LEABHARLANN CHOLÁISTE NA TRÍONÓIDE, BAILE ÁTHA CLIATH | TRINITY COLLEGE LIBRARY DUBLIN
Ollscoil Átha Cliath | The University of Dublin

Terms and Conditions of Use of Digitised Theses from Trinity College Library Dublin

Copyright statement

All material supplied by Trinity College Library is protected by copyright (under the Copyright and Related Rights Act, 2000 as amended) and other relevant Intellectual Property Rights. By accessing and using a Digitised Thesis from Trinity College Library you acknowledge that all Intellectual Property Rights in any Works supplied are the sole and exclusive property of the copyright and/or other IPR holder. Specific copyright holders may not be explicitly identified. Use of materials from other sources within a thesis should not be construed as a claim over them.

A non-exclusive, non-transferable licence is hereby granted to those using or reproducing, in whole or in part, the material for valid purposes, providing the copyright owners are acknowledged using the normal conventions. Where specific permission to use material is required, this is identified and such permission must be sought from the copyright holder or agency cited.

Liability statement

By using a Digitised Thesis, I accept that Trinity College Dublin bears no legal responsibility for the accuracy, legality or comprehensiveness of materials contained within the thesis, and that Trinity College Dublin accepts no liability for indirect, consequential, or incidental, damages or losses arising from use of the thesis for whatever reason. Information located in a thesis may be subject to specific use constraints, details of which may not be explicitly described. It is the responsibility of potential and actual users to be aware of such constraints and to abide by them. By making use of material from a digitised thesis, you accept these copyright and disclaimer provisions. Where it is brought to the attention of Trinity College Library that there may be a breach of copyright or other restraint, it is the policy to withdraw or take down access to a thesis while the issue is being resolved.

Access Agreement

By using a Digitised Thesis from Trinity College Library you are bound by the following Terms & Conditions. Please read them carefully.

I have read and I understand the following statement: All material supplied via a Digitised Thesis from Trinity College Library is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of a thesis is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form providing the copyright owners are acknowledged using the normal conventions. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

An Empirical Investigation of the Association between Musical Aptitude and Foreign Language Aptitude

Lorraine Frances Gilleece

Thesis submitted for the Degree of Doctor of Philosophy

Centre for Language and Communication Studies

University of Dublin

Trinity College

2006

TRINITY COLLEGE
15 MAY 2006
LIBRARY DUBLIN
THESIS
7803

The man that hath no music in himself,
Nor is not moved with concord of sweet sounds,
Is fit for treasons, stratagems and spoils;
The motions of his spirit are dull as night
And his affections dark as Erebus:
Let no such man be trusted. Mark the music.

William Shakespeare (1564 – 1616)

Declaration

I hereby declare that this thesis, submitted in candidature for the degree of Doctor of Philosophy at Trinity College Dublin, is entirely my own work and has not been previously submitted for a degree at this or any other university. I agree that the Library may lend or copy the thesis upon request.

Lorraine Gilleece

Lorraine Gilleece, 30th September 2005

Abstract

Given the joint ubiquity of music and language, and pre-theoretic similarities between the two, it is relevant to consider the relationship between musical ability and linguistic ability, specifically in relation to second language acquisition. The specific question of musical aptitude and its relationship to foreign language aptitude is the central focus of this thesis. In Chapter 2, the dissertation reviews classical and recent research on individual differences, in particular those individual differences which are known to have a major impact on the second language learning process. This review reveals a complex relationship between language aptitude, intelligence and working memory. Chapter 3 examines the relationship between music and other cognitive abilities, focussing primarily on the relationship between music and language ability. This suggests that further analysis of the music-language relationship is indeed justified. From the analysis of past research reported in Chapters 2 and 3, open questions emerge about two important issues: the extent to which music and language aptitude are related, and the extent to which that relationship is mediated by general intelligence. Empirical investigations are carried out to investigate these issues quantitatively and qualitatively.

Phase I of the study examines receptive aptitude in music and language. A sample of 149 subjects completes the Bentley Measures of Musical Aptitude and a language aptitude test based on the Modern Language Aptitude Test, in addition to the Raven Progressive Matrices Test. Phase II of the study examines productive aptitude in music and language. Forty one of the original subjects imitate foreign language words and sentences and short rhythm patterns. Results of both experiments reveal a significant relationship between music and language aptitude, independent of general intelligence. Correlations may be considered small to moderate.

Acknowledgments

I would like to thank my supervisor Professor David Singleton for all his encouragement and kind words throughout my postgraduate studies. I would also like to express my deep gratitude to Dr. Carl Vogel for his assistance throughout my time in Trinity, particularly for taking the time to offer advice and suggestions on this thesis. I am very grateful to Marie Moran in the Royal Irish Academy of Music for her assistance, especially for so willingly spending a weekend correcting all the productive music tests. I cannot over emphasise how much this is appreciated. Dr. Cathal Walsh in the Department of Statistics was also very generous with his time and I deeply appreciate the prompt replies to my numerous emails.

A very big thank you must go to all the pupils and staff of the schools where I carried out my field work, particularly to the following teachers and principals: Mr. Johnson, St. Benildus College, Stillorgan; Ms. Eileen O'Donnell, St. Raphaela's Secondary School, Stillorgan; Mr. John Gildea, St. Bricin's Vocational School, Belturbet; Mr. Ken Duggan, CBS Westland Row; Mr. Sean Casey, Dundrum College and Mr. Brendan Doyle, St. Conleth's College, Ballsbridge. I would also like to thank the language schools who recruited native speakers of various foreign languages for me – Alpha College, North Great George's St.; Mr. Noel Gardner, Centre for English Studies, Dame St. and Mr. Ciaran McCarthy, EF International Language School, Fitzwilliam Square.

Thank you very much to the native speakers of Chinese, Czech, Korean and Spanish who helped me prepare and correct the language aptitude tests. Thanks to Benji for help with Chinese, also to Nam, Ko and Namhee, and Namsook in the Korean Embassy, for help with Korean, Lucie, and also Petra in the Czech Embassy, for help with Czech, and finally, Susana, Nuria and Alma for help with Spanish. I also appreciate the helpful comments of Jeong-Young Kim, University of Helsinki, on the Korean stimuli used in Phase II.

Thanks to Martha Dalton for teaching me how to use the Praat software and for answering any phonetics questions. Thank you also to Dr. Ann Devitt and Brian

Murphy in the Computational Linguistics Group for help with Perl and general computer queries! I also appreciate the help of Denise O'Leary in reading earlier drafts of this thesis. Thanks to Drs. Colm O'Reilly and Sheila Gilheany in the Centre for Talented Youth, DCU for providing me with the Raven Progressive Matrices Test booklets.

This research was made possible through various sources of funding. I am very grateful for the financial support of Trinity College, from whom I have received the following funding: Foundation Scholarship, Trinity College Postgraduate Award, Trinity Trust Travel Grant. I also appreciate being awarded the Dr. A. J. Leventhal Scholarship. Finally, I would like to acknowledge receipt of a postgraduate scholarship from the Irish Research Council for the Humanities and Social Sciences.

Finally and most especially, I must thank my family. Thank you very much to mum and dad for everything over the years – from finding digs in first year, to advice on buying a lap-top, from frozen dinners to advice on teaching, I am very grateful for everything. Thank you Richard for all your help – posters in Powerpoint, statistics, neuroanatomy, …., but more than anything, for chocolate and ice-cream at times when the relationship between music and language seemed particularly elusive!

Table of Contents

Declaration ..... *i*
Abstract ..... *ii*
Acknowledgments ..... *iii*
Table of Contents ..... *v*
List of Figures ..... *x*
List of Tables ..... *xii*
List of Equations ..... *xiii*

Chapter 1 Introduction **1**
Chapter 2 Individual Differences **19**
2.1 Introduction 19
2.2 Individual Differences in Psychological Research 20
2.3 Individual Differences in First Language Acquisition 26
2.4 Differences between Second Language Learners 28
2.5 General Aptitude 33
2.5.1 Stability of Aptitude 33
2.5.2 Components of Aptitude 34
2.5.3 Other Approaches to Aptitude 38
2.5.4 Aptitude and Age 40
2.5.5 Language Aptitude Tests 41
2.5.6 Relationship between Language Aptitude and Other Cognitive Skills 43
2.5.7 Importance of Aptitude in Different Contexts 45
2.5.8 Conclusion 46
2.6 Intelligence 48
2.6.1 Introduction 48
2.6.2 Early Measurements of Intelligence 53
2.6.3 Current Views on Intelligence 57
2.6.4 Conclusion 60
2.7 Working Memory 61
2.7.1 Introduction 62
2.7.2 Phonological Loop and Language Acquisition 65
2.7.2.1 Capacity of Phonological Loop 66
2.7.2.2 Phonological Store or Rehearsal System 68
2.7.3 Working Memory & Intelligence 71
2.7.4 Other Working Memory Components 75
2.7.5 Summary of Discussion of Working Memory 79
2.8 Conclusion 80
Chapter 3 Music, Language and the Brain **82**
3.1 Introduction 82
3.2 Music and Spatial Reasoning 83
3.2.1 Explaining the "Mozart Effect" 84
3.2.2 Replication of the "Mozart Effect" Experiments 86
3.2.3 Extension of Findings 88
3.2.4 Conclusion 90
3.3 Music and Linguistic Intelligence 90
3.3.1 Music and First Language Acquisition 92
3.3.2 Music and Second Language Acquisition 94
3.3.3 Other Recent Research 103

