicited by musical stimuli) and verbal AWM (elicited by linguistic stimuli).

There are two possible explanations for their positive correlations. One explanation is that the brain regions activated by both AWM systems are the same, sharing the same neural resources. This means that the same region can efficiently process both musical and linguistic information. The other explanation posits that both systems process musical and linguistic stimuli through similar mechanisms, including encoding, maintaining, manipulating, and suppressing irrelevant information. However, the neurocognitive evidence supporting these explanations remains debated. Some scholars argue that while the two systems are managed by different brain hemispheres, they share similar rehearsals for stimulus processing (e.g., Imm et al., 2008; Jeong & Ryu, 2016). Conversely, others believe that although the same brain region manage both systems, their approaches to processing stimuli differ (e.g. Koelsch et al., 2009). Regardless of which hypothesis may prove more accurate in future research, combined with the findings of this current study, it is reasonable to speculate that the AWM capacity enhanced by L2 learners through music can transfer to reading and other language skills. Thus, it may become clearer why verbal AWM may serve as a mediator between MP and RC.

5.2 Group Differences in MP, RC, PA, and AWM by Musical Training Backgrounds

In this study, significant differences were found only in L2 PA, but not MP, RC, and AWM, among participants who were divided into three groups with no musical training, basic training, and advanced training. L2 PA performance was significantly lower in the untrained group than in the basic and advanced training groups. This suggests a significant positive correlation between L2 PA and musical training; however, the direction of their relationship remains unclear. Fortunately, two longitudinal studies referenced in the literature review further found that musical training positively affects L2 PA (Herrera et al., 2011; Patscheke et al., 2016), thereby indicating the direction of their relationship is that musical training predicts improvements in L2 PA. In contrast, a surprising but interesting finding in this study is that no significant difference was observed in MP between the three groups. This implies that musical training background did not correlate with participants' MP in this study. However, much empirical evidence indicate more musical training can improve an individual's MP (e.g., Bigand & Poulin-Charronnat, 2006; Morrongiello & Roes, 1990). In response, three possible explanations are provided in terms of participants' professional background and learning environment, quality of musical training, and individual differences in musical aptitude.

First, since participants are undergraduate English majors whose daily studies focus more on language rather than music-related activities, they may lack sufficient opportunities to practice and reinforce MP skills. Therefore, the positive effects of musical training may be more easily observed in phonological rather than music perception. Second, despite reporting similar time spent on musical training, participants may have significant differences in the quality of their training. For instance, intermittent or unsystematic training pattern may not effectively enhance MP. Finally, variations in innate musical aptitude may obscure the effects of training. For example, some participants may naturally possess higher MP, making the effects of training less pronounced in these individuals.

From a fine-grained perspective, a follow-up MANOVA on four subtests of L2 PA revealed that the advantages gained by participants with basic or advanced musical training were evident in sound segmentation and word blending subtests. This suggests that participants' ability to blend word parts and segment sounds benefits more from musical training than their ability to recognise rhymes and perceive syllables. This finding challenges Patscheke et al.'s (2016) observation that musical training has a stronger impact on participants' ability to perceive syllables, detect rhymes, and blend word parts than on their ability to segment sounds/phonemes.

Two reasons potentially contributing to the discrepancies in results are the age of participants and the content of musical training. According to the participants' teacher, recognising rhymes and counting syllables in words are highly emphasised in their ESL classrooms. Therefore, adult English majors in this study may already have acquired proficiency in these PA skills due to their extensive formal L2 learning experience. As a result, these two subtests might have been easier for them, making the additional benefits from musical training relatively less pronounced. In contrast, Patscheke et al. (2016) focused on five-year-old pre-schoolers who received extensive exposure to nursery rhymes as part of their musical training. Nursery rhymes, characterised by simple lyrics and end rhymes, naturally enhance children's ability to recognise rhymes in L2 words (Herrera et al., 2011). Additionally, children in Patscheke et al.'s (2016) study had far less exposure to formal L2 education compared to the adult Chinese English majors in this current study. Therefore, children benefited more from musical training to improve their ability to count syllables in words.

An on-going debate in the literature is regarding whether varying levels of musical training significantly affect participants' performance on digit span forward and backward tests (Saarikivi et al., 2019; Nie et al., 2022). This study found no significant differences between the groups in the overall AWM test. It should be noted that Saarikivi et al. (2019), Nie et al. (2022), and this study each offers unique perspectives, with differences in results attributed to various factors, including research design, intervention methods, and criteria for classifying musical training (Table 5.1). Future research can leverage the strengths of these three designs by understanding participants' prior musical involvement, implementing systematic music interventions for selected individuals while establishing a control group, regularly measuring participants to assess effects over time, and employing robust modelling methods.

Table 5.1 Three Studies Exploring the Relationship Between Musical Training and Two Digit Span Tasks

	Research Design	Classification Criteria for Training	Intervention	Measuring Time Points	Findings
Saarikivi et al., 2019	cohort sequential design	having more than 2 years of instrumental training	no	three times within six years (once every two years)	• Digit span forward task: the music-trained group outperformed the untrained group • Digit span backward task: no significant between-group differences

Nie et al., 2022	longitudinal design	experimental vs control group: music intervention or not	one-year systematic music intervention	before and after intervention	• Digit span forward task: no significant between-group differences • Digit span backward task: the music-trained group outperformed the untrained group
Current study	cross-sectional design & PLS modelling	self-reported varying degrees of musical training	no	single time point	There were no significant differences between the three groups (no, basic, and advanced training) in the overall auditory working memory test.

Taken together, the lack of a significant correlation between musical training and both MP and RC is likely due to various complex factors. However, the predictive power of MP on RC reflects the similar processing mechanisms for musical and linguistic stimuli, or the shared neural resources between them, highlighting the mediating role of AWM.

5.3 Pedagogical Implications for L2 Reading Teaching

The PLS model demonstrates that participants with better MP tend to have superior AWM, which contributes to enhanced RC. Therefore, for typically developing L2 adult learners, it may be beneficial for teachers to adopt music-related reading instruction.

Currently, the only two studies focusing on music-based L2 reading instruction have been implemented with children (Augustine, 2015; Fonseca-Mora et al., 2015). In both studies, children were encouraged to sing the reading materials. Ultimately, these catchy songs improved their reading abilities through spontaneous rehearsals. In Augustine's (2015) research, children not only sang L2 songs but also incorporated additional physical movements or imitations. Corrigall and Trainor (2011) and Edwards et al. (2009) regarded music as an ideal way for children to explore texts and printing. When children listen and sing, they understand that characters printed, like musical notation, are meaningful. Additionally, as Pica (2010) emphasised, body language, through which verbal expressions of thoughts and feelings begin, is a powerful form of communication. From this perspective, L2 learners can benefit from physical movements that come with music. However, she also points out that explicit language learning can be enhanced through physical activities, but their impact is far greater on children than on adults.

These two studies aimed at children have obtained good results. It should be noted that adult L2 learners typically already possess mature learning methods, and their learning goals and schedules may differ significantly from those of children. Therefore, requiring them to spend extensive time on purely musical training may be impractical or not aligned with their learning needs. Instead, incorporating music elements appropriately into L2 reading instruction may boost learning outcomes while actively engaging students' interests.

RC is influenced by both L2 learners' existing linguistic knowledge and their use of reading strategies (Macaro et al., 2015). Therefore, teachers' reading instruction could closely focus on these two aspects. Essential components of linguistic knowledge include vocabulary and grammar. Much evidence suggests that the effectiveness of using music to teach vocabulary (e.g. Chen, 2020; Coyle & Gómez, 2014; Mannarelli & Serrano, 2024; Yousefi et al., 2014) and grammar (e.g. Alinte, 2013; Fagerland, 2006).

However, no research has yet explored the integration of reading-specific strategy instruction with music elements for adult L2 learners. Yapp et al. (2021) synthesised ten essential L2 reading strategies that teachers should prioritise after reviewing literature on reading strategy and consulting Harvey and Goudvis's (2007) reading strategy handbook. Therefore, drawing on Yapp et al.'s (2021) summary, the researcher identifies specific reading strategies that are well-suited for integration with music, and provides teachers and researchers with three original suggestions for future practice and research. First, the two reading strategies of skimming and scanning can be trained with background music. Specifically, using music with different rhythms could help learners adjust their reading pace to better apply both techniques. For example, playing upbeat music could encourage learners to skim through the text quickly to grasp the main ideas, while slower music may prompt a more focused scanning approach to find specific information. Second, lyrical content could potentially be used to activate background knowledge by associating certain melodies with specific themes. For example, playing a piece of music that is thematically related to the reading material may help learners recall relevant information and concepts. Finally, after reading or studying a text on a particular topic, learners might benefit from listing all in-text key words or phrases related to the theme. Using the newly acquired vocabulary, they could create lyrics that ensure connections between the words are established within the song. Then, by singing the lyrics, students might better grasp and remember the vocabulary associated with the reading topic. Through these methods, teachers may effectively combine reading strategies with music, enhancing L2 learners' reading skills while potentially increasing learning enjoyment.

5.4 Therapeutic Potential for L2 Reading Difficulties

Although generalising MP's prediction for RC to atypical development requires caution, the theoretical foundation established in this study provides a potential basis for hypothesising the potential effectiveness of music therapy for L2 reading disorders. The research aims not only to present statistical results but also to discuss practical applications, showcasing that musical training could be beneficial for a wider population.

Existing literature supports the therapeutic effects of music on people with L1 reading disabilities (e.g., Colwell & Murlless, 2002; Leloup et al., 2021). Notably, L1 and L2 reading disabilities are classified in the same way due to a well-established reading model. It is known as simple view of reading (SVR), proposed and developed by Gough and Tunmer (1986) and Hoover and Gough (1990). In essence, the SVR posits that L1 reading comprehension hinges on the ability to decode words and to understand the meaning conveyed. Afterwards, some scholars argue that the SVR also works well for L2 readers and can assist teachers in diagnosing L2 reading problems (Sparks, 2015; Verhoeven & Leeuwe, 2012). Therefore, they believe that reading difficulties manifest in two ways, regardless of whether it is L1 or L2. Specifically, one is characterised by difficulties in spelling and decoding printed texts, occurring despite normal intelligence and sufficient education, and without any obvious sensory or neurological impairments. The other involves children who can read aloud accurately and fluently but have trouble understanding the content.

Therefore, despite the current lack of research in this area, it is reasonable to speculate that musical training may also have analogous therapeutic effects on individuals with L2 reading disabilities. Forgeard et al. (2008) conducted a study investigating the relationship between MP and language-related skills. They found that in typically developing L1 readers, MP directly influences reading ability. In contrast, for readers with L1 reading difficulties, MP indirectly affects reading ability through PA. Different from these findings, this current study reveals that in typically developing L2 readers, MP not only directly but also partially impacts RC through AWM. Therefore, for readers with L2 reading difficulties, the influence of MP on RC may operate through PA, AWM, or both. This highlights the key role of MP in RC and implies that the way in which musical training alleviates L1 and L2 reading difficulties may be interconnected yet subtly different.

Colwell and Murlless (2002) compared the effects of music activities and regular reading programmes on the reading accuracy of pupils with reading disabilities and found that participants in both conditions improved their reading accuracy, but in the music context, students were more attentive and had fewer instances of breaking classroom rules and interrupting learning compared to regular reading activities. Interestingly, participants used sign language frequently in the music context. Schunk (1999) found that incorporating sign language in music activities further enhance L2 development. Thus, spontaneous sign language that comes with musical training might serve as a facilitator for students with L2 reading disabilities.

Register et al. (2007) compared the effectiveness of an intensive music course on the reading abilities of both dyslexic and typically developing children. They found that participants with disorders showed significantly more improvement on post-reading comprehension tests compared to those without. Admittedly, the typically developing group's higher pre-test scores may have caused a ceiling effect, limiting their improvement potential. However, a more likely reason is the smaller class size of the reading disabilities group, which facilitated greater engagement and concentration among students. In contrast, the typically developing group had larger class sizes but relatively lower attendance rates. This observation also provides valuable insights for teachers to treat reading L2 difficulties, that is, to engage and capture students' attention when designing and implementing musical training interventions. Register et al. (2007) provided a highly transparent experimental procedure, offering detailed descriptions of the music course design. These courses are easily replicable and extended by other L2 reading teachers.

In addition to simply introducing additional musical training, exercises that combine reading exercises with music can also work well (Bonacina et al, 2015; Leloup et al., 2021). Bonacina et al (2015) examined a rhythmic reading training scheme, in which reading exercises were placed against the rhythmic background. 14 middle school students with reading disorders participated in nine biweekly 30-minute individual training, and their post-intervention reading improvements were compared to those of the control group of 14 dyslexic students who did not receive the intervention. Their study found that this training significantly improved participants' reading speed of pseudowords and reading accuracy of high-frequency long words. Similarly, Leloup et al.'s (2021) training programme is called repeated reading with vocal music masking, wherein participants read texts repeatedly while background music featuring human voices is played. Their longitudinal research also achieved encouraging results that dyslexic children's reading gains achieved through this programme were well retained over time.

This programme of background music-assisted reading provides a valuable reference for interventions in L2 reading difficulties. However, the selection of background music and texts should account for cultural and linguistic differences to ensure that the content is relevant and understandable for L2 learners. Additionally, L2 dyslexic learners tend to experience greater cognitive load in processing L2 compared to L1. While background music with music may benefit some students, it could be distracting for others. Therefore, when designing similar training programmes and interventions, it is crucial to fully consider the specific challenges and needs of L2 learners, and to make appropriate adjustments and optimisations. Overall, all this needs to be tested empirically since this is a complex area, but that your research suggests it may be a fruitful area for exploration

5.5 Limitations and Suggestions

The researcher tried the best to adopt rigorous methods appropriately applied to research questions. However, the findings should still be interpreted with caution due to the following limitations, which future research can address for improvement.

The first limitation is that RC tests exclusively consist of multiple-choice questions, which may not adequately measure a student's ability to analyse, interpret, and synthesize information from the text. However, this decision primarily aims to prevent participants from experiencing excessive pressure, ensuring they can maintain focused during the test-taking process. Simultaneously, to provide as rich assessment as possible, the researcher carefully selected two different types of texts, each of which examines a different purpose of reading: magazine articles focus on detailed understanding, while news reports focus on main idea summary. This design helps to examine students' grasp of different genres and skills used in different reading tasks. Nevertheless, future research can include more diverse question types, such as open-ended or essay questions, which further evaluate students' critical thinking skills, textual inference abilities, and overall comprehension rather than just identifying correct answers among provided options.