3.4 The Brain, Music and Cognition 109
3.4.1 Introduction 109
3.4.2 Anatomical Differences Arising From Musical Training 111
3.4.4 Music and Language in the Brain 112
3.5 Conclusion 118
Chapter 4 Methodology 120
4.1 Introduction 120
4.2 Phase I - Receptive Skills 121
4.2.1 Introduction 121
4.2.2 Design of Test Battery 122
4.2.2.1 Music 122
4.2.2.2 Language 124
4.2.2.3 Non-verbal Intelligence (gF) 127
4.2.3 Scoring of Tests 128
4.2.4 Subjects 128
4.2.5 Administration of Tests 130
4.2.6 Analysis 131
4.3. Phase II - Productive Skills 132
4.3.1 Introduction 132
4.3.2 Design of Test Battery 135
4.3.2.1 Music 135
4.3.2.2 Language 136
4.3.3 Scoring of Tests 139
4.3.3.1 Music 139
4.3.3.2 Language 141
4.3.4 Subjects 145
4.3.5 Administration of Tests 146
4.3.6 Analysis 146
4.4 Phase III –Receptive & Productive 147
4.5 Reliability 147
4.5.1 Inter Rater Reliability 148
4.5.1.1 Korean 148
4.5.1.2 Spanish 150
4.5.1.3 Kappa Statistic 152
4.5.1.4 Conclusion 153
4.5.2 Intra Rater Reliability 153
4.5.2.1 Korean 153
4.5.2.2 Spanish 155
4.5.2.3 Kappa Statistic 156
4.5.2.4 Conclusion 158
4.5.3 Reliability of Scale Analysis 158
4.5.3.1 Korean 159
4.5.3.2 Spanish 159
4.5.3.3 Music (Prod) 159
4.5.3.4 Summary 159
4.6 Conclusion 159
Chapter 5 Results and Discussion 161
5.1 Introduction 161
5.2 Phase I – Examination of Receptive Skills in Music and Language 161
5.2.1 Distribution of scores 161

5.2.2 Correlations 166
5.2.2.1 Correlations between components of tests 168
5.2.2.2 Correlations between tests 171
5.2.3 Subjects divided by Raven scores 179
5.2.3.1 Correlations within high and low Raven groups 186
5.2.3.2 Summary and conclusion of Raven groups 187
5.2.4 Subjects divided by gender 188
5.2.5 Partial Correlation 191
5.2.6 Regression 192
5.2.6.1 Introduction 192
5.2.6.2 Multiple Regression 195
5.3 Phase II – Examination of Productive Skills 200
5.3.1 Subjects 201
5.3.2 Distribution of scores 201
5.3.3 Correlations 203
5.3.3.1 Correlations between components of Language test 204
5.3.3.2 Correlations between tests 204
5.3.4 Subjects divided by Raven scores 208
5.3.4.1 Correlations within high and low Raven groups 211
5.3.4.2 Summary of Low/High Raven split for Phase II 213
5.3.5 Subjects divided by gender 214
5.3.6 Partial Correlations 215
5.3.7 Regression 215
5.3.8 Comparison of PVI measures with Native Speaker scores 217
5.3.8.1 Why the differences? 219
5.3.8.2 Why values are not near English or Spanish values? 221
5.3.8.3 Comparison with Native Speaker Scores for Rhythm 221
5.3.8.4 PVI scores vs. ns scores 223
5.3.9 Conclusion 224
5.4 Phase III: Links between Receptive and Productive Skills 224
5.4.1 Correlations 225
5.4.1.1 Correlations Between Tests 225
5.4.2 Subjects divided by Raven scores 228
5.4.2.1 Correlations within high and low Raven groups 228
5.5 Multiple Testing 229
5.6 Korean and Spanish as tests of Language Aptitude 230
5.7 Conclusion 235
Chapter 6 Conclusion 236
6.1 Introduction 236
6.2 Summary of research 236
6.3 New approaches and findings 237
6.4 Relationship to existing literature 240
6.5 Limitations of this study 241
6.6 Suggestions for future work 241
6.7 Summary 243
Bibliography 244
Appendix 1 275
Correlations 275
Pearson Correlations 275
Spearman's Rho 276

Groups divided by Raven scores 277
Groups divided by gender 278
Regression 280
Outliers omitted 280
Gender as predictor 282
Stepwise regression 283
Appendix 2 285
Normal Q-Q Plots 285
Correlations 288
Pearson Correlations 288
Spearman's Rho 290
Group divided by Raven scores 292
Regression 293
Outliers omitted 294
Stepwise regression 295
Appendix 3 297
Correlations 297
Pearson Correlations 297
Spearman's Rho 298
Group divided by Raven scores 299
Appendix 4 300
Bentley Measures of Musical Aptitude 300
Test items 300
Answer sheet 304
Transcript of Instructions 305
Receptive Language Aptitude Test 308
Test items 308
Answer sheet 314
Transcript of Instructions 318
Appendix 5 320
Music productive test 320
Spanish sentences 322
Korean Words 322
Appendix 6 324
Korean rater instructions 324
Spanish rater instructions 325
Appendix 7 326
Calculation of Kappa statistic for Korean raters 326
Korean rater 1 326
Algorithm 326
Items examined 329
Output of algorithm 329
Korean rater 2 330
Algorithm 330
Items examined 334
Output 335
Spanish rater 1 335
Items examined 335
Output 336
Spanish rater 2 337

Algorithm 337
Items examined 340
Output 341
Appendix 8 344
Appendix 9 347

List of Figures

Figure 1. Tree diagram for sentence of English 10
Figure 2. Tree diagram for musical phrase 11
Figure 3. Syntactic structure in music 12
Figure 4. Gender and origin of subjects in Phase I 130
Figure 5. Screenshot from Anvil Studio 136
Figure 6. Utterance segmented in Praat 144
Figure 7. Subjects participating in Phase II 145
Figure 8. Boxplot of Korean scores 149
Figure 9. Korean raters' mean scores for each subject 149
Figure 10. Boxplot of Spanish scores 150
Figure 11. Spanish raters' mean scores for each subject 151
Figure 12. Spanish raters' mean scores for each sentence 152
Figure 13. Bubbleplot for scores of Korean Rater 1 153
Figure 14. Bubbleplot for scores of Korean Rater 2 154
Figure 15. Bubbleplot for scores of Spanish Rater 1 155
Figure 16. Bubbleplot for scores of Spanish Rater 2 156
Figure 17. Distribution of Music (Recep) 162
Figure 18. Distribution of Language (Recep) 163
Figure 19. Distribution of Raven scores 163
Figure 20. Distribution of Raven grades 164
Figure 21. Distribution of Raven Grades in School D 165
Figure 22. Distribution of Raven Grades in School E 165
Figure 23. Association between music, language and intelligence 173
Figure 24. Connections between music, language and memory 174
Figure 25. Language (Recep) scores for different Raven categories 182
Figure 26. Music (Recep) scores for different Raven categories 183
Figure 27. Cognitive processes in music and language 185
Figure 28. Music (Recep) in Males and Females 189
Figure 29. Language (Recep) in Males and Females 189
Figure 30. Raven in Males and Females 190
Figure 31. Music (Recep) and Language (Recep) scores 194
Figure 32. Music, Language and Raven scores 195
Figure 33. Residuals vs. order of data 197
Figure 34. Residuals vs. fitted values 197
Figure 35. Normal probability plot of residuals 198
Figure 36. Cook's Distances, Phase I 199
Figure 37. Distribution of Music (Prod) scores 202
Figure 38. Distribution of Language (Prod) scores 202
Figure 39. Distribution of Spanish scores 203
Figure 40. Distribution of Korean scores 203
Figure 41. Spanish Pronunciation and Rhythm scores 204
Figure 42. Language (Prod) scores according to Raven category 209
Figure 43. Music (Prod) scores according to Raven category 210
Figure 44. A normal probability plot of residuals 215
Figure 45. Cook's Distances, Phase II 216
Figure 46. Findings of Grabe and Low (2002) 218
Figure 47. PVI values found in Phase II compared to literature 218
Figure 48. NS PVI values contrasted with NNS values 220

Figure 49. Boxplot of Spanish Pronunciation and Rhythm NS scores 222
Figure 50. PVI scores ranked according to NS scores 223
Figure 51. Scatterplot of Spanish and Korean scores..... 230
Figure 52. Language (Prod) and Language (Recep) scores 234

List of Tables

Table 1. Effects of Musical Training on Pronunciation Test Scores of 6th Graders .. 99
Table 2. Effects of Musical Training on Pronunciation Test Scores of 7th Graders .. 99
Table 3. Effects of Musical Training on Pronunciation Test Scores of 8th Graders .. 99
Table 4. Summary of Regression Models, Phase I 200

List of Equations

Equation 1. Calculation of normalised Pairwise Variability Index 143
Equation 2. Calculation of Krippendorff's alpha 156
Equation 3. Calculation of partial correlation 191

Chapter 1 Introduction

In a recent special edition of Nature Neuroscience devoted to music and the brain, editor John Spiro (2003, p.661) notes that:

> "Music, like language, is a universal feature across all human societies, both ancient and modern. And just as the ability to understand spoken language emerges effortlessly in infants, the ability to appreciate music likewise requires no explicit training."

While music and language may go hand in hand in humans, music is of course also found in species that do not possess language; e.g. while bird song is familiar to most, birds are not the only species to produce music. Cetaceans such as whales use patterns of sounds for communication. Humpback, bowhead, fin, and blue whales all produce stereotypical, patterned, and regularly repeated vocalizations of varying complexity that have been characterised as "song" (Gedamke *et al.*, 2001, p.3038). Researchers are still uncertain as to the function of whale song but suggest that as it is produced by males, primarily on breeding grounds, it is likely used as a reproductive advertisement display (see e.g. Helweg *et al.*, 1992; Watkins *et al.*, 1987).