Second, participants' musical training background is based on self-reported data and the questionnaire items regarding their involvement in musical activities are somewhat simplistic. This introduces potential biases, as individuals might overestimate or underestimate their involvement in music activities due to memory inaccuracies. Furthermore, the questionnaire mainly focuses on the time spent on music activities, which is a rough indicator and does not reflect the quality or specific content of these activities. Participants may have attended different types of music classes or engaged in various music activities, which can have different impacts on their language-related skills. For example, Christiner and Reiterer (2015) found that singing training and musical instrument training contribute differently to the comprehension and imitation of unfamiliar L2 accents, with vocalists significantly outperforming instrumentalists. Hence, this effect is also likely to manifest in L2 reading, but such differences cannot be captured in this study. Future research should consider conducting in-depth interviews in addition to questionnaires. This approach would allow for a more thorough exploration of participants' musical training experiences, such as specific courses, types and frequency of activities. Moreover, researchers can obtain richer and more accurate information about participants' music motivations, challenges, and achievements by asking follow-up questions and seeking clarifications.

Finally, the removal of two subtests each from MP and PA may reduce the model's explanatory power for RC. Although the MP and PA measurement tools selected are widely used and reliable, in practical applications, some subtests may not meet statistical significance standards. This does not imply fundamental flaws in the measurement tools themselves but rather suggests that their applicability and effectiveness may vary across specific research contexts. Therefore, removing these subtests aims to ensure the model accurately reflects the complexity of the research phenomenon, thereby enhancing the credibility of findings. Future research should consider expanding into other L2 environments and diverse populations to reevaluate and select appropriate research instruments.

6. Conclusion

The main goal of this quantitative study was to explore the structural relationships among music perception (MP), L2 silent reading comprehension (RC), phonological awareness (PA), and auditory working memory (AWM). Additionally, this study explored whether participants with different musical training backgrounds (no training, basic training, advanced training) significantly differ in their performance on MP, PA, AWM, and RC. Data was collected from supervised tasks completed by third-year Chinese undergraduate ESL students majoring in English teaching.

It was found that MP, RC, PA, and AWM were significantly correlated with each other. Among AWM, PA, and MP, AWM showed the highest correlation with RC. The partial least squares model further indicated that PA directly predicted RC, whereas MP predicted RC both directly and indirectly through AWM. In brief, the model suggests that PA, MP, and AWM are all important for RC, and MP works through a more complex pathway involving AWM. For participants with different musical training backgrounds, significant between-group differences were only observed in PA. Basic and advanced training groups significantly outperformed the untrained group in the sound segmentation and word parts blending subtests, while there was no significant difference in performance between these two training groups.

The present study adds to the growing research on the intersection between music and language. It narrows its focus on the acoustic dimension of L2 silent reading comprehension. Most importantly, this study develops Foncubierta et al.'s (2020) research by clarifying their confusion about some reading concepts (i.e. comprehension and fluency), adapting well-established research instruments, and further performing rigorous statistical testing for them. Furthermore, a sound sample size, a pilot study, and groundbreaking statistical modelling with partial least squares contribute to a better understanding of how music perception interacts with L2 silent reading comprehension and its two acoustic precursors in the Chinese ESL context.

It is also worth mentioning that this study provides evidence to support further investigation into the effectiveness of music-related reading interventions, thereby having significant potential pedagogical and therapeutic implications. Pedagogically, if teachers can integrate music elements into L2 reading instruction, this may help to improve adult learners' reading comprehension while stimulating their interest. Two interesting avenues for future research on the intersection of music and L2 reading are suggested. First, reading strategy instruction with music elements proposed in the discussion chapter are appropriate for longitudinal intervention studies, thereby assessing the potential enduring effects of music-assisted pedagogy in enhancing reading proficiency. Additionally, it is advisable to examine how individual differences influence the effectiveness of music-assisted reading instruction. For example, variations in learning styles and motivation may affect learners' acceptance and adaptation to these pedagogies, aiding in the development of personalised strategies for music-assisted L2 education. Therapeutically, existing evidence that music therapy is effective in L1 reading acquisition combined with the discovery of this study that music perception can predict L2 reading learning provides a theoretical basis for treating L2 reading disabilities with music.

In conclusion, this study supports the significant roles of PA, MP, and AWM in ESL silent reading comprehension. While limitations exist, the findings pave the way for future research to explore these relationships in broader contexts. One promising area is the exploration of these relationships across different languages. Investigating whether similar patterns emerge in learners of L2s with different orthographic and phonological characteristics could provide deeper insights into the universality of these findings. Additionally, examining learners at varying proficiency levels could help determine if the influence of PA, MP, and AWM on RC changes as learners advance in their language skills. This could inform targeted interventions tailored to specific proficiency levels.

References

Alinte, C. (2013). Teaching grammar through music. *The Journal of Linguistic and Intercultural Education*, *6*, 7-27.

Antonakis, J., Bendahan, S., Jacquart, P., & Lalive, R. (2014). Causality and endogeneity: Problems and solutions. *The Oxford handbook of leadership and organizations*, *1*(6), 93-117.

Armbruster, B. B., Lehr, F., & Osborn, J. (2003). *Put reading first: The research building blocks of reading instruction kindergarten through grade 3* (2nd ed.). Washington, DC: Partnership for Reading.

Arndt, C., Schlemmer, K., & Van der Meer, E. (2023). The Relationship of Musical Expertise, Working Memory, and Intelligence. *Music Perception: An Interdisciplinary Journal*, *40*(4), 334-346.

Ashtiani, F. T., & Zafarghandi, A. M. (2015). The effect of English verbal songs on connected speech aspects of adult English learners' speech production. *Advances in language and literary studies*, *6*(1), 212-226.

Assche, E. V., Duyck, W., & Hartsuiker, R. J. (2012). Bilingual word recognition in a sentence context. *Frontiers in psychology*, *3*, 1-8.

Augustine, C. (2015). How the use of music and movement impacts the learning of reading skills by preschoolers. *Malaysian Journal of Music*, *4*(2), 74-90.

Ažman, S. (2011). *Fonološko zavedanje in branje pred otrokovim vstopom v šolo ter v prvem in drugem razredu* (Doctoral dissertation), Univerza v Ljubljani, Pedagoška fakulteta.

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556-559.

Baddeley, A., Logie, R., Nimmo-Smith, I., & Brereton, N. (1985). Components of fluent reading. *Journal of memory and language*, *24*(1), 119-131.

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, *51*(6), 1173-1182.

Bell, E., Bryman, A., & Harley, B. (2022). *Business research methods*. Oxford: Oxford University Press.

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*(1), 100-130.

Birch, B. M., & Fulop, S. (2020). *English L2 reading: Getting to the bottom*. New York: Routledge.

Bolduc, J., Gosselin, N., Chevrette, T., & Peretz, I. (2021). The impact of music training on inhibition control, phonological processing, and motor skills in kindergarteners: a randomized control trial. *Early Child Development and Care*, *191*(12), 1886-1895.

Bonacina, S., Cancer, A., Lanzi, P. L., Lorusso, M. L., & Antonietti, A. (2015). Improving reading skills in students with dyslexia: the efficacy of a sublexical training with rhythmic background. *Frontiers in psychology*, *6*, 1-8.

Bonacina, S., Krizman, J., White-Schwoch, T., Nicol, T., & Kraus, N. (2020). Distinct rhythmic abilities align with phonological awareness and rapid naming in school-age children. *Cognitive processing*, *21*, 575-581.

Brandt, A., Gebrian, M., & Slevc, L. R. (2012). Music and early language acquisition. *Frontiers in psychology*, *3*, 1-17.

Bray, J. H., & Maxwell, S. E. (1985). *Multivariate analysis of variance* (No. 54). Beverly Hills, CA: Sage Publications.

Bus, A. G., & Van IJzendoorn, M. H. (1999). Phonological awareness and early reading: A meta-analysis of experimental training studies. *Journal of educational psychology*, *91*(3), 403-414.

Cao, F., Vu, M., Lung Chan, D. H., Lawrence, J. M., Harris, L. N., Guan, Q., ... & Perfetti, C. A. (2013). Writing affects the brain network of reading in Chinese: A functional magnetic resonance imaging study. *Human brain mapping*, *34*(7), 1670-1684.

Cardillo, G. C. (2008). Relationships among prosodic sensitivity, musical processing, and phonological awareness in pre-readers. In *Speech Prosody: Campinas, ISCA Archive* (pp. 595-598).

Castro, A., Alves, D. C., Correia, S., & Soares, C. C. C. (2018). *Phonological awareness screening and assessment tool for European Portuguese speaking children*. Paper presetned at the 10th European Congress of Speech and Language Therapy, Cascais, Potugal.

Chen, I. S. J. (2020). Music as a mnemonic device for foreign vocabulary learning. *English Teaching & Learning*, *44*(4), 377-395.

Chin, W. W. (1998). The partial least squares approach to structural equation modeling. *Modern methods for business research*, *295*(2), 295-336.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale: Lawrence Erlbaum Associates.

Christiner, M., & Reiterer, S. M. (2013). Song and speech: examining the link between singing talent and speech imitation ability. *Frontiers in psychology*, *4*, 1-11.

Christiner, M., Renner, J., Groß, C., Seither-Preisler, A., Benner, J., & Schneider, P. (2022). Singing Mandarin? What short-term memory capacity, basic auditory skills, and musical and singing abilities reveal about learning Mandarin. *Frontiers in Psychology*, *13*, 1-14.

Colwell, C. M., & Murlless, K. D. (2002). Music activities (singing vs. chanting) as a vehicle for reading accuracy of children with learning disabilities: A pilot study. *Music therapy perspectives*, *20*(1), 13-19.

Coumel, M., Groß, C., Sommer-Lolei, S., & Christiner, M. (2023). The contribution of music abilities and phonetic aptitude to L2 accent faking ability. Languages, *8*(1), 68-86.

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

Danne, M. C., Campbell, J. R., Grigg, W. S., Goodman, M. J., & Oranje, A. (2005*). Fourth-Grade Students Reading Aloud: NAEP 2002 Special Study of Oral Reading. The Nation's Report Card* (NCES 2006-469). *Washington, DC: National Center for Education Statistics*.

Degé, F., Kubicek, C., & Schwarzer, G. (2015). Associations between musical abilities and precursors of reading in preschool aged children. *Frontiers in psychology*, *6*, 1-10.

Degé, F., Müllensiefen, D., & Schwarzer, G. (2020). Singing abilities and phonological awareness in 9-to 12-Year-Old children. *Jahrbuch Musikpsychologie*, *29*, 1-20.

Degé, F., & Schwarzer, G. (2017). Music lessons and verbal memory in 10-to 12-year-old children: Investigating articulatory rehearsal as mechanism underlying this association. *Psychomusicology: Music, Mind, and Brain*, *27*(4), 256-266.

Dean, P. (2020, 24 March). *Uni Surrey: Open Resources. Gorilla Experiment Builder.* https://app.gorilla.sc/openmaterials/50646.

Diamantopoulos, A., & Winklhofer, H. M. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of marketing research*, *38*(2), 269-277.

Diamantopoulos, A., & Siguaw, J. A. (2006). Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. *British journal of management*, *17*(4), 263-282.

D'Souza, A. A., Moradzadeh, L., & Wiseheart, M. (2018). Musical training, bilingualism, and executive function: working memory and inhibitory control. *Cognitive research: principles and implications*, *3*, 1-18.

Cambridge University Press. (n.d.). CAE Reading and Use of English Part 5. Exam English. https://www.examenglish.com/CAE/cae_reading_use_of_english5.htm.

Christiner, M., & Reiterer, S. M. (2015). A Mozart is not a Pavarotti: Singers outperform instrumentalists on foreign accent imitation. *Frontiers in human neuroscience*, *9*, 1-8.

Corrigall, K. A., & Trainor, L. J. (2011). Associations between length of music training and reading skills in children. *Music perception*, *29*(2), 147-155.

Culp, M. E. (2017). The relationship between phonological awareness and music aptitude. *Journal of Research in Music Education*, *65*(3), 328-346.

Edwards, L. (2012). *Music and movement: A way of life for the young child*. New Jersey: Pearson Higher Ed.

Eccles, R., Van der Linde, J., Le Roux, M., Holloway, J., MacCutcheon, D., Ljung, R., & Swanepoel, D. W. (2021a). Effect of music instruction on phonological awareness and early literacy skills of five-to seven-year-old children. *Early Child Development and Care*, *191*(12), 1896-1910.

Eccles, R., Van der Linde, J., Le Roux, M., Swanepoel, D. W., MacCutcheon, D., & Ljung, R. (2021b). The effect of music education approaches on phonological awareness and early literacy: A systematic review. *The Australian Journal of Language and Literacy*, *44*(1), 46-60.

Fagerland, R. (2006). Can Music Be Used Effectively To Teach Grammar?. *A Journal for Minnesota and Wisconsin Teachers of English to Speakers of Other Languages*, *41*, 41-58.

Fennell, A. M., Bugos, J. A., Payne, B. R., & Schotter, E. R. (2021). Music is similar to language in terms of working memory interference. *Psychonomic Bulletin & Review*, *28*, 512-525.

Field, A. (2013). *Discovering statistics using SPSS*. London: Sage Publications.

Foncubierta, J. M., Machancoses, F. H., Buyse, K., & Fonseca-Mora, M. C. (2020). The acoustic dimension of reading: Does musical aptitude affect silent reading fluency?. *Frontiers in neuroscience*, *14*, 1-13.

Fonseca-Mora, M. C., Jara-Jiménez, P., & Gómez-Domínguez, M. (2015). Musical plus phonological input for young foreign language readers. *Frontiers in psychology*, *6*, 1-9.

Fonseca-Mora, M. C., Machancoses, F. H., Gryb, O., & Reiterer, S. (2021). Musical aptitude, working memory, general intelligence and plurilingualism: When adults learn to read fluently in a foreign language. *Cogent Education*, *8*(1), 1-16.

Forgeard, M., Schlaug, G., Norton, A., Rosam, C., Iyengar, U., & Winner, E. (2008). The relation between music and phonological processing in normal-reading children and children with dyslexia. *Music perception*, *25*(4), 383-390.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research*, *18*(1), 39-50.

Gagnon, R., & Nicoladis, E. (2021). Musicians show greater cross-modal integration, intermodal integration, and specialization in working memory than non-musicians. *Psychology of Music*, *49*(4), 718-734.

Garcia, L. M., Feldberg, C., Tartaglini, M. F., Hermida, P. D., Irrazabal, N., Somale, V., & Stefani, D. (2020). Memory and music: The role of musical training in verbal and visual working memory in musician and non-musician adult subjects: Preliminary results: Prevention (nonpharmacological)/Lifestyle factors (eg, smoking, etc.). *Alzheimer's & Dementia*, *16*, Article e038592.

Gordon, R. L., Fehd, H. M., & McCandliss, B. D. (2015). Does music training enhance literacy skills? A meta-analysis. *Frontiers in psychology*, *6*, 1-16.

Gough, P. B., & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and special education*, *7*(1), 6-10.

Grabe, W., & Stoller, F. L. (2019). *Teaching and researching reading*. London: Routledge.

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual review of psychology*, *48*(1), 163-189.

Guo, X., Ohsawa, C., Suzuki, A., & Sekiyama, K. (2018). Improved digit span in children after a 6-week intervention of playing a musical instrument: an exploratory randomized controlled trial. *Frontiers in psychology*, *8*, 1-9.

Hair, J. F., Ringle, C. M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing theory and Practice*, *19*(2), 139-152.

Hair, J. F., Astrachan, C. B., Moisescu, O. I., Radomir, L., Sarstedt, M., Vaithilingam, S., & Ringle, C. M. (2021). Executing and interpreting applications of PLS-SEM: Updates for family business researchers. *Journal of Family Business Strategy*, *12*(3), 1-8.

Hansen, M., Wallentin, M., & Vuust, P. (2013). Working memory and musical competence of musicians and non-musicians. *Psychology of Music*, *41*(6), 779-793.

Hargreaves, D. J. (2002). What are musical identities, and why are they important? In R. MacDonald, D. Hargreaves, and D. Miell (Eds.), *Musical Identities* (pp. 1-20). Oxford: Oxford University Press.

Harris, R. J. (2001). *A primer of multivariate statistics*. Mahwah, NJ: Lawrence Erlbaum Associates.

Harvey, S., & Goudvis, A. (2007). *Strategies that work: Teaching comprehension for understanding and engagement*. Portland: Stenhouse publishers.

Hastings and Prince Edward District School Board (HPEDSB). (2013). Quick Phonological Awareness Screening. https://www.uen.org/syc/downloads/Handout6_QPAS.pdf.

Heikkola, L. M., & Alisaari, J. (2019). The Effects of Songs on L2 Proficiency and Spoken Fluency: A Pedagogical Perspective. In L. Pekka, M. Mutta, and P. Peltonen (Eds.), *Fluency in L2 Learning and Use* (pp. 166-185). Bristol: Multilingual Matters.

Herrera, L., Lorenzo, O., Defior, S., Fernandez-Smith, G., & Costa-Giomi, E. (2011). Effects of phonological and musical training on the reading readiness of native-and foreign-Spanish-speaking children. *Psychology of Music*, *39*(1), 68-81.

Hill, A. D., Johnson, S. G., Greco, L. M., O'Boyle, E. H., & Walter, S. L. (2021). Endogeneity: A review and agenda for the methodology-practice divide affecting micro and macro research. *Journal of Management*, *47*(1), 105-143.

Holmes, B. C., & Allison, R. W. (1986). The effect of four modes of reading on children's comprehension. *Literacy Research and Instruction*, *25*(1), 9-20.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and writing*, *2*, 127-160.

Hudson, R. F., Pullen, P. C., Lane, H. B., & Torgesen, J. K. (2008). The complex nature of reading fluency: A multidimensional view. *Reading & Writing Quarterly*, *25*(1), 4-32.

Imm, J. H., Kang, E., Youn, T., Park, H., Kim, J. I., Kang, J. I., ... & Park, H. J. (2008). Different hemispheric specializations for pitch and audioverbal working memory. *Neuroreport*, *19*(1), 99-103.

Jain, C., Devi, N., Parthasarathy, S., & Kavitha, S. (2019). Effect of musical training on psychophysical abilities and working memory in children. *Journal of Indian Speech Language & Hearing Association*, *33*(2), 71-74.

Janurik, M., Surján, N., & Józsa, K. (2022). The relationship between early word reading, phonological awareness, early music reading and musical aptitude. *Journal of Intelligence*, *10*(3), 50-64.

Jekiel, M., & Malarski, K. (2021). Musical hearing and musical experience in second language English vowel acquisition. *Journal of Speech, Language, and Hearing Research*, *64*(5), 1666-1682.

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language learning*, *64*(1), 160-212.

Jeong, E., & Ryu, H. (2016). Nonverbal auditory working memory: Can music indicate the capacity?. *Brain and Cognition*, *105*, 9-21.

Kato, S. (2009). Suppressing inner speech in ESL reading: Implications for developmental changes in second language word recognition processes. *The Modern Language Journal*, *93*(4), 471-488.

Kempert, S., Götz, R., Blatter, K., Tibken, C., Artelt, C., Schneider, W., & Stanat, P. (2016). Training early literacy related skills: To which degree does a musical training contribute to phonological awareness development?. *Frontiers in psychology*, *7*, 1-16.

Kintsch, W., & Kintsch, E. (2005). Comprehension. In S. Paris and S. Stahl (Eds.), *Children's reading comprehension and assessment* (pp. 89-110). Mahwah, NJ: Lawrence Erlbaum Associates.

Kock, N., & Lynn, G. (2012). Lateral collinearity and misleading results in variance-based SEM: An illustration and recommendations. *Journal of the Association for information Systems*, *13*(7), 546-580.

Koda, K. (2007). Reading and language learning: Crosslinguistic constraints on second language reading development. *Language learning*, *57*, 1-44.

Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Müller, K., & Gruber, O. (2009). Functional architecture of verbal and tonal working memory: an FMRI study. *Human brain mapping*, *30*(3), 859-873.

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-II* (2nd ed.). San Antonio: Pearson PsychCorp.

Leloup, G., Anders, R., Charlet, V., Eula-Fantozzi, B., Fossoud, C., & Cavalli, E. (2021). Improving reading skills in children with dyslexia: efficacy studies on a newly proposed remedial intervention—repeated reading with vocal music masking (RVM). *Annals of Dyslexia*, *71*, 60-83.

Li, P., Zhang, Y., Fu, X., Baills, F., & Prieto, P. (2022). Melodic perception skills predict Catalan speakers' speech imitation abilities of unfamiliar languages. In *Proceedings of the International Conference on Speech Prosody 2022* (pp. 876-880).

Liu, J. (2012). *Aligning TEM-4 with the CEFR* (Master's thesis). Henan Normal University.

Loui, P., Kroog, K., Zuk, J., Winner, E., & Schlaug, G. (2011). Relating pitch awareness to phonemic awareness in children: implications for tone-deafness and dyslexia. *Frontiers in psychology*, *2*, 1-5.

Lucas, J. R., & Gromko, J. E. (2007). The relationship of musical pattern discrimination skill and phonemic awareness in beginning readers. *Contributions to Music Education*, *34*, 9-17.

Mandell, J. (2009). *Electronic music and medical education*. Accessed September 1, 2019, from http://jakemandell.com.

Mannarelli, P., & Serrano, R. (2024). 'Thank you for the music': examining how songs can promote vocabulary learning in an EFL class. *The Language Learning Journal*, *52*(1), 1-15.

Melby-Lervåg, M., & Lervåg, A. (2011). Cross-linguistic transfer of oral language, decoding, phonological awareness and reading comprehension: A meta-analysis of the correlational evidence. *Journal of research in reading*, *34*(1), 114-135.

Morrongiello, B. A., & Roes, C. L. (1990). Developmental changes in children's perception of musical sequences: Effects of musical training. *Developmental Psychology*, *26*(5), 814-820.

Mullis, I. V., & Martin, M. O. (2019). *PIRLS 2021 Assessment Frameworks*. Amsterdam, the Netherlands: International Association for the Evaluation of Educational Achievement.

Nie, P., Wang, C., Rong, G., Du, B., Lu, J., Li, S., ... & Tervaniemi, M. (2022). Effects of music training on the auditory working memory of Chinese-speaking school-aged children: a longitudinal intervention study. *Frontiers in Psychology*, *12*, 1-9.

Nisha, K. V., Neelamegarajan, D., Nayagam, N. N., Winston, J. S., & Anil, S. P. (2021). Musical aptitude as a variable in the assessment of working memory and selective attention tasks. *Journal of Audiology & Otology*, *25*(4), 178-188.

Nunnally, J., & Bernstein, I. (1994). *Psychometric Theory (3rd edition)*, New York: MacGraw-Hill.

Okhrei, A., Kutsenko, T., & Makarchuk, M. (2016). Performance of working memory of musicians and non-musicians in tests with letters, digits, and geometrical shapes. *Biologija*, *62*(4), 207-215.

Olson, C. L. (1974). Comparative robustness of six tests in multivariate analysis of variance. *Journal of the American Statistical Association*, *69*(348), 894-908.

Özata, H., Babür, N., & Haznedar, B. (2016). Phonological awareness in reading acquisition. In B. Haznedar and F. Ketrez (Eds.), *The Acquisition of Turkish in Childhood* (pp. 243-271). Amsterdam: John Benjamins.

Paap, K. R., & Greenberg, Z. I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive psychology*, *66*(2), 232-258.

Patel, A. D. (2011). Why would musical training benefit the neural encoding of speech? The OPERA hypothesis. *Frontiers in psychology*, *2*, 1-14.

Patel, A. D. (2012). Language, music, and the brain: a resource-sharing framework. In P. Rebuschat, M. Rohrmeier, J. Hawkins, and I. Cross (Eds.), *Language and Music as Cognitive Systems* (pp. 204-223). Oxford: Oxford University Press.

Patscheke, H., Degé, F., & Schwarzer, G. (2016). The effects of training in music and phonological skills on phonological awareness in 4-to 6-year-old children of immigrant families. *Frontiers in psychology*, *7*, 1-13.

Phillips, S., & Kelly, K. (2018). *Assessment of learners with dyslexic-type difficulties*. London: Sage.

Pica, R. (2010). Linking literacy and movement. *Young children*, *65*(6), 72-73.

Pilonieta, P. (2012). How Fast Is Too Fast? Fluency Instruction in the Classroom. *California Reader*, *45*(3), 8-12.

Politimou, N., Dalla Bella, S., Farrugia, N., & Franco, F. (2019). Born to speak and sing: Musical predictors of language development in pre-schoolers. *Frontiers in Psychology*, *10*, 1-18.

Rammsayer, T., & Altenmüller, E. (2006). Temporal information processing in musicians and nonmusicians. *Music Perception*, *24*(1), 37-48.

Register, D., Darrow, A. A., Swedberg, O., & Standley, J. (2007). The use of music to enhance reading skills of second grade students and students with reading disabilities. *Journal of Music Therapy*, *44*(1), 23-37.

Reynolds, C. R. (1997). Forward and backward memory span should not be combined for clinical analysis. *Archives of Clinical Neuropsychology*, *12*(1), 29-40.

Rimmer, C., Dahary, H., & Quintin, E. M. (2024). Links between musical beat perception and phonological skills for autistic children. *Child Neuropsychology*, *30*(3), 361-380.

Saad, M., Ong, M. H. A., Ali, A. M., Samat, N., & Ahmad, R. (2020). Moderating Work Efficiency and Performance Through Best-Practice Certification: Managementâ€™ s Perspective. *International Journal of Economics, Management and Accounting*, *28*(1), 123-140.

Saarikivi, K. A., Huotilainen, M., Tervaniemi, M., & Putkinen, V. (2019). Selectively enhanced development of working memory in musically trained children and adolescents. *Frontiers in Integrative Neuroscience*, *13*, 1-12.

Sarstedt, M., Ringle, C. M., Cheah, J. H., Ting, H., Moisescu, O. I., & Radomir, L. (2020). Structural model robustness checks in PLS-SEM. *Tourism Economics*, *26*(4), 531-554.

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*, *36*(2), 110-124.

Seashore C.E., Lewis D., & Saetveit J.G. (1960). *Seashore measures of musical talent*. New York: The Psychological Corporation.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: an individual differences approach. *Journal of experimental psychology: General*, *125*(1), 4-27.

Shen, Y., Lin, Y., Liu, S., Fang, L., & Liu, G. (2019). Sustained effect of music training on the enhancement of executive function in preschool children. *Frontiers in Psychology*, *10*, 1-14.

Silas, S., Müllensiefen, D., Gelding, R., Frieler, K., & Harrison, P. M. (2022). The associations between music training, musical working memory, and visuospatial working memory: An opportunity for causal modeling. *Music Perception: An Interdisciplinary Journal*, *39*(4), 401-420.

Skinner, B. F. (1938). *The behaviour of organisms*. New York: Appleton-Century-Crofts.

Skubic, D., Gaberc, B., & Jerman, J. (2021). Supportive development of phonological awareness through musical activities according to Edgar Willems. *SAGE open*, *11*(2), 1-11.

Snowling, M. J., & Hulme, C. (2012). Annual Research Review: The nature and classification of reading disorders–a commentary on proposals for DSM-5. *Journal of child psychology and psychiatry*, *53*(5), 593-607.

Sousa, J., Martins, M., Torres, N., Castro, S. L., & Silva, S. (2022). Rhythm but not melody processing helps reading via phonological awareness and phonological memory. *Scientific Reports*, *12*(1), 1-11.

Spiegel, M. F., & Watson, C. S. (1984). Performance on frequency-discrimination tasks by musicians and nonmusicians. *The Journal of the Acoustical Society of America*, *76*(6), 1690-1695.

Sparks, R. L. (2015). Language deficits in poor L2 comprehenders: The simple view. *Foreign Language Annals*, *48*(4), 635-658.

Steinbrink, C., Knigge, J., Mannhaupt, G., Sallat, S., & Werkle, A. (2019). Are temporal and tonal musical skills related to phonological awareness and literacy skills?–Evidence from two cross-sectional studies with children from different age groups. *Frontiers in Psychology*, *10*, 1-16.

Stevens, E. A., Park, S., & Vaughn, S. (2019). A review of summarizing and main idea interventions for struggling readers in grades 3 through 12: 1978–2016. *Remedial and Special Education*, *40*(3), 131-149.

Strait, D. L., Hornickel, J., & Kraus, N. (2011). Subcortical processing of speech regularities underlies reading and music aptitude in children. *Behavioral and Brain Functions*, *7*, 1-11.

Sun, B., Yang, J., & Liang, Z. (2024). How music facilitates second language (L2) learning: a systematic review from 2002 to 2022. *Innovation in Language Learning and Teaching*, 1-19.

Sun, Y., Lu, X., Ho, H. T., & Thompson, W. F. (2017). Pitch discrimination associated with phonological awareness: Evidence from congenital amusia. *Scientific Reports*, *7*(1), 1-11.

Swaminathan, S., & Schellenberg, E. G. (2020). Musical ability, music training, and language ability in childhood. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(12), 2340-2348.

Talamini, F., Altoè, G., Carretti, B., & Grassi, M. (2017). Musicians have better memory than nonmusicians: A meta-analysis. *PloS one*, *12*(10), 1-21.