Like human language, whale song exhibits complex organisation (see Payne & Webb, 1971). It involves units combining to form sub-phrases, which in turn combine to form phrases. The same phrase is repeated many times and this is known as a theme. A collection of themes is known as a song. Despite these complexities, whale song and other animal communication systems are generally not considered to be language, as they fail to exhibit many of the essential features of language (see Aitchison, 1998; Hockett & Altman, 1968). Hockett was one of the first to propose a number of design features which he hypothesised were the essential features of language. The number of design features which he considered important has changed over the years with his longest list consisting of sixteen features. Aitchison (1998) proposes the following list of ten design features:

- Use of vocal-auditory channel: This is not unique to humans as many animals such as cows, birds and foxes use the vocal-auditory channel in

communication. Nor is the use of the vocal-auditory channel all-important, as language can be written, or sign-language may be used, with no loss of information. Therefore, this characteristic is not very useful in distinguishing animal from human communication.

- Arbitrariness: The symbols in a human language are neutral, i.e. there is no connection between the word and what it means; e.g. a dog may be called DOG, or equally CHIEN or MADRA. While animals generally have a link between the message and the signal used to convey it, this is not always the case; e.g. gulls indicate aggression through use of an arbitrary symbol. Thus, arbitrariness is not sufficient for distinguishing between human and animal communication systems.
- Semanticity: This relates to the use of symbols to 'mean' or to refer to objects or actions. While it has been proposed that semanticity is exclusively human, Aitchison notes that the vervet monkey may have particular meanings associated with different sounds; e.g. it may mean 'snake' when it chutters (Aitchison, 1998, p.28).
- Cultural transmission: Aitchison proposes that a greater proportion of communication is genetically inbuilt in animals than in humans. Human children learn a lot of their language from those around them. See below (*p.12*) for a brief discussion of the innate language learning mechanisms, where it is noted that while a certain amount of human language may be innate, it needs to be triggered off by input.
- Spontaneous usage: Humans use language spontaneously but then it appears that some animals do too.
- Turn-taking: Humans quickly learn that in conversation, it is necessary for the speakers to take turns. Turn-taking is not exclusively human, as birds sometimes sing duets.
- Duality: Language is organised on two levels – the meaningless level of individual sounds and the meaningful level of individual sounds combined into words. Duality is also present in bird song.
- Displacement: This refers to the ability to refer to things removed in time and place. Clearly, this is easy with human language but seems to be absent or difficult in animal communication systems. Even the honey bee, who is

capable of giving some information about nectar in a different location, cannot use displacement to the same degree as humans (von Frisch, 1950, 1954). Therefore, displacement seems to be a feature which is important in human language but only partially present in animal communication systems.

- Structure-dependence: This refers to the ability of a native speaker of a language to recognise different units which are structurally equivalent; e.g. "the big man" is a unit consisting of three words which can function as the subject of a sentence. A single word such as "John" may equally function as the subject. Native speakers know this intuitively. Aitchison (1998, p.30) notes that based on what is now known about animal communication systems, it appears that they do not use structure-dependent operations. However, above it was mentioned that whale song, with its use of sub-phrases, phrases and song seems to exhibit some sort of structure-dependence. Clearly more research is required in this area to investigate the syntax of whale song before any definitive statements can be made.
- Creativity: This refers to the ability to produce and understand an indefinite number of novel utterances. Animals normally have a fixed number of signals which convey a set number of messages. Work with dolphins showed that while they may have a very sophisticated system of communication, it was not 'creative' in the human sense (Evans & Bastian, 1969).

Aitchison (1998, p.34) concludes that if it is necessary to use all the design characteristics of human language 'naturally' in order to qualify as using language, it must be said that animals do not qualify. Some animals possess some of the features but no animal system clearly exhibits semanticity or structure-dependence. Also, animals seem to lack the ability to communicate creatively.

Thus, while music and language both constitute systems of communication, language is generally considered to be unique to humans. Music, on the other hand, is not uniquely human. The aim of this chapter is to present an introduction to the types of research which investigate the relationship between music and language in

humans. A brief overview is sufficient to demonstrate the varied nature of this research and to highlight the exciting findings currently emerging. Studies which are directly relevant to this thesis will be discussed in greater detail in later chapters. Following the general overview of research in the area of music and language, the structure of the thesis is laid out.

The similarities between music and language should not be underestimated. Both constitute systems of communication with sophisticated structural properties such as recursion and duality of patterning (Trehub, 2003). Both music and language are claimed to be innate facilities (Chomsky, 1964; Pinker, 1994; Trehub, 2000, 2001) which are shaped by cultural influences; e.g. a child born in Ireland is likely to learn English as a mother tongue whereas a child born in France is likely to develop French as a first language. Similarly, the child will grow accustomed to the music of his/her society, whether that is Western tonal music based on scales or Middle Eastern music which employs maqâmât.[1] In addition, there is growing evidence that both music and language may involve similar processing locations in the brain. These similarities are introduced briefly below.

On the other hand, there are clear differences between the two modalities; e.g. the capacity of language to offer adaptive advantages is clearer than that of music. Language facilitates communication between humans, whether through oral or written language or through sign language. The significance of music, however, is less clear cut (Hauser & McDermott, 2003). Music cannot communicate meaning as precisely as language and, while there are aural and written forms, 'sign' music does not exist in the way that sign language exists. Before presenting the similarities between music and language, a number of other differences between the two modalities are discussed.

One major difference between music and language relates to semantics. As argued by Lerdahl and Jackendoff (1983, p.5),

[1] In Arab music, melody is usually based on scales or modes known as maqâmât. Maqâmât are generally treated as scales consisting of 24 equal quarter-tones. See Music Dictionary Online http://www.dolmetsch.com/defsm.htm (last verified 18.07.2005).

> "whatever music may "mean", it is in no sense comparable to linguistic meaning; there are no musical phenomena comparable to sense and reference in language, or to such semantic judgements as synonymy, analyticity, and entailment…."

Trehub (2003, p.669) also notes the lack of semanticity in music although it exhibits duality of patterning; i.e.

> "although discrete, meaningless elements are combined to produce meaningful structures, the resulting musical pieces are not meaningful in the same way that verbal utterances are."

Music lacks a semantic negation operator. At the pragmatic level it appears to lack the possibility of lying.

One example, however, of meaning in music is the Devil's Chord. The Catholic Church in the Middle Ages banned a particular musical interval known as a tritone (e.g. C and F#) on the grounds that it was the "devil's chord". Kay Gardner (1997, p.117) suggests that

> "… the tritone, when sung at length as harmony by a group of meditators, will take singers and listeners to a place where they will be in touch with Divinity. Perhaps this is a reason why it was so threatening in ages past."

If music was completely meaningless, there would surely be no reason to ban certain chords or compositions. Even today, censorship still exists in the field of music. In his book, Taboo Tunes, Peter Blecha (2004, pp.6*ff.*) addresses the attitudes of the Taliban in Afghanistan to music. He notes that after seizing the city of Kabul in 1996, they barred women from singing in public. The explanation given by one Muslim cleric was that

> "according to Islam, one of the worst sins is to encroach on a person's consciousness. When someone listens to music, his state of mind changes."

Even the practice of keeping caged songbirds in the home was banned by the Taliban because of the melodies they produce. Such ideas, as embedded in Western tradition, originated in the work of Plato (360 B.C.E.), who believed that music had no place in a well-organised state. He believed that art simply served to excite the passions which might destroy the proper balance of the individual soul. This in turn would upset the efficient running of the state and as such, music should not be permitted in a well-organised republic. Plato suggested that art was an imitation of

reality, and as artists might get their imitations wrong, art was in danger of spreading misinformation. Hence,

> "hymns to the gods and praises of famous men are the only poetry which ought to be admitted into our State."
>
> (Book X)

While the actions of the Taliban highlight their acknowledgment of the power of music, music achieves its effects through the overall composition. Each note does not 'mean' something in the way that each word of a language has meaning.

Another difference between music and language relates to the extent to which each corresponds to a "module" of the mind. Fodor (1983) was one of the first to assert the existence of a language module. He proposes that modular cognitive systems are "domain-specific, innately specified, hardwired, autonomous and not assembled" (*ibid*, p.37). Music is not proposed by Fodor to constitute a module. Other researchers also point to the autonomy of the language faculty; e.g. Fromkin (1997) asserts that there is growing empirical evidence to support the autonomy view of language as a separate 'organ'. She, like Fodor, cites Franz Joseph Gall (1791, 1810) who was among the first to argue against the view that the brain was an unstructured organ and in favour of discrete anatomical areas which were directly responsible for specific cognitive functions. Later, Broca (1861) and Wernicke (1874) discovered specific areas involved in language production and comprehension; areas which are still called Broca's and Wernicke's regions respectively. Other evidence put forward by Fromkin to support the notion of a language module comes from research with 'idiot savants'. These individuals show severe deficits in general intellectual abilities, yet show remarkable talents in particular areas such as languages. She concludes that

> "the more we look, …, the more we find that knowledge and processing of language is separate from the ability to acquire and process other kinds of knowledge, that the asymmetry between general knowledge and linguistic knowledge shows language to be independent of general intellectual ability, and that language itself, as well as other cognitive systems are distinct both anatomically and functionally."
>
> (Fromkin, 1997, p.23)

Fromkin's arguments in favour of modularity now seem overly simplistic. Language no longer seems to be completely separate from other cognitive abilities such as reasoning, Working Memory abilities or musical abilities. While Fromkin suggests that the fact that particular areas of the brain subserve particular functions is evidence in favour of modularity, this now appears to be an oversimplification. This issue is further examined in Chapter 3, but for now it suffices to say that it seems that these areas, once believed to be solely and specifically used for language processing, may have more general functioning than was previously believed (see e.g. Müller & Basho, in press). In fact, these areas are used in processing stimuli other than language. Of course, in itself the fact that there is not a single location solely used for language processing does not mean that a language module does not exist. There does not necessarily have to be a one-to-one relationship between location and function; i.e. it is possible that the language faculty uses neural resources in different parts of the brain and that these resources are also used by other cognitive abilities. The main point here is that pinpointing the location of various processes in the brain is a complex task and much remains to be known about the interactions of various processes and the functions of the different parts of the brain.

Similarly, Fromkin's argument relating to 'idiots savants' does not seem adequate to be the sole argument in favour of a language module. 'Idiot savants' are rare, abnormal individuals. It is necessary to be cautious before generalising from their situation to that of the normal population.

Singleton (1998) for his part, argues against encapsulation of the language module. He highlights the influence of higher level contextual factors on language processing; e.g. a native speaker of Finnish studying in France failed to understand her L1 as she was expecting to hear French rather than Finnish. Clearly her comprehension of Finnish was blocked by contextual factors outside the language domain. A study by Emmorey *et al.* (1993) shows that usage of American Sign Language (ASL) results in improved visuospatial memory. They suggest that these findings provide additional evidence against the modularity theory of language, as image generation and rotation are central to both American Sign Language and also to visual memory. Thus, this points to an interaction between language and other

cognitive faculties. Hence, language cannot be completely encapsulated; there is a limit to its modularity. The issue of American Sign Language and Working Memory is addressed again in *Section 2.7.4*. Chapter 3 considers evidence from areas as diverse as education and neurology. The conclusion there is that language no longer seems to be completely separate from other cognitive abilities such as reasoning, mathematical ability, etc.

Despite the differences between music and language, Lerdahl and Jackendoff (1983) attempt to analyse music using a generative-transformational grammar, an approach more commonly associated with linguistics. They believe that this is warranted given the significant parallels between the theories of music and language; i.e. in both cases there is a combination of psychological concerns and a formal nature to the theory (*ibid.*, p.5). Lerdahl and Jackendoff were successful in using similar structures to parse music and language fragments. They found similarities between prosodic tree structures used to describe the phonology of a language and time-span reduction trees which they developed to describe music.

Carstairs-McCarthy (1999) assigns great importance to the phonology of a language in determining syntax. This is a novel approach as it prioritises the role of the sounds of a language in determining abstract structural and semantic features like syntax and categorical differences in denotation. Carstairs-McCarthy develops the idea that the phonology of a language may in part account for its syntax. He suggests that the descent of the larynx led to changes in the vocal tract which allow syllabically organised vocalisation. This in turn permitted an increase in vocabulary which then required a reliable syntax. The source of this syntax is the same neural mechanism as that controlling syllable structure. Hence, some features of syntax seem to be by-products of characteristics of the syllable; e.g. grammatical subjects may be by-products of onset margins.[2] This is an interesting, albeit unusual, approach to syntax insofar as it proposes that syntax is influenced by structural constraints imposed by sound patterns. It would seem plausible that some sort of

[2] A maximal syllable can be divided into a central nucleus (generally vocalic) and two margins (generally consonantal), namely an onset preceding the nucleus and a coda following it, see http://www.cogsci.ecs.soton.ac.uk/cgi/psyc/newpsy?11.082 (last verified 20.07.2005).

parallel might exist in music, whereby musical structures are constrained by the range of possible notes.

Other researchers have also examined the parallels between the syntax of music and language. Trehub (2003) notes that both are rule-governed and that both embody the property of recursion. Bod (2002) attempts to provide a unified model of structural organization in music and language which could be used for parsing both music and language fragments. His approach uses tree diagrams for both music and language.

It is well known that tree diagrams are often used to describe the structure of language fragments; e.g. Figure 1 presents a tree diagram for the sentence "She saw the man with the wooden leg". Figure 1 is a basic tree diagram which simplifies over the functional categories of X-bar theory or any particular framework for articulating linguistic theories. In addition, it does not show the nuances of constraint-based frameworks such as HPSG.[3] In drawing trees for language fragments, the constituents are labelled with syntactic categories. This is possible, because in language there are constraints on which constituents may be combined; e.g. a determiner (DET) and a noun (N) together form a noun phrase (NP). A verb (V) and a noun phrase together form a verb phrase (VP). A sentence (S) is formed from a noun phrase and a verb phrase.

[3] Head-driven Phrase Structure Grammar (see e.g. Pollard & Sag, 1994)

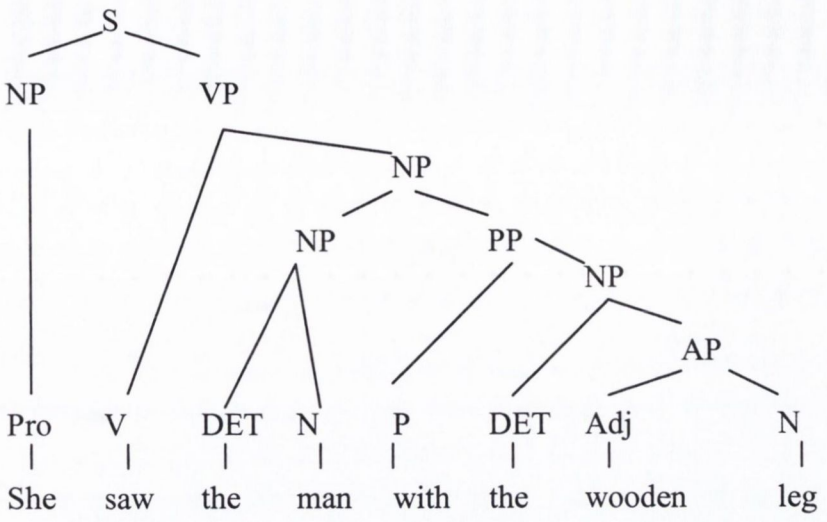


Figure 1. Tree diagram for sentence of English

Music may also be ascribed a tree structure. As in music any two notes have the potential to be combined, there are no labels on the constituents. Lerdahl and Jackendoff (1983) propose a number of rules for segmenting music and then apply these rules to various fragments. According to Lerdahl and Jackendoff, there are two types of grouping rules which govern how a musical passage is segmented. These are, firstly, grouping well-formedness rules (GWFR), which establish the formal structure of grouping patterns. Secondly, there are grouping preference rules (GPR), which establish which of the formally possible structures that can be assigned to a piece, correspond to the listener's actual intuitions. Some examples of grouping well-formedness rules are as follows (*ibid.*, pp.37*ff.*):

GWFR1: Any contiguous sequence of pitch-events, drum beats, or the like can constitute a group, and only contiguous sequences can constitute a group.

GWFR2: A piece constitutes a group

GWFR3: A group may contain smaller groups.

There are two additional well-formedness rules governing smaller groups within a larger group. Some examples of grouping preference rules are as follows:

GPR1: Avoid analyses with very small groups – the smaller, the less preferable.

GPR2: (Proximity) Consider a sequence of four notes *n1 n2 n3 n4*. All else being equal, the transition *n2 – n3* may be heard as a group boundary if

a. (Slur/Rest) the interval of time from the end of *n2* to the beginning of *n3* is greater than that from the end of *n1* to the beginning of *n2* and that from the end of *n3* to the beginning of *n4*, or if

b. (Attack-Point) the interval of time between the attack points of *n2* and *n3* is greater than that between the attack points of *n1* and *n2* and that between the attack points of *n3* and *n4*.

Other types of rules are also discussed, e.g. metrical well-formedness rules, which a listener uses to associate a metrical structure with a musical surface. There are also rules governing reduction, which allow a listener to hear the basic structure underlying a highly ornamented piece of music. Figure 2 shows one possible tree diagram for a musical phrase. The example is presented in Lerdahl and Jackendoff (1983). Other trees are possible, e.g. tree structures for each bar of the music, but following Lerdahl and Jackendoff's rules to achieve the most psychologically plausible tree, results in the tree presented in Figure 2.

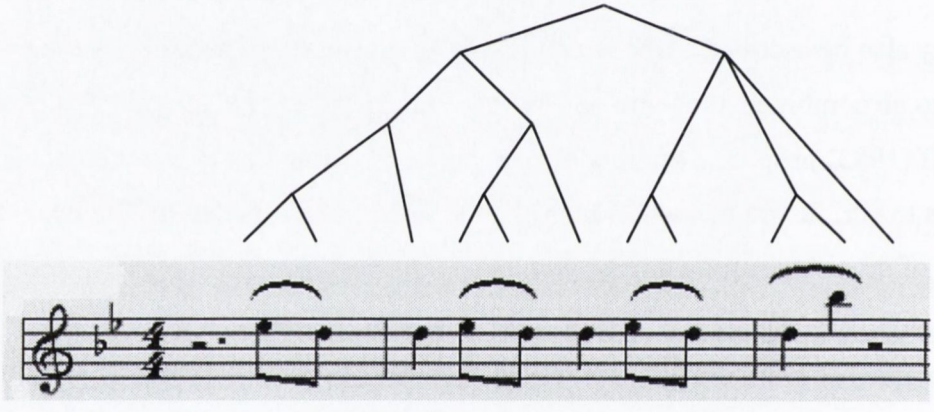


Figure 2. Tree diagram for musical phrase

Bod (2002) suggests that a general underlying parsing model could exist for music and language. He notes that the same data-oriented parsing[4] model with the same parameter setting achieves maximum accuracy in parsing both musical and linguistic structures. This is a parsing model based on simplicity and likelihood. He points out that the perceptual system strives for the simplest structure but is biased by the likelihood of previous structures. Combining these two principles allows accurate prediction of the perceived tree structure in language and music.