Talamini, F., Carretti, B., & Grassi, M. (2016). The working memory of musicians and nonmusicians. *Music Perception: An Interdisciplinary Journal*, *34*(2), 183-191.

Talamini, F., Grassi, M., Toffalini, E., Santoni, R., & Carretti, B. (2018). Learning a second language: Can music aptitude or music training have a role?. *Learning and Individual Differences*, *64*, 1-7.

Tan, L. H., Spinks, J. A., Feng, C. M., Siok, W. T., Perfetti, C. A., Xiong, J., ... & Gao, J. H. (2003). Neural systems of second language reading are shaped by native language. *Human brain mapping*, *18*(3), 158-166.

Tasnim, Z. (2022). Songs for EFL/ESL Class: How to Teach Listening Skill through Music. *MEXTESOL Journal*, *46*(3), 1-12.

Taylor, A. C., & Dewhurst, S. A. (2017). Investigating the influence of music training on verbal memory. *Psychology of Music*, *45*(6), 814-820.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schröger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: an event-related potential and behavioral study. *Experimental brain research*, *161*, 1-10.

Verhoeven, L., & Van Leeuwe, J. (2012). The simple view of second language reading throughout the primary grades. *Reading and writing*, *25*, 1805-1818.

Vidal, M. M., Lousada, M., & Vigário, M. (2020). Music effects on phonological awareness development in 3-year-old children. *Applied Psycholinguistics*, *41*(2), 299-318.

Wang, C., Tao, S., Tao, Q., Tervaniemi, M., Li, F., & Xu, P. (2020). Musical experience may help the brain respond to second language reading. *Neuropsychologia*, *148*, 1-9.

Wechsler, D., Coalson, D. L., and Raiford, S. E. (2008). *WAIS-IV Technical and Interpretive Manual*. San Antonio, Texas: Pearson.

Yapp, D. J., de Graaff, R., & van den Bergh, H. (2021). Improving second language reading comprehension through reading strategies: A meta-analysis of L2 reading strategy interventions. *Journal of Second Language Studies*, *4*(1), 154-192.

Yates, C. M., Justus, T., Atalay, N. B., Mert, N., & Trehub, S. E. (2017). Effects of musical training and culture on meter perception. *Psychology of Music*, *45*(2), 231-245.

Yeşil, B., & Ünal, S. (2017). An investigation on the effects of music training on attention and working memory in adults. *Anatolian Journal of Psychiatry/Anadolu Psikiyatri Dergisi*, *18*(6), 531-535.

Yousefi, A., Yekta, R. R., & Farahmandian, H. (2014). The effect of modern lyrical music on second language vocabulary acquisition. *Mediterranean Journal of Social Sciences*, *5*(23), 2583-2586.

Zentner, M. & Strauß, H. (2017). Assessing musical ability quickly and objectively: Development and validation of the Short-Proms and the Mini-Proms. *Annals of the New York Academy of Sciences*, *1400*, 33-45.

Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological bulletin*, *131*(1), 3-29.

Appendix A

Background Information Questionnaire

1. What is your gender?
   A. Male
   B. Female
   C. Prefer to self-describe as non-binary, gender-fluid, agender, etc.

2. What is your year of birth?
   A. 2006
   B. 2005
   C. 2004
   D. 2003
   E. 2002
   F. 2001
   G. 2000
   H. 1999 and earlier

3. What is the level of your English?
   A. Native-like/mother-tongue-like
   B. Advanced/professional
   C. Intermediate/fluent
   D. Beginner

4. Do you have any English qualifications?
   A. College English Test Band 4
   B. College English Test Band 6
   C. Test for English Majors Band 4
   D. Other – please specify: ____________

5. Besides Mandarin and English, do you speak, or have you studied any other languages for more than half a year? If so, please specify.

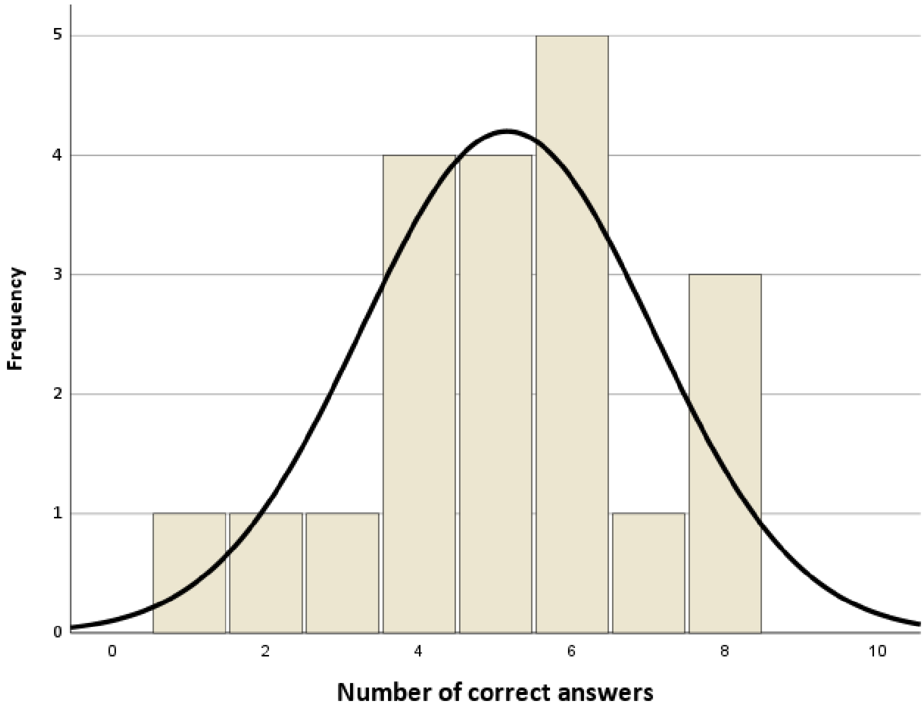
6. Were you involved in any formal music-related activities before university enrolment?
   A. Never.
   B. I just took the compulsory music lessons when I was in primary and/or secondary school.
   C. Besides compulsory music lessons in school, I used to regularly participate in after-class activities (e.g., singing, playing the instrument, composing) for more than a year.

7. Have you been engaging in any music-related activities (e.g. singing, playing the instrument, composing) since university enrolment until now? If yes, how many hours per week do you spend on these activities?
   A. None
   B. Less than 1h
   C. 2-5h
   D. 5-10h
   E. More than 10h

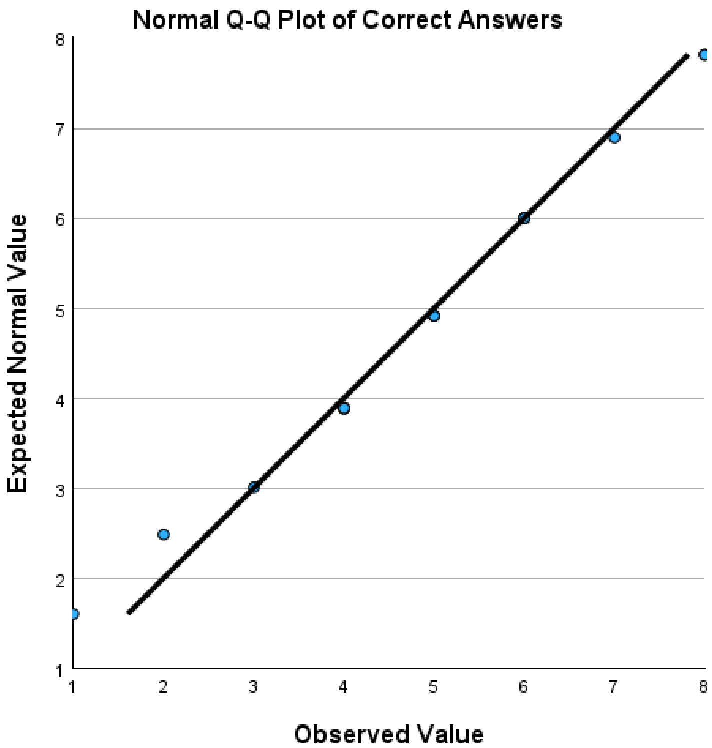
8. Do you usually listen to music? If yes, how much time is spent per week?
   A. Less than 1h
   B. 2-5h
   C. 5-10h
   D. More than 10h

Appendix B
Pilot Study Data

B.1 Histogram of pilot participants' correct answers to the reading test



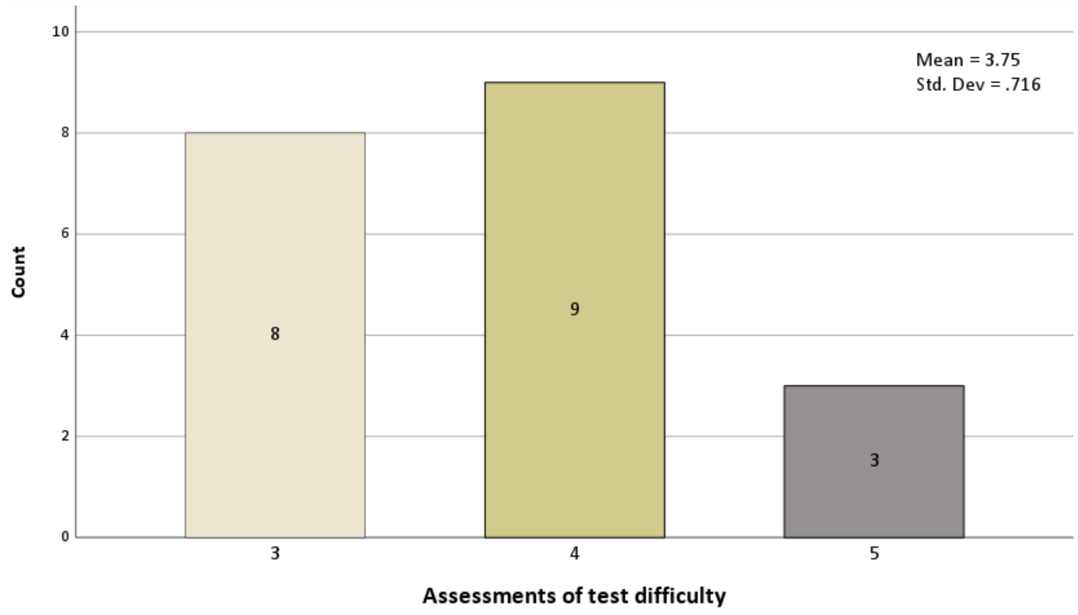
B.2 Q-Q plot of pilot participants' correct answers to the reading test



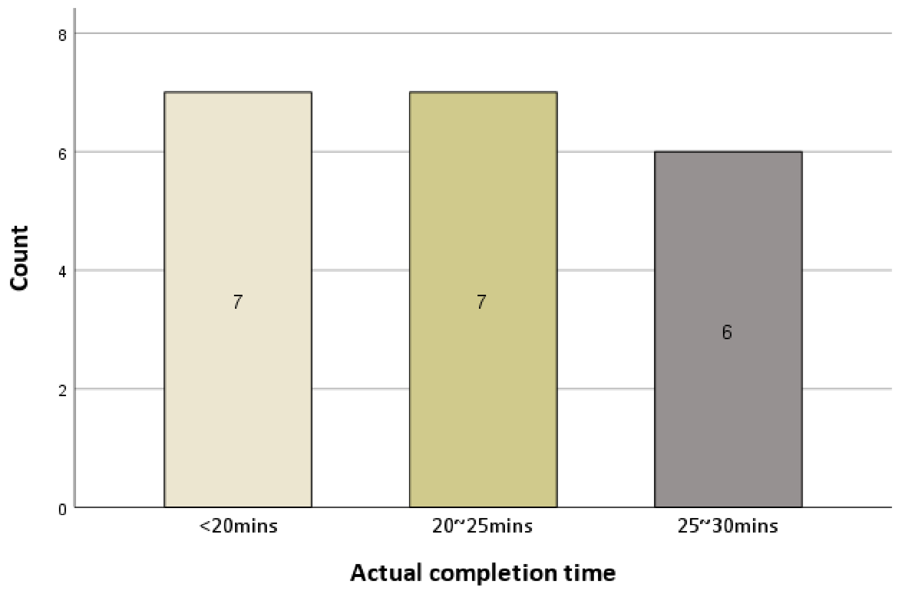
B.3 Descriptive statistics for pilot participants' correct answers to the reading test

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
Correct	20	1	8	5.15	1.899	-.341	.512	-.011	.992
Valid N (listwise)	20								

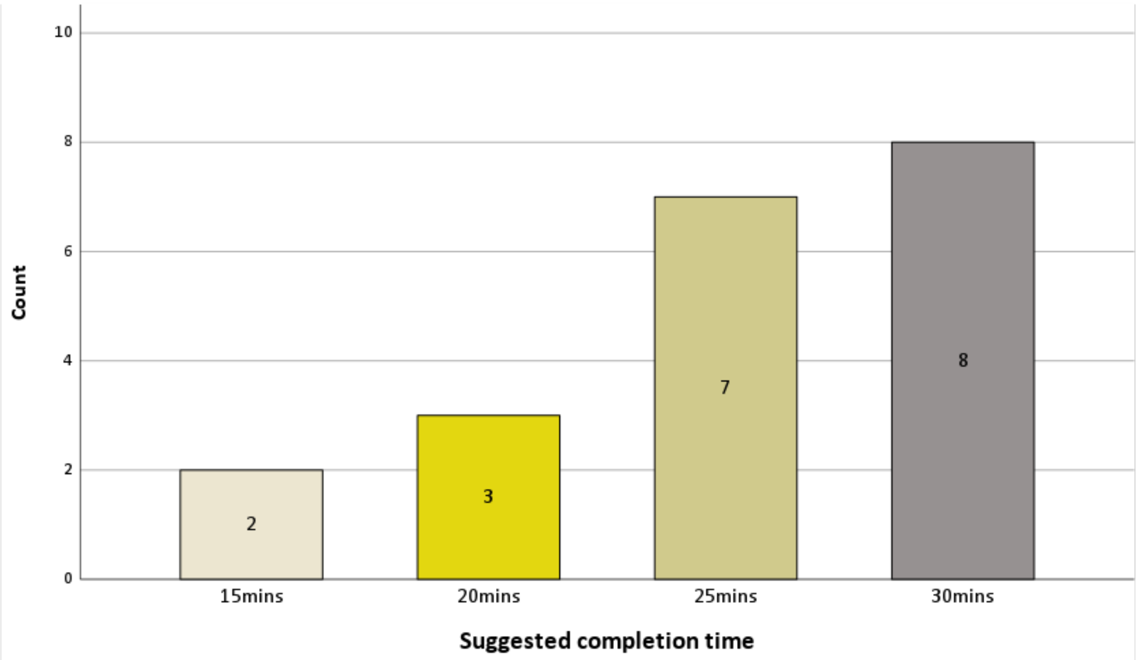
B.4 Histogram of pilot participants' assessments of reading test difficulty



B.5 Histogram of pilot participants' actual completion time



B.6 Histogram of pilot participants' suggested completion time



B.7 Descriptive statistics for pilot participants' ratings on two speakers

	N	Minimum	Maximum	Mean	Std. Deviation
Rating on the UK speaker	20	1	2	1.30	.470
Rating on the USA speaker	20	1	2	1.40	.503
Valid N (listwise)	20				

Appendix C

English Silent Reading Comprehension Test

You are going to read a magazine article about an African film festival. For questions 1-6, choose the answer which you think best fits according to the text.