An alternative detailed model of how musical structures may be parsed is presented by Sutcliffe (in progress). He suggests that as music evolved over the centuries,

[4] Data-Oriented Parsing (DOP) is based on the idea that newly perceived input is understood in terms of previously perceived input. In other words, DOP analyses new data by probabilistically combining fragments from a corpus of previously analysed data (see homepage of Rens Bod http://staff.science.uva.nl/~rens/whatisdop.html, last verified 23.08.2005).

composers subconsciously emulated the syntactic structures which exist in language. This approach to the syntax of music is based on chord progressions. The advantage of this approach is that it makes explicit what constitutes a phrase and how a piece of music may be segmented into its constituent parts. Sutcliffe proposes that the phrase in music is equivalent to the sentence in language, with the musical phrase built up as follows:

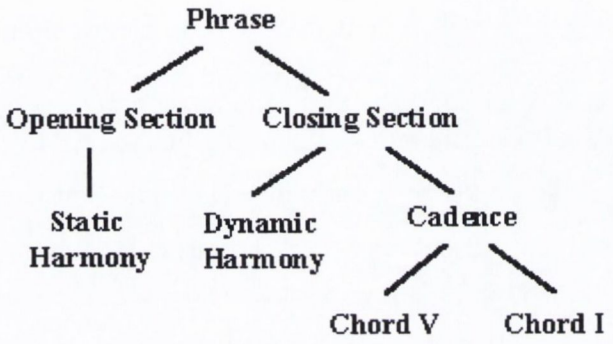


Figure 3. Syntactic structure in music

Static harmony is described as an oscillation of the tonic or dominant chord with other chords. As the harmony oscillates around the tonic, there is no sense of movement. The tonic is the note which denotes the key of the piece of music; e.g. in the key of F major, the tonic is F. The chord on that degree of the scale is usually described as chord I. Sutcliffe gives an example of static harmony in a piece of music written in the key of F major. For four bars, the phrase is composed of tonic harmony (F major) with just one movement to C major and then back to F major. Harmony that is dynamic is made up of chord progressions and creates a sense of movement. The following chord progression is an example of dynamic harmony from the same piece of music written in F major: F major, G minor, C major, F major.

Both language and music appear to be innate in humans (Chomsky, 1964; Pinker, 1994; Trehub, 2000, 2001), although the language faculty may need input of some sort in order to become active. One argument used in favour of an innate language faculty comes from the "Poverty of Stimulus" argument. Chomsky (1964) proposed that children must be born with a "Language Acquisition Device" which facilitates their learning of their mother tongue because, he suggests, children do not

receive adequate input to learn their first language without any innate mechanism. Input received by children is often imperfect, yet all normal children learn to use their mother tongue correctly (modulo individual differences, as discussed in Chapter 2). According to Chomsky the basics of language are present in the "Language Acquisition Device" but need to be triggered off by input. Evidence often presented to support the necessity of input comes from "wolf-children", i.e. children deprived of language input who fail to learn a native language (Curtiss, 1988; Grimshaw *et al.*, 1998; Rymer, 1993). A problem with this form of evidence is that "wolf-children" have generally suffered severe deprivation and abuse before they were rescued. Their language deficits may be the result of overall developmental problems. Evidence which suggests that musical ability may also be innate comes from the fact that members of a particular culture can recognise unmusical sequences without any specific musical training in the same way as native speakers recognise ungrammatical sentences (Hauser & McDermott, 2003). Even very young children have been shown to be sensitive to fundamental features of music; e.g. 2- and 6- month old infants listen longer to sequences of consonant intervals than to sequences of dissonant intervals (Trainor & Heinmiller, 1998). Trehub (2003, p.670) proposes that

> "it is reasonable to conclude, then, that the rudiments of music listening are gifts of nature rather than products of culture".

Connections between music and language have historically aroused much interest, and speculation as to the origins of the two has been seen throughout history. Rousseau (1998, p.318) suggests that "verses, songs and speech have a common origin". According to Rousseau, music and speech became separate due to the introduction of harmony to music. Others have proposed that music evolved as a sexually selected system, designed to attract mates and signal mate quality (Darwin, 1871; G. F. Miller, 2000). Another suggestion is that music is simply a by-product of various perceptual and cognitive mechanisms that serve other functions such as language (Pinker, 1997). Evidence to this effect comes from a study by Ramus and colleagues who showed that both human infants and cotton-top tamarin monkeys can discriminate on the basis of rhythm between sentences from different languages. Because non-human primates lacking in language are capable of this, it suggests that the capacity for discriminating language rhythm must have evolved for

more general auditory purposes (Ramus *et al.*, 2000). Hauser and McDermott (2003, p.667) conclude that

> "some aspects of rhythm perception for music may be tapping domain-general auditory mechanisms that may well have been in place before our species began producing music."

Research with rhesus monkeys shows that they have some musical sensitivity (Wright *et al.*, 2000). Hauser and McDermott (2003, p.665) argue that as monkeys do not produce music on their own, this shows that at least some aspects of music perception are determined by general properties of the auditory system. One problem with this argument however, is that while monkeys may not produce music as such, it is known that rhesus monkeys communicate by means of elaborate facial and vocal expressions (Ghazanfar & Logothetis, 2003; Hauser *et al.*, 1993; Hinde & Rowell, 1962). They distinguish between 'coo' calls and 'threat' calls (Rowell & Hinde, 1962). Whereas 'coo' calls are long, tonal signals, 'threat' calls are noisy, pulsatile calls of short duration. Clearly, in order to distinguish between these calls, monkeys employ certain abilities which seem closely related to musical skills. Hence, the monkeys may have some basic music-like abilities or indeed their abilities may be the result of the general auditory system as Hauser and McDermott suggest.

An interesting point of convergence between music and language is the epic poem, such as Homer's Iliad or Odyssey. Beye (1972) considers how poems of up to 16,000 lines in length could be passed down orally from one generation to the next. He suggests that while words are the basic units of ordinary speech, it would not be possible for the bards to have remembered so many words. Therefore, it seems that the minimal element of the Greek epic was the formulaic phrase. Each idea or action could be described by a phrase which had a fixed metrical quantity (*ibid.*, p.5) and there were very few phrases of identical metrical quantity which expressed identical meaning. This reduced the effort of the bard because he was not choosing individual words but larger phrases. It is interesting to compare this to the ability of musicians to memorise long pieces of music. Research in this area suggests that musicians make use of the structure of the piece in memorising and performing it (Williamson & Egner, 2004; Williamson & Valentine, 2002). Some bars appear to be of greater importance structurally than others as they serve to divide the piece

into segments. Williamson and his colleagues have shown that as musicians become more expert, they spend more time on the structurally significant bars which seem to facilitate encoding to and retrieval from memory. Therefore, a parallel can be drawn between the bards' memorisation of epic poems and expert musicians' memorisation of music. Rather than attempting to memorise individual words or notes, longer sections are used, i.e. formulaic phrases in the case of the poets and structurally significant bars in the case of musicians.

Both language and music are closely related to culture. Rousseau (1998, p.321) points out the difficulty in appreciating music which is unfamiliar in a particular culture, saying that

> "the most beautiful songs to our taste, will always only indifferently touch an ear that is not at all accustomed to them; it is a language for which one has to have the dictionary."

While some theorists agree with this view that musical meaning is determined exclusively by cultural convention (e.g. Blacking, 1973; Feld & Keil, 1994; Walker, 1996), an empirical study by Balkwill and Thompson (1999) reveals that emotion in music is communicated through a combination of universal and cultural cues. They propose that listeners may rely on either of these cues, or both, to arrive at an understanding of musically expressed emotion. Their investigation examined the ability of Western listeners, unfamiliar with Indian music, to rate the degree of sadness, joy, anger or peace in Hindustani raga excerpts. As there was nothing shared culturally between the Western listeners and the Indian community in which the music originated, the ability of subjects to successfully identify the emotions would suggest that there is indeed something universal in music which is not conveyed by cultural means. Results showed that the Western listeners were indeed able to identify these emotions quite successfully, leading Balkwill and Thompson to conclude that

> "the expression of emotion in the music of a given culture should be most salient to listeners of the same culture due to the shared understanding of the conventional representation of emotion within that tonal system. When familiar cultural cues are absent, the listener must pay more attention to basic perceptual cues, such as tempo and complexity. These cues allow

listeners to gain a general understanding of the intended emotion" (*ibid*, p.45).

The overlap between language and music in the brain is discussed in detail in Chapter 3. There it is shown that Broca's area, in addition to processing language, now appears to also be involved in the processing of musical syntax.

Given the similarities between music and language, it is not surprising that they have long provided the focus of much research. Should a substantive link be shown to exist between the two, education is one area which could potentially benefit. Already educators are attempting to exploit the extra-musical benefits of music education but there is still relatively little evidence to explain scientifically how music may have these effects. In the field of applied linguistics, researchers have long been asking questions as to the nature of the influence of musical aptitude on language aptitude (see e.g. Dexter & Omwake, 1934; Eterno, 1961). However, relatively little work was actually done in the area, with results that were somewhat inconclusive. This thesis attempts to advance research in this area, specifically as detailed later in Chapters 4 and 5.

Until relatively recently, the idea of a musical ear or an ear for languages was often consigned to folklore, and it was considered that researchers should get on with more important work on other aspects of second language acquisition research. Now, however, evidence shows that individuals do indeed differ in their aptitude for music and languages and that music and language abilities are not as disparate as was once believed. Thus, it appears that the musical ear may not simply be an old wives' tale. Clearly the assertion that music and language may share some common processes seems to contradict the hypothesis that language is modular. It now seems unrealistic to support the notion of a completely encapsulated language module, as evidence presented in Chapter 3 shows definite links between language and other abilities.

Given current interest in language and music and their relationship to each other, it seems like an interesting time to reconsider the language-music relationship with respect to the second language learning process. Language aptitude is beginning to

be better understood and the neural aspects of language and music are the focus of much research, although many issues remain unresolved in this area.

While the music-language relationship is receiving attention from many neuroscientists, it is an area which still generates relatively little interest among researchers in the field of second language learning. Given the paucity of information available on how musical aptitude influences second language learning, it seems prudent to investigate the relationship further. The few existing studies in the area have proved very interesting with preliminary findings being very positive (see *Sections 3.3.1, 3.3.2*). The study presented in this dissertation aims to contribute something to overcoming the current dearth of knowledge in this field.

The main aim of this thesis is to contribute to the growing body of knowledge on the relationship between music and language abilities in humans. More specifically, it is hoped to shed some light on the particular relationship between music and language aptitude in order to consider further how musical aptitude may impact on the language learning process. The focus of this study is second language learning.

The specific goals of this research are twofold. Firstly, I wish to reconsider earlier findings of an association between musical and linguistic aptitude and conduct an experiment along the same lines as some of the earlier studies. Early research focussed primarily on receptive skills, generally using pencil and paper tests. This approach is followed here, although the tests used are not identical to those in other studies. Identical tests were not used because they were unavailable; e.g. it proved very difficult to source the Seashore Measures of Musical Talent or the original version of the Modern Language Aptitude Test. Similar tests were readily available, i.e. the Bentley Measures of Musical Ability and a variation of the Modern Language Aptitude Test, and hence were employed in this study. Thus, though not an exact replication, a very similar methodology is used. Earlier results of other studies were not reanalysed – again because of difficulties in sourcing the original materials and resulting data in machine-readable format.

Secondly, I wish to broaden the scope of enquiry to consider productive skills in the two domains. Chapter 2 discusses Working Memory and productive tests of

Working Memory ability which are believed to be reasonably accurate predictors of language aptitude. To the best of my knowledge productive tests are not commonly used to test musical aptitude and I aim to test the viability of this approach. The issue of validity arises in relation to novel tests devised in the course of this study and is considered in Chapter 5.

The thesis attempts to explore the music-language relationship from the point of view of how any such relationship could influence the language learning process. It is set out as follows:

Chapter 2 introduces the concept of individual differences (IDs) and how these influence the learning process. Important individual differences which are closely tied up with musical aptitude and language aptitude are considered. These include IDs which may act as confounding variables in any study of musical aptitude or language aptitude, e.g. Working Memory ability and intelligence.

Chapter 3 focuses on the music-language relationship specifically and the work which has been done to date in this area. The empirical work carried out for this thesis is closely related to these earlier studies, so it is useful to consider their outcomes before embarking on any new research.

Chapter 4 presents details of the methodology of the current study.

Chapter 5 details the findings of this study. Results point to a moderate to large correlation between music and language aptitude in both the receptive and productive domains. Correlations remain significant even when intelligence is factored out.

Chapter 6 offers a tentative conclusion and some suggestions for further work. The primary result of the thesis is evidence of a positive correlation between music and language aptitude in a sample of Irish secondary school students.

Chapter 2 Individual Differences

2.1 Introduction

It is widely accepted that second language learners experience differential success in mastering a second language (L2). This is in stark contrast to first language (L1) acquisition where all normally endowed children succeed in the acquisition of their mother tongue, given normal circumstances. Of course individual differences in L1 verbal ability do exist, but these generally only emerge in the context of verbal ability tests or specific examination settings. Native speakers of a language, without a disability, normally master their L1 sufficiently to function in everyday situations. Second language learners, on the other hand, even after many years of learning, are often very far from reaching a native-like level of competence. Many different explanations have been proposed to explain why some learners struggle with a second language and only meet with limited success while others master all aspects, including pronunciation, syntax and pragmatics. Factors which have been proposed to account for differential success include age, language aptitude, social-psychological factors, personality, cognitive style, hemisphere specialization and learning strategies.

While there is no denying the existence of such important differences between learners, it remains unclear how they interact to influence the learning process. A greater understanding of such individual differences (IDs) is essential in trying to arrive at a thorough understanding of the second language acquisition (SLA) process. Indeed, over thirty years ago, Selinker (1972) said that a theory of second language learning that does not provide a central place for individual differences among learners cannot be considered acceptable. Current research also acknowledges the importance of IDs and how they interact with methods of instruction to determine the success or otherwise of the learner (see e.g. Robinson, 2002b); for example it is mentioned below in *Section 2.2* that cognitive styles interact with methods of instruction, with introverts found to learn better in a traditional classroom while extraverts seem to prefer discovery learning.

The aim of this chapter is to present an overview of individual difference research, focussing primarily on individual differences which have been shown to be significant for second language learning purposes. This chapter sets out to give a detailed account of those individual differences which are directly relevant to this thesis, while also briefly mentioning other significant IDs which may not be directly relevant to the current thesis. The particular individual difference which is under investigation in this work, i.e. the musical aptitude of learners, will be examined more thoroughly in Chapter 3.

The current chapter begins by presenting ID research in psychology which focuses on personality differences (*Section 2.2*). It then turns briefly to individual differences in first language acquisition in order to demonstrate that there can be individual variation even in the process of first language learning where all children were originally believed to pass through very similar stages (*Section 2.3*). The remainder of the chapter then examines differences which are significant for second language learners in general and are of particular significance for this study. The IDs considered in detail are as follows:

- Language Aptitude (*Section 2.5*)
- Intelligence (*Section 2.6*)
- Working Memory (*Section 2.7*)

2.2 Individual Differences in Psychological Research

Differences between individuals concern not only those interested in second language acquisition but also developmental psychologists, social psychologists and those investigating learning theory. Obviously, individual differences exist in all areas of the human condition such as athleticism, appetite, vision and height, to name but a few, but these are of less relevance to SLA. Research in psychology sheds much light on individual differences which may be directly relevant to SLA researchers. As is often the case in research on SLA, theories of developmental, social and educational psychology were traditionally based on broad generalisations. It was assumed that all children pass through the various stages of development in a specified order, all individuals of a species can be conditioned in a

common manner and all people generally behave in a similar way. Physiological psychologists generally assume that everyone's nervous system has a similar structure and will operate in the same way.

However, everyday experience suggests that differences between individuals are significant – not all children develop at the same rate, not all people can learn a task at the same rate. Such general observations seem to be supported by much scientific evidence. It seems that differences in personality, e.g. whether a person is introverted or extraverted, to name but one, may account for differences in behaviour and rate of learning. Leith (1974, cited in Engler, 2003) for example, found that extraverted children learned best in situations that emphasise discovery learning, i.e. where students explore for themselves and develop their own theories. Introverts were found to learn better using a traditional approach where the teacher presents concepts more formally. The introversion-extraversion distinction may impact directly on language learning; e.g. a study by Dexter and Omwake (1934), discussed in detail in *Section 3.3.2* below, found that introverted language learners received higher scores than extraverts when rated for native-like accent in the language they were learning. The introversion-extraversion distinction relates to the question of cognitive styles, which themselves have been shown to be important for second language learning success (e.g. Ehrman, 1997; Ehrman & Leaver, 2003; Ehrman *et al.*, 2003; Schmeck, 1988). Cognitive styles are the preferred way in which individuals process information or approach a task (Larsen-Freeman & Long, 1991, p.192). The term cognitive style is now often used interchangeably in the literature with other terms such as learning style or personality type.

Terms such as extraversion and introversion come from studies of personality. There are many different types of personality theory, ranging from phenomenological theories, through theories arising from factor analysis of personality questionnaires to theories attempting to find the biological basis of personality. A brief overview of some of the best known theories of personality in each of these categories follows.

Phenomenological theories of personality attempt to account for personality differences through differences in people's conscious experience. George Kelly

(see Kelly, 1955, 1963) and Carl Rogers (see Rogers, 1959, 1967; Rogers *et al.*, 1989) focus on how the individual perceives himself and others. According to Kelly, a lot of our time is spent trying to evaluate other people in order to predict their likely behaviour. As different people notice different aspects of behaviour, individuals develop different sets of personal constructs according to which they evaluate those around them. A person's background and values also influence the way in which they construe behaviours. Therefore, a person's experience determines the values of his/her personality traits. Little (1991, p.19) examines how personal constructs impact on the learning process. He notes that

> "any learning task requires the learner to assimilate new knowledge to his current system of constructs. When the new knowledge is additional information about a subject the learner is already familiar with, learning may proceed without any great difficulty."

This contrasts with the situation where the new knowledge conflicts with the learner's existing system of constructs; a situation which may result in learning being "not only difficult, but painful" (*ibid.*). This view of learning as a process, whereby new information is incorporated into the existing system, is based on Piaget's theory of child development (see *Section 2.6.1* below).

Freud's (1917/1964, 1932, 1957) psychoanalytical theory also appeals to the phenomenological etiology of differences between individuals, grounding them in differences in childhood experience. He proposes that events affect the subconscious mind, which in turn affects the actions of the individual at a later date. According to Freud, the mind can be divided into sections called the ego (conscious and capable of rational thought), the id (unconscious) and the superego (containing a person's moral values). This division of the mind into conscious and unconscious sections allows Freud to account for the fact that memories can be forgotten, yet still influence behaviour, i.e. they remain in the id but affect the ego. The id can then be further divided into two basic instincts – the eros (life force) which is satisfied through sexual activity and the thanatos (death instinct) which is a destructive force. Freud proposes that during childhood there is conflict between these two basic instincts. He suggests that there is also conflict between the basic instincts of a child and the attempts of adults to ensure that the child learns to satisfy his instincts in socially acceptable ways. Hence, childhood experiences are

important in determining future adult behaviour. While Freud's theory was one of the first formally stated modern theories of personality, it has to a large extent been discredited in main stream psychology. A thorough critique of Freud's work is beyond the scope of this thesis but more detailed discussion may be found, for example, in Kline (1981), Erdelyi (1985) and Fisher and Greenberg (1996).