After a bumpy 225km drive from a meagre airstrip in Tindouf, south western Algeria, a sprawling single-story town begins to emerge from the desert's dust. As the sun climbs in the cloudless sky, visitors are rewarded with their first glimpse of Dakhla refugee camp. It isn't the most obvious setting for a film festival, but for seven years, just before the glitz and glamour of Cannes, the Sahrawi people of Dakhla have hosted actors and film-makers from around the world for this six-day event. This year, for the first time, direct flights were laid on from London, giving the opportunity for overseas visitors to play a part in this extraordinary occasion. But despite the energy and excitement, the background to the film festival is a serious one, as the Sahrawi people have been living for thirty years in this isolated desert outpost, having been forced to flee their native Western Sahara.

Western Sahara, Africa's last colony, was taken over by Morocco when the Spanish withdrew in 1976, despite a ruling from the International Court of Justice. This was followed by a brutal 16-year war, during which time tens of thousands of Sahrawis fled across the Algerian border to refugee camps. In 1991, a ceasefire agreement was drawn up, in which a referendum on self-determination was promised to decide the fate of the country and its people. However, almost twenty years later, the gears of diplomacy have turned slowly and nothing has happened. Meanwhile the refugees have been left stranded in five refugee camps dotted around the vast, inhospitable desert.

Dakhla, home to nearly 30,000 of these refugees, is the most remote of these camps, being located 175 km from the nearest city. Unlike its namesake, the beautiful coastal city in Western Sahara, this Dakhla has no paved roads and is entirely dependent on outside supplies for food and water. Temperatures regularly top 120 degrees, there is minimal vegetation and there are frequent sandstorms. Locally it is known as the Devil's Garden. Despite these obvious setbacks, the town is clean and well organised, with wide sandy streets. Houses and tents are grouped in neat family compounds. There are hospitals, funded by aid agencies, and a good standard of education. For the duration of the festival, an articulated lorry is parked in the central compound, and a multiplex-sized screen is mounted on its side. Around it are stalls and tents housing workshops and exhibitions.

The aim of the festival is to raise international awareness on the plight of the refugees. However, it also offers a rare chance for the refugees to go to the movies and experience some educational opportunities. It is hoped that it might foster a new generation of Sahrawi film-makers, especially as this year, the festival also celebrated the opening of a permanent film, radio and television school in a neighbouring camp.

The program of films for this year included over forty films from around the world. Films range from international blockbusters to various works on and by the Sahrawi people. The themes mostly centre on experiences of struggle and hope, but there were lighter moments, such as an animated film for the children and a flash of Rachel Weisz's naked bottom during the ancient Egyptian epic Agora which proved to be a highlight for many older boys. However, the runaway favourite was 'a Victime', a documentary about Ibrahim Leibeit, a 19-year-old Sahrawi who lost his leg to a land mine last year.

Films are screened at night, so the daytime is taken up with exhibitions, camel races and football matches. One afternoon the London-based charity 'Sandblast' put on a joint workshop with a film-maker, giving refugees the opportunity to learn about filmmaking and create their own video messages. These were put online so that their extended families in Western Sahara, from whom they have been separated for more than 33 years, could watch them. Helen Whitehead, a film-maker from London said, 'Working together really broke down language and cultural barriers. It was very rewarding, and we came across some real talent.'

More than 500 visitors flew into Tindouf on charter planes and braved the rough drive to the settlement. All the visitors to the festival stay with Sahrawi families, sharing their homes and partaking of their food. Living with these displaced people gives overseas participants an invaluable insight into the conditions in which the refugees live. Alongside the film buffs there are real celebrities such as actors Victoria Demayo and Helena Olano. They are mostly B and C listers from the Spanish film industry, although the real stars do take an interest. Director Javier Cardozo was a visitor last year, and Penelope Cruz is a long-term supporter, but pulled out of attending the festival this year at the last minute. Will the celebrity backing make a difference to the plight of the refugees? Possibly. Cardozo's suggestion that the Spanish, as the ex-colonial masters of Western Sahara, were responsible for the situation received significant coverage in the Spanish Media and put some pressure on the government to take some action. However, although the campaign in Spain is growing steadily, the focus of attention cannot only be on the Spanish government.

On the final day of the gathering, there is a dusty red-carpet ceremony in which the White Camel award for best picture is presented to Jordi Ferrer and Paul Vidal for ‘El Problema’, their 2009 film about Western Sahara. Actors, activists and festival organisers gather on stage in high spirits to show their solidarity with the refugees. But as the stalls are dismantled and the trucks are driven away, the thoughts of the visitors turn to the people they are leaving behind. They may never get the chance to see the world or fulfil their dreams of becoming actors or film-makers. For them, there is nowhere to go. Dakhla is essentially a desert prison.

1. In the first paragraph, the writer emphasises:
   A. the enthusiasm that the festival instils
   B. the sensational nature of the festival
   C. the festival’s increasing media attention
   D. the festival’s unlikely location

2. According to the writer, the refugees have been in the desert for so long because:
   A. International agencies do not know they are there
   B. the Moroccan government disagree with the UN
   C. a proposed vote is yet to take place
   D. there is a war in their home country

3. What does the writer say about the original city of Dakhla?
   A. It is by the sea.
   B. It has good health and educational facilities.
   C. It does not have proper roads.
   D. It gets food and water from aid agencies.

4. What is said about the films shown at the festival?
   A. They mostly show the personal experiences of the Sahwari people.
   B. All of the films are serious in content.
   C. The variety of films suited a wide range of tastes.
   D. The international films were more popular than the local films.

5. What was the British visitors’ response to the workshops?
   A. They were surprised by the refugee’s film knowledge.
   B. The workshops enabled them to communicate with local people.
   C. The workshops taught the visitors a lot about local culture.
   D. They showed the local films to their families via the internet.

6. What point does the writer highlight in the final paragraph?
   A. There is a contrast between the visitors’ freedom and the refugees’ confinement.
   B. The film festival only gives the refugees unattainable dreams.
   C. The visitors only care about the refugees for the duration of the festival.
   D. The festival is a poor copy of the more famous film festivals.

You are going to read four news reports about an abandoned baby. For questions 7-10, choose from the articles A-D. The reviews may be chosen more than once.

Article A

It has been alleged that the mother charged with attempted murder after dumping her newborn baby into a drain has admitted to abandoning the baby boy. The woman remains in custody after bail was formally refused at Blacktown Local Court. The newborn baby was discovered by passing cyclists on a day when temperatures surpassed 40 degrees Celsius. Mr Otte, who discovered the baby and only cycles the route once a month, said, 'That baby had no chance if we and the other people hadn't been there. Something made us find that baby today'. The child was already undernourished, and dehydration would have taken effect and the baby would not have survived the day. Passersby outside court cried 'shame' as the accused woman's relatives crossed the street in Blacktown.

Article B

A woman has been charged with the attempted murder of her newborn son, who was left in a drain on Tuesday before being discovered the following Sunday. The incident has shocked us all. In the searing heat, the baby had little chance of survival, and the mother must have been aware of this as she callously shoved him through the tiny gap, dropped him into the darkness and left him to his fate. But although crimes like this are a rarity, they don't happen in isolation. Australia criminalises child abandonment, thus making it nigh on impossible for a depressed mother to give up her infant without causing it harm. Meanwhile pregnant women are shuffled through the system, rarely seeing the same caregiver twice. This model of care treats the pregnancy, but ignores the patient, and it is this ill-equipped, indifferent system that makes a crime this one possible.

Article C

A mother has been charged with attempted murder after she allegedly abandoned her newborn son in a roadside drain, police confirmed today. The baby had been alone in the deep drain for five days when, by a stroke of luck, cyclists caught the faint sound of his muffled cries above the heavy noise of motorway traffic. Sweltering temperatures in Sydney have settled around 30°C over the past week and it is believed the week-old boy would have died had he not been found. Karen Healy, National President of the Australian Association of Social said that this was a highly unusual case, as parents who abandon their children tend to do so in high-traffic areas like churches or hospitals where the child will be taken care of. This scenario, in which the mother clearly wanted the baby to be hidden and it was only by the grace of God that the infant survived, suggests an element of shame or possible mental illness which was not heeded by pre- and post-natal health providers. The 30-year-old mother is currently receiving therapy while she remains in police custody.

Article D

A newborn baby boy has been rescued from an eight-foot drain beside a bike track in Australia after passing cyclists heard the sounds of wailing. Graham Bridges, who was among the people who helped rescue the baby, walked the bike track regularly and said it was usually very popular with riders on a Sunday morning. Inspector David Lagats said 'We all thought the worst but he's still alive. It was a long drop down, but he's wrapped up pretty well, so that will have cushioned his fall.' The concrete moulding of the drain formed a layer of insulation which protected the baby from the weather, which, during the week, reached temperatures of thirty degrees. The baby was taken to hospital in a stable condition, thanks, in part, to the fact that newborn babies have reserves of fluids and body sugars which they can resort to as they adapt to the new way of feeding. Lisa Charet, from the state department of family and community services said she was concerned for the mother's welfare. "We can give her the help and support that she needs. She must be feeling enormously distressed if she feels that this is the only course of action available to her."

Which article:

7. differs from the others with regards to the threat to the baby's health? **D**
8. shares the same attitude to the mother as article B? **A**
9. shares article B's view that the mother was not entirely responsible for her actions? **C**
10. shares article C's attitude towards the baby's rescue? **A**

Appendix D

Phonological Awareness Test

1. Rhyming perception

You will hear words in pairs or triads. Please check if these words you hear rhyme.

though – low	sip / lip / speak
boy / top	wit / sick / bit
pen – hen	fan / can / ran
fun – sun	ball / fall / law

2. Syllable awareness

Please answer how many syllables you hear in the words.

school (1); Saturday (3); playtime (2); giant (2); fish (1); sunflower (3); caterpillar (4); walking (2)

3. Sound segmentation

Please answer how many sounds you hear in the words. For example, you will hear three sounds 'k', 'a', and 't' in the word 'cat'.

Game (3); happy (4); up (2); trip (3); shoe (2); stop (4); full (3); help (4)

4. Word parts blending

Listen to parts of high-frequency words, piece them together, and write them down. For example, if you hear /k/, /æ/, and /t/, the word is 'cat'. If you hear /pɜːf/ and /ɪkt/, the word is 'perfect'. These parts are arbitrarily separated from the words.

un – der – stand – ing	/ˌʌndəˈstændɪŋ/
s – ou – p	/suːp/
b – as – ic	/ˈbeɪsɪk/
gen – e – ros – i – ty	/ˌdʒenəˈrɒsəti/
clev – er	/ˈklevə(r)/
cin – em – a	/ˈsɪnəmə/
s – p – oo – n	/spuːn/
c – om – b – ine	/kəmˈbaɪn/

Appendix E

CUREC Approval

SOCIAL SCIENCES & HUMANITIES
INTERDIVISIONAL RESEARCH ETHICS COMMITTEE
DEPARTMENTAL RESEARCH ETHICS COMMITTEE

Department of Education
15 Norham Gardens, Oxford OX2 6PY
student.curec@education.ox.ac.uk; staff.curec@education.ox.ac.uk

UNIVERSITY OF OXFORD

Tianai Zhang
Department of Education, Social Sciences Division
University of Oxford

27 February 2024

Dear Tianai,

Research ethics approval

Research title: Music and Language: Exploring the Relationship Between Music Perceptive Skills and the Acoustic Dimension of ESL Learners' Silent Reading Comprehension

Research ethics reference: EDUC_C1A_24_083

The above application has been considered on behalf of the Education Departmental Research Ethics Committee (DREC) in accordance with the University's procedures for ethical approval of all research involving human participants.

I am pleased to confirm that, on the basis of the information provided to the DREC, ethics approval has now been granted for this study.

Please note the following:

Personal data: It is the responsibility of the PI to ensure that all personal data collected during the project is managed in accordance with the University's guidance and legal requirements.

In-person activities: Any data collection involving in-person interactions with participants must have an up-to-date fieldwork risk assessment in place; further guidance is available from the Safety Office's website.

Amendments: Please notify the committee if you intend to make any amendments to the information in your ethics application as submitted at date of this approval, as all changes must receive ethical approval prior to implementation. The amendment form is available on the SSH IDREC webpage.

We welcome feedback on your experience of the ethical review process and suggestions for improvement. Please email any comments to staff.curec@education.ox.ac.uk / student.curec@education.ox.ac.uk or ethics@socsci.ox.ac.uk.

Yours sincerely,

DREC Dr Jenny A. Wynn

Cc. Anna-Maria Ramezanzadeh

Appendix F

Information Sheet and Consent Form

Central University Research Ethics Committee Approval Reference: EDUC_C1A_24_083

Participant Information Sheet

You are being invited to take part in a research project. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask me if there is anything that is not clear or if you would like more information. Take time to decide whether you wish to take part. Thank you for reading this.

Why is this research being conducted?

The similarities and differences in terms of auditory, syntax and semantic processing make it controversial whether musical ability can be transferred to language ability. This study aims to investigate the relationship between music-perceptive skills and the acoustic dimension of silent reading among Chinese university learners with English as a second language. If students with stronger musical skills show better language learning abilities, this may provide strong pedagogical evidence for introducing creative teaching approaches such as songs to foreign language curricula.

Why have I been invited to take part?

The purpose of this study is to explore the music perception ability and silent reading ability of Chinese university students. The study will involve about 120 participants, all from Nantong University. Based on your age (18 years old and above), mother tongue (Chinese as a first language), and physical condition (with normal visual and auditory functions), you have been identified as a potential participant. You do not have to have received any musical instruction to participate.

Do I have to take part?

No. It is up to you to decide whether to take part. You can withdraw yourself from the research, without giving a reason, and without negative consequences, by advising me of this decision at any point during the data collection. Once your data has been collected, the deadline by which you can withdraw any information you have contributed to the research is 1st, May 2024. If you wish to withdraw your data before this point, please contact me by email. If any participants decide to withdraw, their data will be immediately deleted, and the researcher and the organization will no longer retain any information about that participant.

What will happen to me if I take part in the research?

Informed consent forms will be sent to potential participants before the start of the study. Those who sign their names will participate in the study. Participants can ask to pause or stop the research activities at any time. This research will take place in a classroom at Nantong University. All participants will be asked to arrive at the classroom with their computers and headphones during the study. The tasks (1 hour in total) will be conducted in two sessions with a 20-minute break in between. During the first session, participants will fill out a questionnaire about their background information, including gender, age, whether they play an instrument, and any musical training they may have received. This will take around three minutes to complete. Then, they will complete a phonological awareness test (10 mins) and an auditory working memory test (15 mins). During the next session, the participants will take a reading comprehension task (25 mins) and a music perception test (15 mins). These tests will be completed on your computers.

What are the possible disadvantages and risks in taking part?

This study involves several online tasks, requiring participants to pay high attention. Participants may experience mental tiredness. For this reason, the tasks will be spread over two days and the researcher will implement breaks during tasks, and provide clear instructions on task duration. Students are free to pause the test at any point if they feel tired.

Are there any benefits in taking part?

While there are no immediate benefits for those people participating in the research, it is hoped that the study will give participants an initial sense of whether there is some subtle connection between their ability to perceive music and some auditory-related subskills in silent reading.

Are there any expenses and payments?

None.

What information will be collected and why is the collection of this information relevant for achieving the research objectives?

Your scores on music perception, auditory working memory, phonological awareness, and silent reading comprehension will be collected. The tests you complete will help me better understand the intersection between music and language. My supervisor and I will have access to the research data. Identifiable data (including consent forms) will be stored in Nexus 365 OneDrive for Business. Other research data will be stored for 3 years after publication or public release of the work of the research. In particular, your music test results will be shared with the Personality, Emotion, and Music Lab at the University of Innsbruck to help them develop tests in the future.

Will the research be published? Could I be identified from any publications or other research outputs?

The findings from the research will be written up in a thesis. It will not be possible for participants to be identifiable from the output. A copy of my thesis/ dissertation will be deposited both in print and online in the Oxford University Research Archive where [it will be publicly available to facilitate its use in future research/ its access will be restricted].

Data Protection

The University of Oxford is the data controller with respect to your personal data, and as such will determine how your personal data is used in the research. The University will process your personal data for the purpose of the research outlined above. Research is a task that is performed in the public interest. Further information about your rights with respect to your personal data is available from the University's Information Compliance web site at https://compliance.admin.ox.ac.uk/individual-rights.

Who has reviewed this research?

This research has received ethics approval from a subcommittee of the University of Oxford Central University Research Ethics Committee. (Ethics reference: EDUC_C1A_24_083).

Who do I contact if I have a concern about the research or I wish to complain?

If you have a concern about any aspect of this research, please contact Tianai Zhang at tianai.zhang@education.ox.ac.uk or Dr. Anna-Maria Ramezanzadeh at anna-maria.ramezanzadeh@education.ox.ac.uk, and we will do our best to answer your query. We will acknowledge your concern within 10 working days and give you an indication of how it will be dealt with. If you remain unhappy or wish to make a formal complaint, please contact the Chair of the Research Ethics Committee at the University of Oxford who will seek to resolve the matter as soon as possible:

The Chair, Social Sciences & Humanities Interdivisional Research Ethics Committee; Email: ethics@socsci.ox.ac.uk; Address: Research Services, University of Oxford, Boundary Brook House, Churchill Drive, Headington, Oxford OX3 7GB

Further Information and Contact Details

If you would like to discuss the research with someone beforehand (or if you have questions afterwards), please contact:

Tianai Zhang
Department of Education
OX2, 6PY
University tel: +44 (0) 1865 274024
University email: tianai.zhang@education.ox.ac.uk

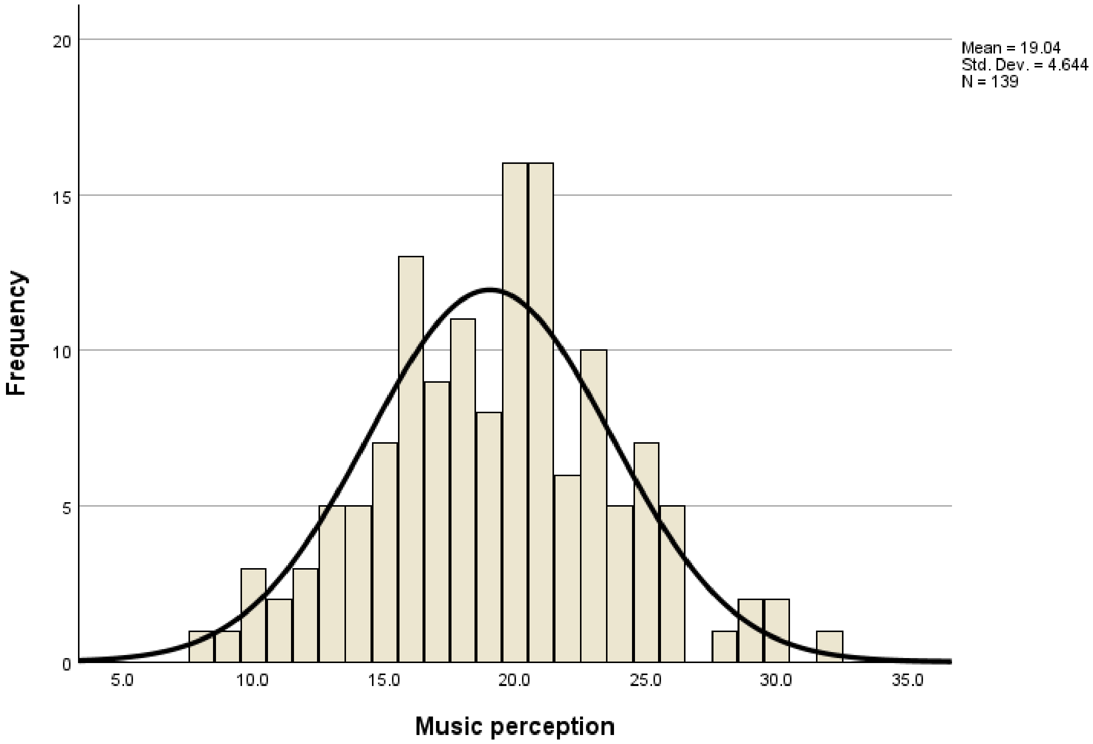
Consent Form

Statement	Please initial each box if you agree with the statement
I confirm that I have read and understand the information sheet for the above research. I have had the opportunity to consider the information, ask questions and have had these answered satisfactorily.	
I understand that my participation is voluntary and that I am free to withdraw at any point until 1st/June/2024, without giving any reason.	
I understand who will have access to the personal data provided, how the data will be stored and what will happen to the data at the end of the project.	
I understand that I will not be identifiable from any publications or any research outputs (e.g., reports for specific organisations, presentations, videos, websites).	
Use of quotations: Please indicate your preference (select *one* option):	
a) I do not wish to be quoted. **or**	
b) I agree to the use of quotations in research outputs if I am not identifiable. **or**	
c) I agree to the use of direct quotations, attributed to my name, in research outputs.	
I give permission for you to contact me again to clarify information.	
I understand how to raise a concern or make a complaint.	
I agree to take part.	

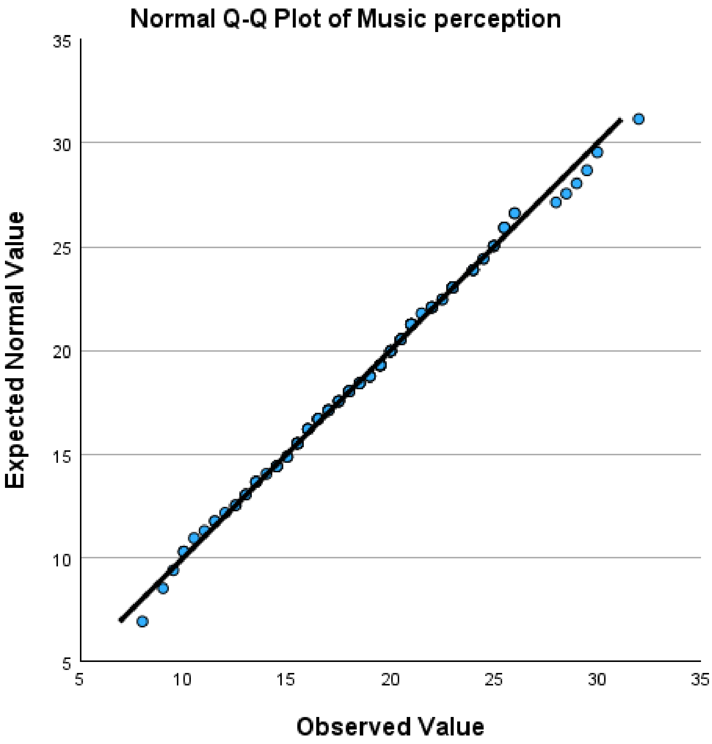
Appendix G

Normality Check for Music Perception (MP), L2 Silent Reading Comprehension (RC), Phonological Awareness (PA), and Auditory Working Memory (AWM) Tests

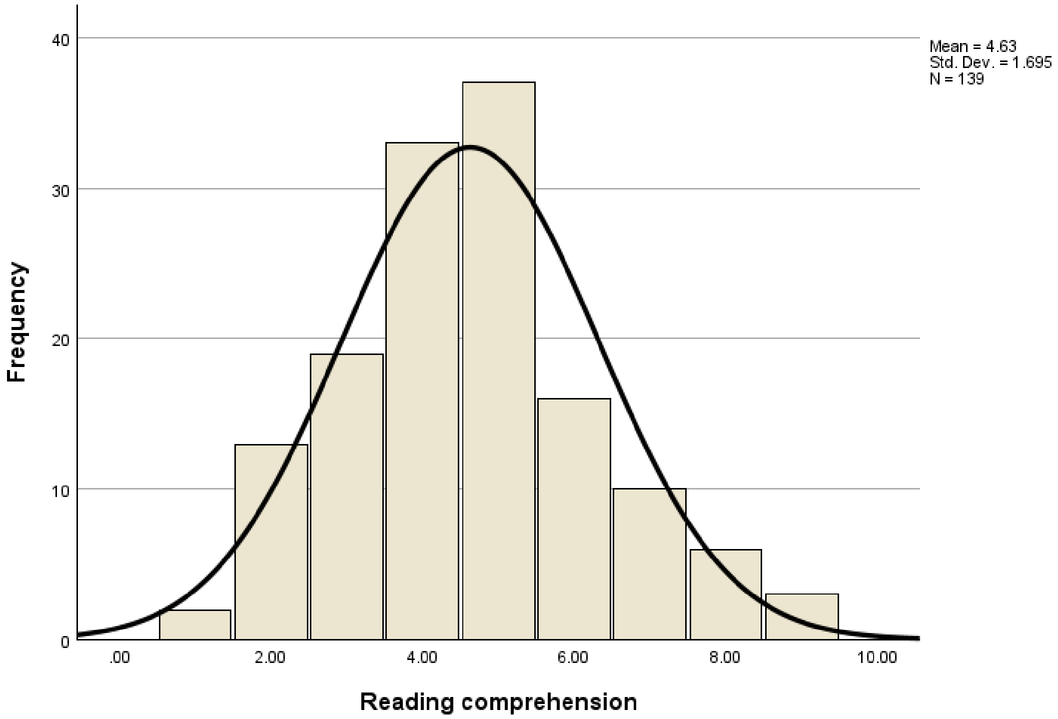
E.1 Histogram of participants' total scores on the MP test



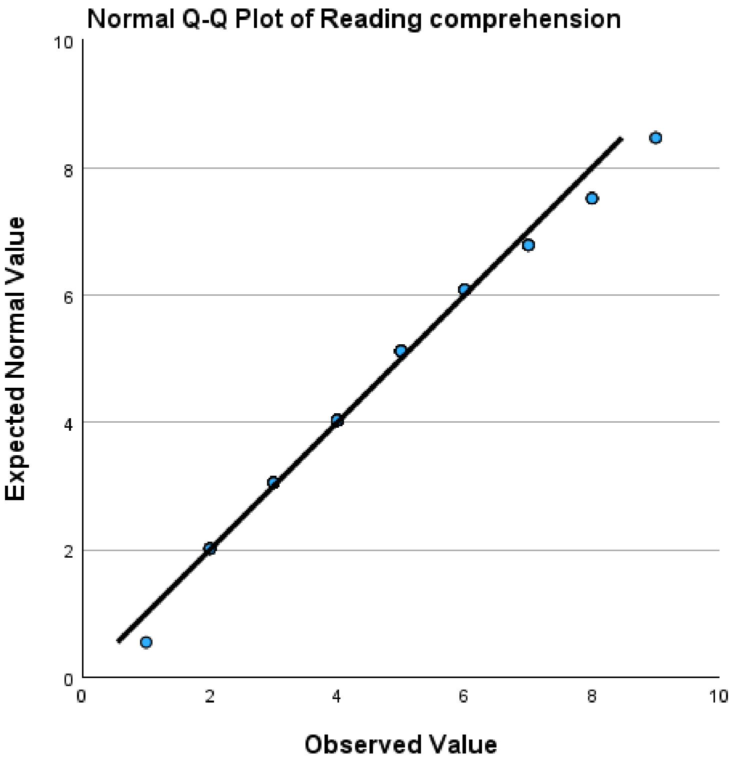
E.2 Q-Q plot of participants' total scores on the MP test