The theories of Kelly, Rogers and Freud attempt to account for individual differences through detailed analyses of individuals. This approach may be contrasted with trait theory which allows people's personalities to be compared in terms of personality traits. Trait theories assume that there is a certain constancy about the way in which people behave, i.e. they assume that behaviour is not completely determined by the situation. If, for example, I am asked to describe a friend, I might describe him as trustworthy and honest. In doing so, I am suggesting that another person who encounters the same individual is also likely to find that he displays these characteristics. Thus, it can be said that his trustworthiness and honesty are enduring traits rather than situation specific qualities. As a result of traits being relatively stable over time, we can consider how individuals differ from each other on various traits.

Early work in trait theory came from Eysenck who recognised two major personality factors (H. Eysenck, 1973; H. Eysenck & Levey, 1972). He describes these two factors by detailing their opposite poles, i.e. each factor has two extremes and an individual will lie somewhere along the continuum. Eysenck's two major factors are:

- Introversion vs. Extraversion
- Neuroticism vs. Emotional Stability

The extraversion-introversion distinction reflects the degree to which a person is outgoing. The neuroticism-emotional stability distinction reflects the stability of an individual's behaviour over time. Eysenck later added psychoticism as a factor of personality (H. J. Eysenck & Eysenck, 1976). Psychoticism does not have two extreme values. Rather it is present to varying degrees in an individual. It is characterised by the loss or distortion of reality and the inability to distinguish between reality and fantasy.

The major advantage of Eysenck's model is his attempt to provide biological explanations for his factors of personality. He proposed that the introversion-extraversion trait is related to arousal thresholds in the Ascending Reticular Activating System (ARAS)[5] of the brain. According to Eysenck, the ARASs of introverts and extraverts may operate at different levels, as a result of which the cortices of introverts are habitually more electrically active than those of extraverts. The higher level of arousal in introverts may create a constraint on their behaviour, leading them to be more reserved and careful. The lower levels of arousal in extraverts may lead to an absence of constraints, which in turn causes them to act more impulsively (see Engler, 2003).

Eysenck suggests that the emotional stability-neuroticism dimension is related to differences in visceral brain (VB)[6] activation. While not all of Eysenck's hypotheses are supported by modern research, his model is still important as it provides testable hypotheses which may be falsified. In addition, some of Eysenck's factors emerge in the current most widely held theory of personality - the Big Five.

The Big Five or the Five-Factor model of personality is probably the most widely accepted trait theory at the moment. The five factors - openness, conscientiousness, extraversion, agreeableness and neuroticism - emerged from studying natural language, where it was found that despite approximately 18,000 words in English which refer to personality attributes, only a much smaller number of clusters was needed to cover all the terms. Cattell (1965) argues for sixteen personality factors, whereas Eysenck, as mentioned above, suggests that three personality factors are sufficient. Others propose that five clusters are necessary (e.g. Goldberg, 1981; Norman, 1963). The possibility that five factors are sufficient was investigated empirically through the development of personality questionnaires. Much of this

[5] The Ascending Reticular System carries input from all sensory organs. It projects to the thalamus and cortex. This system controls the amount of electrical activity in the cortex and is important for such functions as arousal from sleep by an alarm clock, bright lights or movement (see http://pathology.mc.duke.edu/neuropath/nawr/cranial-nerves.html, last verified 31.08.2005).

[6] The visceral brain refers to a system of functionally related neural structures in the brain that are involved in emotional behaviour (http://wordnet.princeton.edu/perl/webwn?s=visceral+brain, last verified 31.08.2005). It includes the limbic system and the hypothalamus, both of which are involved in motivation and emotional behaviour.

work in measuring the dimensions of the Big Five has been done by Robert McCrae and Paul Costa (P. Costa & McCrae, 1985; Paul Costa & McCrae, 1992; McCrae & Costa, 1989, 1990).

Cooper (1998) discusses the Five-Factor model of Personality, which can be described by the acronym OCEAN (openness, conscientiousness, extraversion, agreeableness, neuroticism) but his criticism of this model is that facets of personality deemed to belong to different factors can in fact turn out to be highly correlated. In addition, some of the five factors themselves are highly correlated. Thus, he concludes that the five factors do not seem to be independent. Engler (2003) too acknowledges that there is still some disagreement concerning the exact nature of each of the five factors. However, the Five-Factor model would seem to be the one receiving most support in the literature at the present time.

Having attempted to establish the structure of personality, Cooper (1998, p.76) points out that it is necessary to ask what causes certain behaviours to vary together to form these traits and also what causes an individual to develop a particular personality, e.g. environmental factors or genetic and biological factors. As mentioned above, Eysenck was particularly interested in the biological factors which determine personality. As scientific and medical equipment become increasingly sophisticated, it becomes much easier to examine activities in the brain and hormone levels in the body. Hence, we can better understand the physiological processes underlying personality. However, as Butt (2004) points out, simply knowing that a person has a certain level of a particular hormone in his/her body does not necessarily enable us to determine why he/she is anxious, for example. Such a hormone level may also be present even if the person is not anxious or the person may be anxious without having the corresponding hormone level. Thus, direction of causality is hard to determine and it is difficult to say if personality determines physiology or vice versa.

One non-biological account of how personality factors may interact comes from self-presentational theory. This theory attempts to account for how personality factors may interact to influence anxiety levels. According to this theory, social anxiety

> "arises whenever people are motivated to make a desired impression on others but are not certain that they will do so."
> (Schlenker & Leary, 1982, p.645)

Hence, social anxiety is a function of motivation and the doubt that one will make a good impression. Social anxiety increases as the level of motivation increases and the level of doubt increases. This is one example of how a personality trait may vary in accordance with a number of factors.

The aim of this section was to present an introduction to research in the area of personality which investigates individual differences. It is seen that much remains to be learned about personality traits and how they vary from one individual to the next. As much remains unknown about individual differences, it is quite an exciting field of research.

Even from the short discussion above, it should be clear that some personality differences are of greater importance than others in the context of language learning. Anxiety levels and the ability to deal with anxiety are certainly important factors in language learning success. As such, anxiety has received much attention in the literature on SLA (e.g. Gregersen & Horwitz, 2002; Kitano, 2001). As mentioned above, personality traits are also closely related to cognitive styles. Cognitive styles and their relationship to language learning strategies have been extensively studied in the literature (e.g. Chapelle, 1995; Oxford, 1990; Şakar, 2003). The remainder of this chapter deals specifically with IDs which are relevant for language learning.

2.3 Individual Differences in First Language Acquisition

Having established that the personalities of individuals differ in significant ways, it may be worth re-examining other aspects of human behaviour and development to determine if in fact performance is as uniform as was once believed. One area worthy of consideration is first language acquisition, where it was traditionally proposed that children pass through stages of development in a fairly well ordered

sequence. It now seems that children can take qualitatively different routes into first language acquisition. Shore (1995) claims that not all children acquire language in the same way and that the differences among them are not simply a result of some proceeding faster along a common pathway.

The most common distinction between types of L1 acquisition is that of "referential" vs. "expressive". Referential development involves the child learning individual words, often nouns, and then later combining these into multiword combinations and sentences. On the other hand, the expressive pattern involves phrases from the beginning. These phrases are generally of personal-social utility, e.g. "lemme see". The child gradually recognises the individual words and acquires the ability to recombine them in different ways to make novel utterances.

It has been suggested that this difference between referential and expressive acquisition is related to the child's position in the family (Pine, 1995); e.g. first-borns may receive more child-directed simplified speech so they have the opportunity to learn individual words. Later-born children on the other hand, may encounter more language that is not specifically tailored to their level, with the result that they pick up more "unanalysed whole" language samples. This has been shown in empirical research where second-born children were found to have a significantly higher percentage of frozen phrases and deictic personal pronouns in their first 100 words than first-borns (*ibid.*).

Shore (1995) gives examples of two children who progress in very different ways. Julia (see Bates *et al.*, 1988), said 34 words at 13 months, 80% of which were nouns. At 20 months she knew 290 words, 90% of which were content words. Her first combinations were generally made up of two nouns; e.g. "water bottle" meaning that the bottle is in the water.[7] Julia contrasts dramatically with Maia (see Adamson *et al.*, 1984), whose first words were often surrounded by jabbering in sentence-like intonation contours. At 20 months, her utterances showed sentence-like intonation contours but articulation of individual sounds was poor. At 25 months, approximately half of her "words" were stock phrases such as "Lemme

[7] Of course, one difficulty inherent in research on first language acquisition is knowing precisely and definitively what a child means by a particular utterance.

see", "Ready to go" and "Where ya goin'?". She used pronouns rather than nouns to refer to objects.

In addition to the referential vs. expressive distinction, children may be classified as nominal or pronominal, analytic or holistic, risk takers or conservative. Children who often use formulaic utterances may be described as holistic because they appear to pick up phrases as whole units. They contrast with children who use a more analytic strategy and build up sentences out of individual words. Interestingly, similar terms resurface in the context of second language acquisition where learners may be classified as sequential-global or field independent-dependent (see e.g. Bruner *et al.*, 1956; Hatch, 1974; Pettigrew, 1958; Witkin *et al.*, 1971).

Having seen the importance attributed to individual differences in the study of psychology, as well as the increasing interest being shown in this area in relation to first language acquisition, it is not surprising that individual differences are the subject of a large body of research in the area of second language acquisition. Individual differences in SLA serve as the focus of the remainder of this chapter.