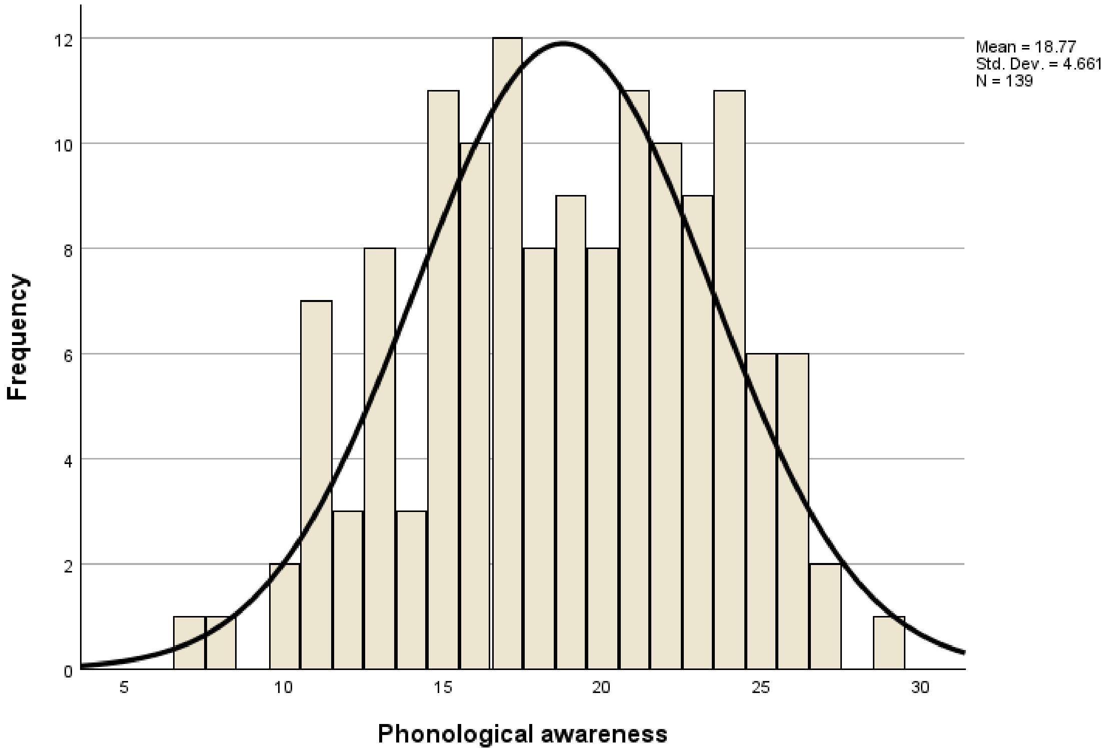
E.3 Histogram of participants' total scores on the RC test



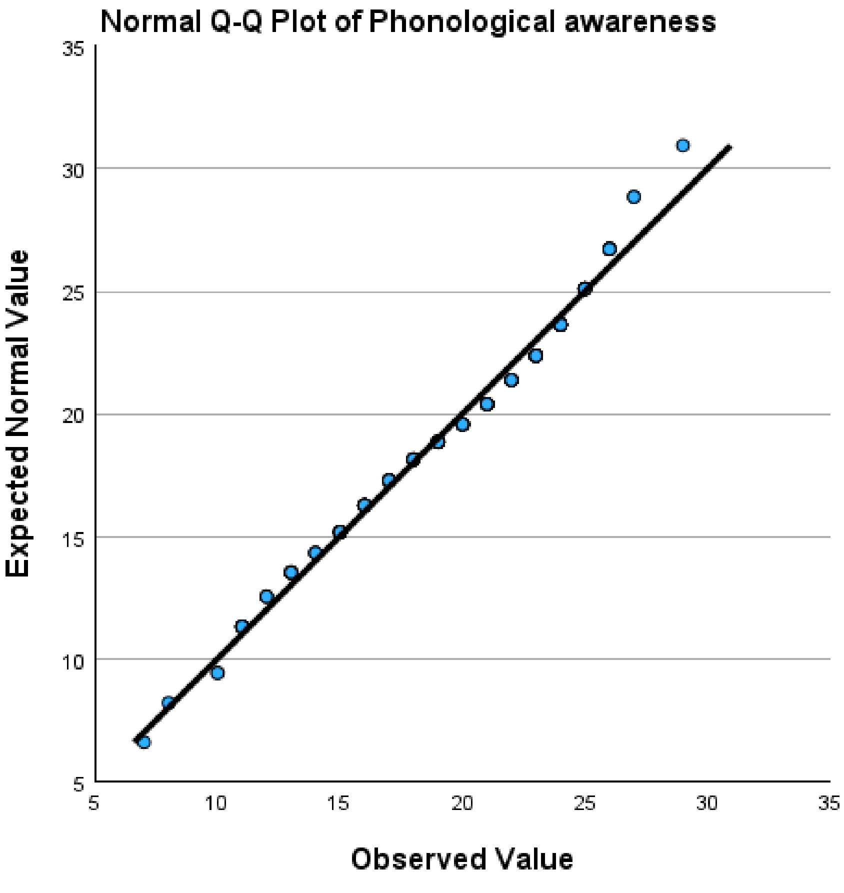
E.4 Q-Q plot of participants' total scores on the RC test



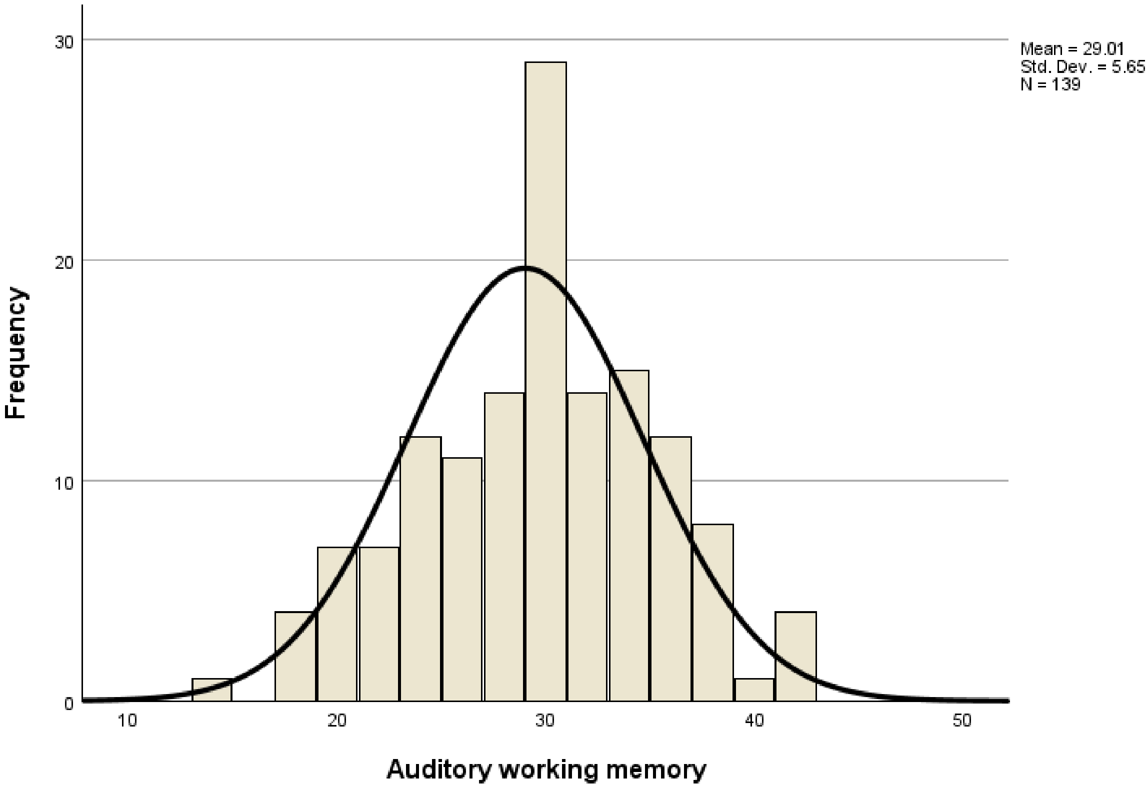
E.5 Histogram of participants' total scores on the PA test



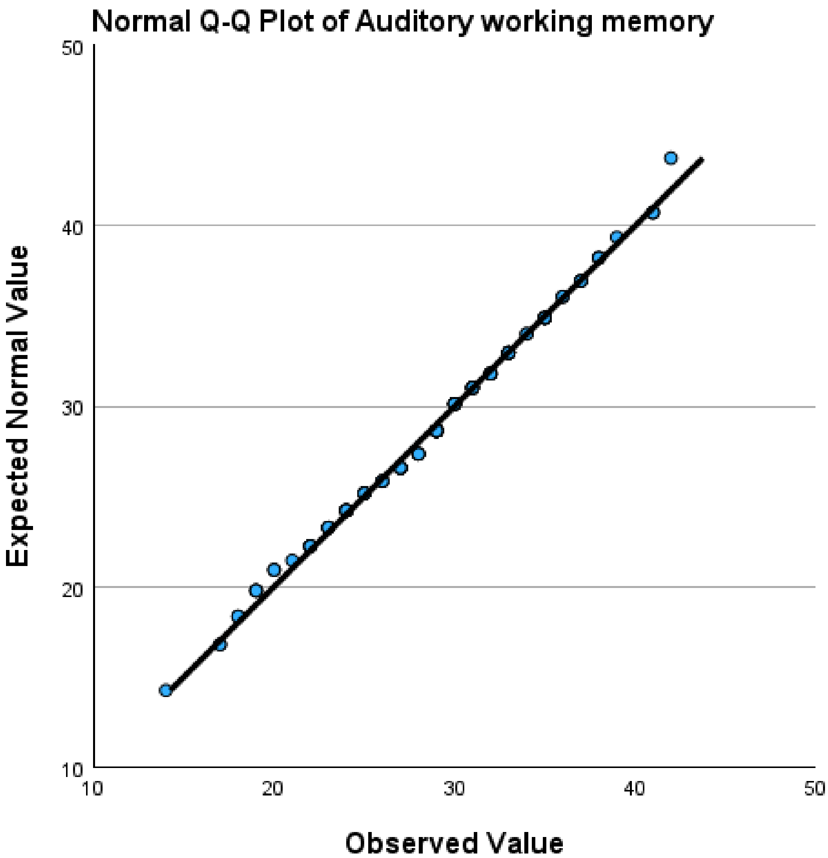
E.6 Q-Q plot of participants' total scores on the PA test



E.7 Histogram of participants' total scores on the AWM test



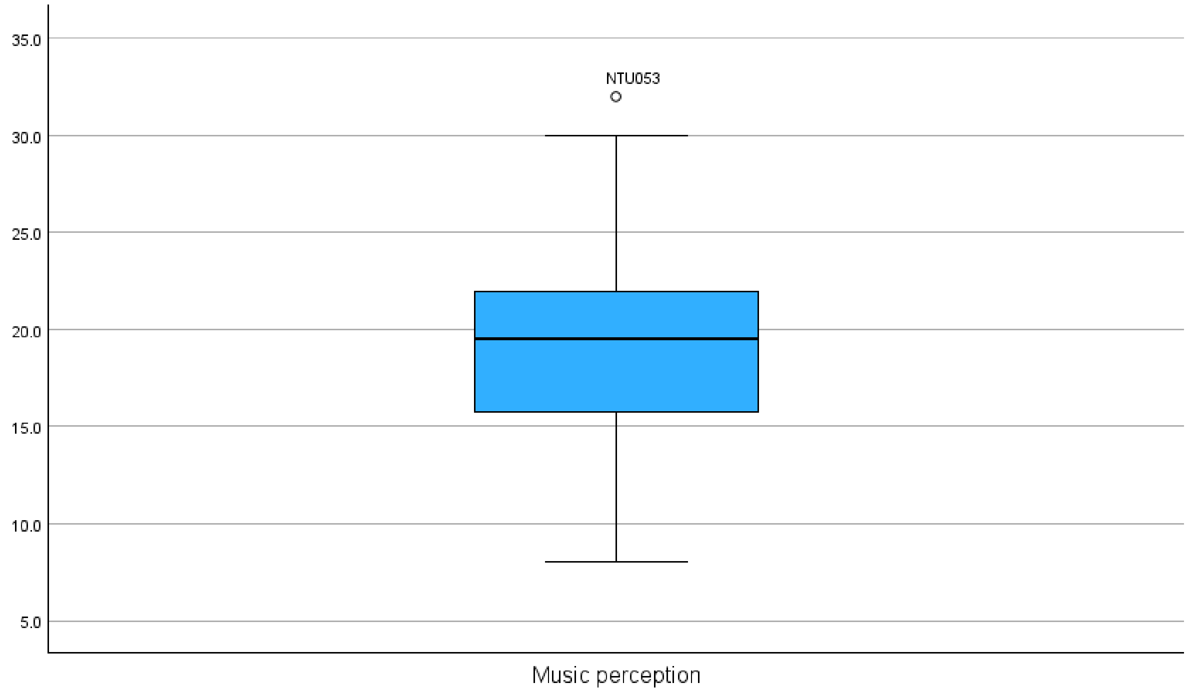
E.8 Q-Q plot of participants' total scores on the AWM test



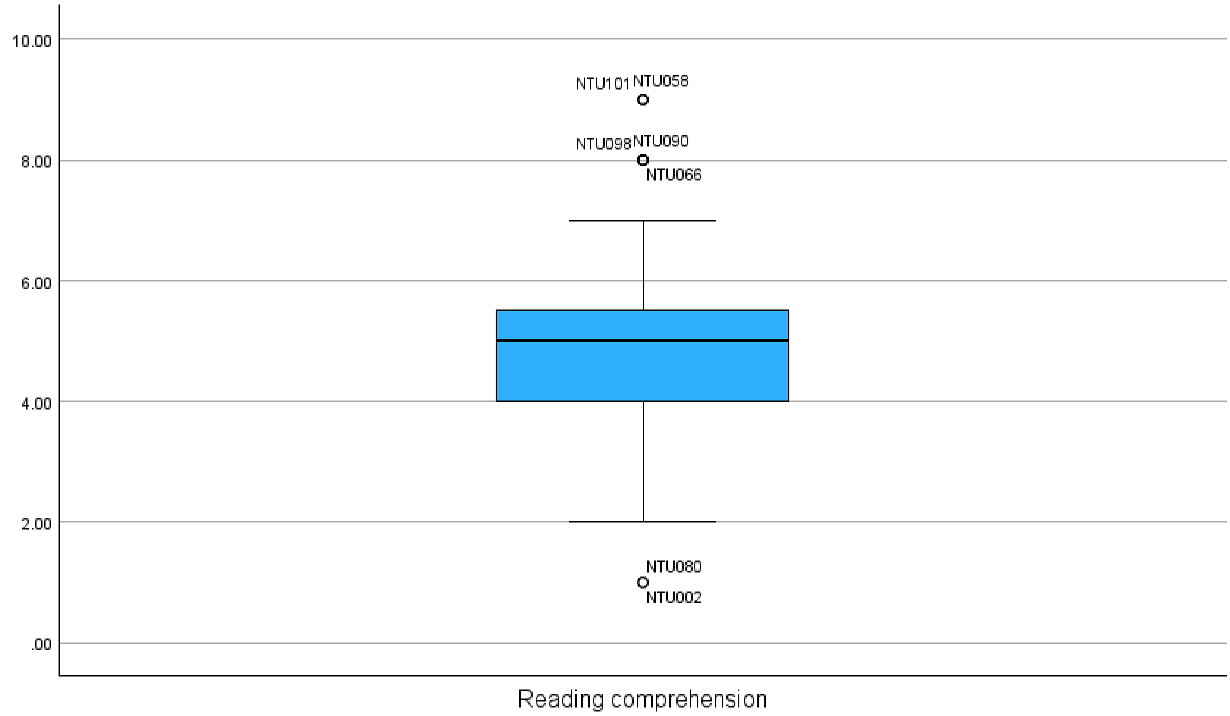
Appendix H

Outliers

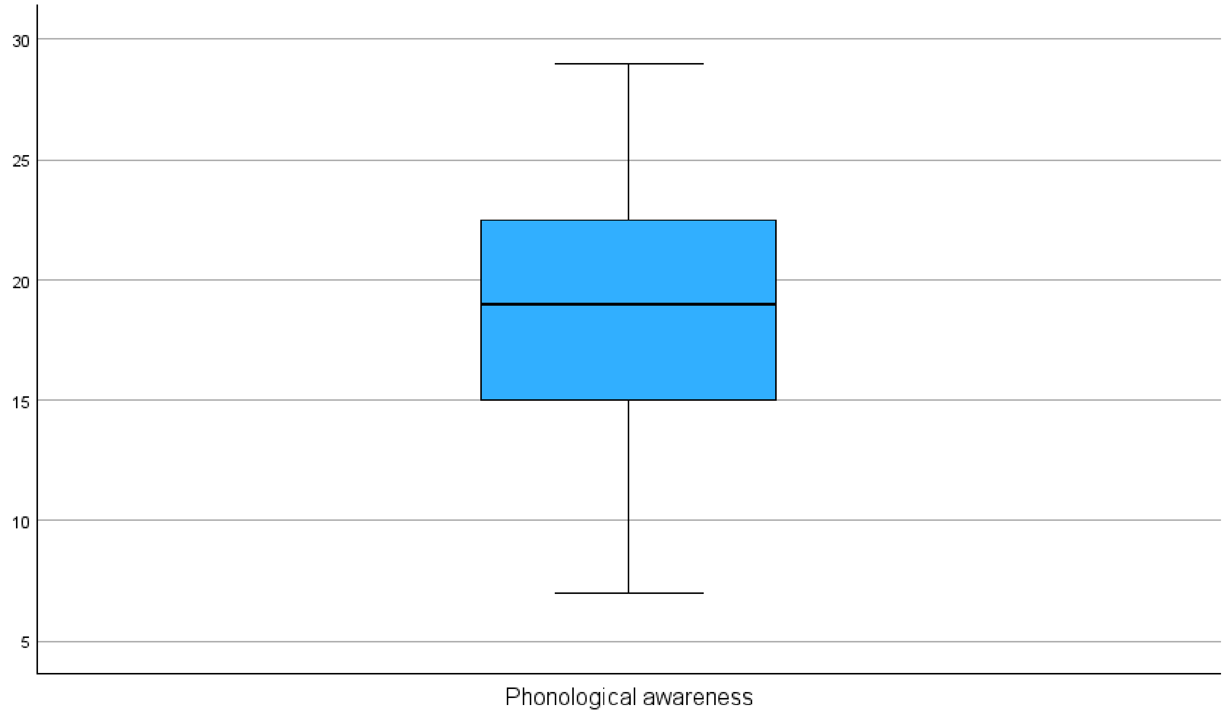
F.1 Boxplot of MP test



F.2 Boxplot of RC test



F.3 Boxplot of PA test



F.4 Boxplot of AWM test

Appendix I

13*13 Correlation Matrix of MP, RC, PA, and AWM Subtests

Pearson Correlation	Melody discrimination	Tunning discrimination	Accent discrimination	Tempo discrimination	Detailed understanding	Mean idea summary	Rhyming perception	Syllable awareness	Sound segmentation	Word parts blending	Digit span forward	Digit span backward	Letter number sequencing
Melody discrimination	1												
Tunning discrimination	.267**	1											
Accent discrimination	.344**	.325**	1										
Tempo discrimination	.281**	.230**	.444**	1									
Detailed understanding	.178*	.140	.190*	.242**	1								
Mean idea summary	.018	.167*	.069	.132	.155	1							
Rhyming perception	.035	.007	.015	-.038	.042	.073	1						
Syllable awareness	.101	.145	.045	-.073	-.085	.151	.191*	1					
Sound segmentation	.186*	.049	.093	.067	.157	.297**	.114	.202*	1				
Word parts blending	.162	.122	.215*	.151	.116	.247**	.174*	.278**	.337**	1			
Digit span forward	.136	.029	.045	.140	.205*	.282**	.238**	.103	.275**	.255**	1		
Digit span backward	.139	.043	.078	.131	.208*	.278**	.144	.153	.271**	.260**	.522**	1	
Letter number sequencing	.087	.176*	.126	.146	.104	.376**	.024	.090	.165	.143	.484**	.388**	1

**. Correlation is significant at the 0.01 level (2-tailed).
*. Correlation is significant at the 0.05 level (2-tailed).

Appendix J

MANOVA Assumptions Met

J.1 Multivariate normality check

	Training			Statistic	Std. Error
Music perception	advanced	Mean		20.194	.7247
		95% Confidence Interval for Mean	Lower Bound	18.737	
			Upper Bound	21.651	
		5% Trimmed Mean		20.193	
		Median		20.000	
		Variance		25.738	
		Std. Deviation		5.0732	
		Minimum		9.0	
		Maximum		32.0	
		Range		23.0	
		Interquartile Range		7.8	
		Skewness		.009	.340
		Kurtosis		-.214	.668
	basic	Mean		18.667	.7099
		95% Confidence Interval for Mean	Lower Bound	17.233	
			Upper Bound	20.100	
		5% Trimmed Mean		18.578	
		Median		18.500	
		Variance		21.167	
		Std. Deviation		4.6007	
		Minimum		9.5	
		Maximum		29.5	
		Range		20.0	
		Interquartile Range		5.5	
		Skewness		.207	.365
		Kurtosis		.073	.717
	untrained	Mean		18.177	.5828
		95% Confidence Interval for Mean	Lower Bound	17.005	
			Upper Bound	19.350	
		5% Trimmed Mean		18.266	
		Median		19.000	
		Variance		16.303	
		Std. Deviation		4.0377	
		Minimum		8.0	
		Maximum		25.5	
		Range		17.5	
		Interquartile Range		5.5	
		Skewness		-.339	.343
		Kurtosis		-.329	.674
Reading comprehension	advanced	Mean		4.776	.2547
		95% Confidence Interval for Mean	Lower Bound	4.263	
			Upper Bound	5.288	
		5% Trimmed Mean		4.751	
		Median		5.000	
		Variance		3.178	
		Std. Deviation		1.7826	
		Minimum		1.0	
		Maximum		9.0	
		Range		8.0	
		Interquartile Range		2.0	
		Skewness		.216	.340
		Kurtosis		-.019	.668
	basic	Mean		4.762	.2256
		95% Confidence Interval for Mean	Lower Bound	4.306	
			Upper Bound	5.217	
		5% Trimmed Mean		4.735	
		Median		5.000	
		Variance		2.137	
		Std. Deviation		1.4619	
		Minimum		2.0	

		Maximum		8.0	
		Range		6.0	
		Interquartile Range		2.0	
		Skewness		.337	.365
		Kurtosis		-.690	.717
	untrained	Mean		4.354	.2587
		95% Confidence Interval for Mean	Lower Bound	3.834	
			Upper Bound	4.875	
		5% Trimmed Mean		4.259	
		Median		4.000	
		Variance		3.212	
		Std. Deviation		1.7923	
		Minimum		1.0	
		Maximum		9.0	
		Range		8.0	
		Interquartile Range		2.0	
		Skewness		.619	.343
		Kurtosis		.555	.674
Phonological awareness	advanced	Mean		19.47	.670
		95% Confidence Interval for Mean	Lower Bound	18.12	
			Upper Bound	20.82	
		5% Trimmed Mean		19.62	
		Median		20.00	
		Variance		22.004	
		Std. Deviation		4.691	
		Minimum		10	
		Maximum		26	
		Range		16	
		Interquartile Range		8	
		Skewness		-.327	.340
		Kurtosis		-.894	.668
	basic	Mean		19.74	.602
		95% Confidence Interval for Mean	Lower Bound	18.52	
			Upper Bound	20.95	
		5% Trimmed Mean		19.71	
		Median		20.00	
		Variance		15.222	
		Std. Deviation		3.902	
		Minimum		13	
		Maximum		27	
		Range		14	
		Interquartile Range		6	
		Skewness		.023	.365
		Kurtosis		-.802	.717
	untrained	Mean		17.21	.710
		95% Confidence Interval for Mean	Lower Bound	15.78	
			Upper Bound	18.64	
		5% Trimmed Mean		17.20	
		Median		17.00	
		Variance		24.211	
		Std. Deviation		4.920	
		Minimum		7	
		Maximum		29	
		Range		22	
		Interquartile Range		9	
		Skewness		.072	.343
		Kurtosis		-.584	.674
Auditory working memory	advanced	Mean		28.22	.690
		95% Confidence Interval for Mean	Lower Bound	26.84	
			Upper Bound	29.61	
		5% Trimmed Mean		28.36	
		Median		29.00	
		Variance		23.303	
		Std. Deviation		4.827	
		Minimum		18	
		Maximum		36	
		Range		18	

	Interquartile Range		7	
	Skewness		-.418	.340
	Kurtosis		-.645	.668
basic	Mean		30.05	.802
	95% Confidence Interval for Mean	Lower Bound	28.43	
		Upper Bound	31.67	
	5% Trimmed Mean		30.11	
	Median		29.50	
	Variance		27.022	
	Std. Deviation		5.198	
	Minimum		19	
	Maximum		39	
	Range		20	
	Interquartile Range		8	
	Skewness		-.072	.365
	Kurtosis		-.879	.717
untrained	Mean		28.90	.965
	95% Confidence Interval for Mean	Lower Bound	26.95	
		Upper Bound	30.84	
	5% Trimmed Mean		28.93	
	Median		29.00	
	Variance		44.734	
	Std. Deviation		6.688	
	Minimum		14	
	Maximum		42	
	Range		28	
	Interquartile Range		9	
	Skewness		-.066	.343
	Kurtosis		-.258	.674

The z-scores for skewness and kurtosis of the four variables within each of the three musical training background groups (advanced, basic, untrained) – totaling 12 values – fall within the range of ±1.96, meeting the assumption of multivariate normality.

J.2 Check for homogeneity of covariance matrices

Box's Test of Equality of Covariance Matrices[a]	
Box's M	21.074
F	1.009
df1	20
df2	63861.636
Sig.	.447

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.

a. Design: Intercept + Training

Appendix K

Follow-up MANOVA on PA subtests

K.1 Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Intercept	Pillai's Trace	.968	992.268b	4.000	133.000	<.001	.968
	Wilks' Lambda	.032	992.268b	4.000	133.000	<.001	.968
	Hotelling's Trace	29.843	992.268b	4.000	133.000	<.001	.968
	Roy's Largest Root	29.843	992.268b	4.000	133.000	<.001	.968
Training	Pillai's Trace	.197	3.659	8.000	268.000	<.001	.098
	Wilks' Lambda	.812	3.639b	8.000	266.000	<.001	.099
	Hotelling's Trace	.219	3.618	8.000	264.000	<.001	.099
	Roy's Largest Root	.131	4.374c	4.000	134.000	.002	.115

a. Design: Intercept + Training

b. Exact statistic

c. The statistic is an upper bound on F that yields a lower bound on the significance level.

K.2 Tests of Between-Subjects Effects

Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	Rhyming perception	3.407a	2	1.703	1.220	.298	.018
	Syllable awareness	4.972b	2	2.486	.773	.464	.011
	Sound segmentation	44.523c	2	22.262	4.454	.013	.061
	Word parts blending	53.188d	2	26.594	8.778	<.001	.114
Intercept	Rhyming perception	4845.402	1	4845.402	3470.594	<.001	.962
	Syllable awareness	3032.817	1	3032.817	942.861	<.001	.874
	Sound segmentation	2924.915	1	2924.915	585.235	<.001	.811
	Word parts blending	1800.459	1	1800.459	594.255	<.001	.814
Training	Rhyming perception	3.407	2	1.703	1.220	.298	.018
	Syllable awareness	4.972	2	2.486	.773	.464	.011
	Sound segmentation	44.523	2	22.262	4.454	.013	.061
	Word parts blending	53.188	2	26.594	8.778	<.001	.114
Error	Rhyming perception	189.874	136	1.396			
	Syllable awareness	437.459	136	3.217			
	Sound segmentation	679.707	136	4.998			
	Word parts blending	412.049	136	3.030			
Total	Rhyming perception	5078.000	139				
	Syllable awareness	3482.000	139				
	Sound segmentation	3616.000	139				
	Word parts blending	2271.000	139				
Corrected Total	Rhyming perception	193.281	138				
	Syllable awareness	442.432	138				
	Sound segmentation	724.230	138				
	Word parts blending	465.237	138				

a. R Squared = .018 (Adjusted R Squared = .003)

b. R Squared = .011 (Adjusted R Squared = -.003)

c. R Squared = .061 (Adjusted R Squared = .048)

d. R Squared = .114 (Adjusted R Squared = .101)

K.3 Multiple Comparisons

Bonferroni

Dependent Variable	(I) Training	(J) Training	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Rhyming perception	advanced	basic	.39	.248	.363	-.21	.99
		untrained	.16	.240	1.000	-.42	.75
	basic	advanced	-.39	.248	.363	-.99	.21
		untrained	-.22	.250	1.000	-.83	.38
	untrained	advanced	-.16	.240	1.000	-.75	.42
		basic	.22	.250	1.000	-.38	.83
Syllable awareness	advanced	basic	-.04	.377	1.000	-.95	.88
		untrained	.38	.364	.899	-.50	1.26
	basic	advanced	.04	.377	1.000	-.88	.95
		untrained	.42	.379	.820	-.50	1.34
	untrained	advanced	-.38	.364	.899	-1.26	.50
		basic	-.42	.379	.820	-1.34	.50
Sound segmentation	advanced	basic	-1.08	.470	.070	-2.22	.06
		untrained	.26	.454	1.000	-.84	1.36
	basic	advanced	1.08	.470	.070	-.06	2.22
		untrained	1.34*	.472	.016	.20	2.49
	untrained	advanced	-.26	.454	1.000	-1.36	.84
		basic	-1.34*	.472	.016	-2.49	-.20
Word parts blending	advanced	basic	.46	.366	.635	-.43	1.35
		untrained	1.45*	.353	<.001	.60	2.31
	basic	advanced	-.46	.366	.635	-1.35	.43
		untrained	.99*	.368	.023	.10	1.89
	untrained	advanced	-1.45*	.353	<.001	-2.31	-.60
		basic	-.99*	.368	.023	-1.89	-.10

Based on observed means.

The error term is Mean Square (Error) = 3.030.

*. The mean difference is significant at the .05 level.