2.4 Differences between Second Language Learners

Of course there are innumerable individual differences which influence the language learning process and its outcome. Indeed Gardner and MacIntyre (1992, p.212) propose that

> "there are probably as many factors that might account for individual differences in achievement in a second language as there are individuals."

However, despite the inherent problems in trying to list all possible IDs, it is worth considering some of the different ID taxonomies presented in the literature in order to gain an insight into the research which has taken place to date.

Unfortunately, while it is useful to list and discuss the ways in which learners differ, a thorough understanding of IDs will only become possible following the detailed examination of the interaction between IDs and learning contexts and the interaction between IDs and learner strategies, as noted by Ellis (1994) and Robinson (2002b).

This is because the characteristics of learners may influence their reaction to various teaching styles; e.g. exceptionally shy learners may feel uncomfortable in a communicatively oriented classroom whereas they may react more positively to learning analytically.

Before we can hope to understand the interaction between IDs and other factors, it is necessary to be aware of at least some of the IDs receiving attention from researchers. Rod Ellis (1994) for one, examines in detail the ways in which language learners differ. He investigates the effects that these differences have on learning outcomes and how they affect the process of L2 acquisition. He also looks at how individual learner factors interact with instruction in determining learning outcomes.

Before considering individual difference taxonomies, it is useful to highlight the inherent difficulties in attempting to classify individual differences. Different researchers use different terms to describe similar concepts. This makes direct comparison of taxonomies difficult; e.g. it will be seen below that Altman (1980) includes cognitive styles, especially field dependence/independence, in his taxonomy without further examples. Skehan (1989), on the other hand, gives a number of examples of cognitive factors, one of which is field dependence. Others include extraversion/introversion, risk-taking, intelligence and anxiety. Whereas Skehan includes extraversion/introversion under cognitive and affective factors, Altman places the introversion dimension under the heading of personality factors. Therefore, it is clear that while both researchers consider cognitive factors such as field dependence and introversion to be of significance, they are not always consistent in their classification of such factors.

Another issue of importance relates to the fact that different researchers attach varying degrees of importance to the various individual differences with the result that some IDs are omitted from some taxonomies. Altman notes that his list may not be complete and no claim for completeness is made. In general, in the literature, taxonomies of individual differences seem to be constructed around factors of interest to each researcher, with little reference to previously published

taxonomies. According to Altman (1980, p.5*f.*), the following factors influence individual learner differences in language learning:

1. Age
2. Sex
3. Previous experience with second-language learning
    a. General
    b. Specific to the target language
4. Proficiency in the native language
5. Personality factors, such as:
    a. Introvert-extrovert
    b. Capacity for empathy with foreigners
    c. Goal-oriented vs. role-oriented
    d. Competitive vs. withdrawing
6. Language Aptitude
    a. Auditory ability
    b. Grammatical sensitivity
    c. Inductive learning ability
    d. Rote memory ability
7. Attitudes and Motivation
    a. Motivational orientation
        i. Integrative
        ii. Instrumental
        iii. 'Linguistic hobby'
    b. Intensity of motivation
    c. Source of motivation
        i. Peer group
        ii. Parental pressure
        iii. Internal
    d. Attitudes toward:
        i. The target language
        ii. The target culture or people
        iii. Language learning in general
        iv. The target language teacher
        v. The environment for learning

8. Learning rate
9. General Intelligence
10. Sense Modality Preference
    a. Visual
    b. Auditory
    c. Tactual
    d. Kinaesthetic
11. Sociological Preference
    a. Whole-class environment
    b. Large-group environment
    c. Small-group environment
    d. Independent study
    e. Self-instruction (programmed instruction)
    f. Learning with peers
    g. Learning with teacher
12. Cognitive Styles, especially field dependence/independence
13. Learner Strategies
14. Learner errors

Peter Skehan's (1989) taxonomy of individual differences is rather shorter than that above, yet many of the same factors emerge:

1. Language Aptitude
2. Motivation
3. Language Learning Strategies
4. Cognitive and affective factors
    a. Extroversion/introversion
    b. Risk-taking
    c. Intelligence
    d. Field Independence
    e. Anxiety

Larsen-Freeman and Long (1991) for their part, suggest that the following factors contribute to differences between learners:

1. Age

2. Aptitude
3. Socio-psychological Factors
    a. Motivation
    b. Attitude
4. Personality
    a. Self-esteem
    b. Extroversion
    c. Anxiety
    d. Risk-taking
    e. Sensitivity to rejection
    f. Empathy
    g. Inhibition
    h. Tolerance of Ambiguity
5. Cognitive Style
    a. Field Independence/Dependence
    b. Category Width
    c. Reflexivity/Impulsivity
    d. Aural/visual
    e. Analytic/gestalt
6. Hemisphere Specialization
7. Learning Strategies
8. Other factors
    a. Memory, awareness, will
    b. Language Disability
    c. Interest
    d. Sex
    e. Birth Order
    f. Prior Experience

The individual differences which are of direct relevance to this thesis are language aptitude, intelligence and Working Memory. These are discussed in detail below.

2.5 General Aptitude

Language aptitude is investigated here as an individual difference between learners because the concept is extremely important for the remainder of this thesis. In trying to examine the impact of musical aptitude on language learning, it is language aptitude that I will attempt to measure rather than levels of achievement in language learning. This facilitates consideration of whether or not having good musical aptitude improves a language learner's chance of success.

General Aptitude is defined by Carroll (1981) as the "capacity of learning a task, which depends on some combination of more or less enduring characteristics of the learner." Later, he states that

> "to the extent that cognitive abilities are at least relatively stable and relatively resistant to attempts to change them through education or training, and at the same time are possibly predictive of future success, they are often regarded as aptitudes" (Carroll, 1993, p.16).

With regard to language aptitude, it is proposed that individuals with higher aptitude have a special talent which will enable them to make faster progress. In addition, they may also eventually reach a higher level of proficiency. The idea that high aptitude may be particularly important in the case of adult learners who no longer have the advantages associated with youth is discussed in *Section 2.5.4* below.

2.5.1 Stability of Aptitude

Concerning the stability of language aptitude over time, Politzer (1969) found that training on relevant cognitive tasks did not lead to an improvement in language aptitude scores. Support for this idea of the stability of aptitude comes from Skehan & Ducroquet (1988) who provide evidence of significant relationships between measures of first language development and language aptitude scores in adolescence. They cite the findings of the Bristol Language Project, which show significant correlations between early first language measures and aptitude indices ten years later. If indeed measures of L1 development are significantly related to L2 aptitude, language aptitude must be relatively constant. Carroll (1973) also posits an association between first language learning and foreign language aptitude,

proposing that foreign language aptitude might in fact be the residue of L1 learning ability.

However, not all researchers agree with this proposal that aptitude is stable; e.g. it has been proposed that "aptitude should not be viewed as a static personality trait; novices can become experts with experience" (McLaughlin, 1990, p.173). Others tentatively suggest that "language aptitude is a form of developing expertise rather than an entity fixed at birth" (Grigorenko *et al.*, 2000, p.401). On the other hand, Grigorenko *et al.* also acknowledge that more research is needed to examine which aspects of aptitude may be modifiable.

Given the current uncertainty, the best conclusion may be that

> "although language aptitude appears to be a relatively stable individual characteristic over time, it may be influenced by prior language experience, particularly in childhood" (Harley & Hart, 1997, p.382).

Skehan (1998, p.188) for his part, proposes that

> "although previous language learning is likely to bring into play beneficial changes for future language learning, there is still an underlying endowment which has not changed, and which acts as a constraint on what is possible in terms of the speed of future learning."

Later Skehan seems to revise his opinion somewhat in saying that

> "the truth of this matter is that there is simply not enough evidence to argue for the stability of aptitude with any certainty, but for now, following Carroll, we will assume that aptitude does not change with the seasons!" (Skehan, 2002, p.79).

Hence, the best conclusion may be that aptitude is relatively constant but experience in learning language is certainly useful.

2.5.2 Components of Aptitude

Most aptitude research stems from Carroll's original work in which he identified four factors of language aptitude (Carroll, 1962):

- Phonetic Coding. This is the ability to "code" auditory phonetic material in such a way that this material can be recognised, identified, and remembered over something longer than a few seconds.

- Grammatical Sensitivity. This is the ability to handle "grammar", i.e. the forms of language and their arrangements in natural utterances.
- Rote Memory for Foreign Language Materials. This is the capacity to learn a large number of associations in a relatively short time. This ability is independent of phonetic coding ability as even those who have the phonetic coding ability may still not be able to hear and remember relationships.
- Inductive Language Learning Ability. This is the ability to infer linguistic forms, rules and patterns from new linguistic content itself with a minimum of supervision or guidance.

Auditory abilities were later added by Carroll as another category of abilities important for language learning (Carroll, 1990). This refers to the ability to understand speech which is unclear or under conditions of masking. Auditory abilities identified in Stankov & Horn (1980) include Speech Perception Under Distraction, Temporal Tracking (an ability to recognise and remember the order in which particular sounds occur), Maintaining and Judging Rhythm and Discrimination among Sound Patterns. Speech Perception Under Distraction may be measured by tests such as "Talk Masking" and "Cafeteria Noise Masking". These tests are similar in design to each other and involve the subject identifying words despite background noise.

Interestingly, an auditory ability similar to Speech Perception Under Distraction has recently received attention from those interested in the experience-induced adaptivity of the cognitive processing system, in particular how experience induces adaptations to the system. The ability to pick out particular sounds in spite of other background sounds seems to be particularly developed in orchestral conductors. Conductors have been shown to be particularly good at monitoring the performance of an entire orchestra while simultaneously following single musicians. Their ability in this area is better than pianists and non-musicians and may be due to their prolonged professional experience which has altered the processing of auditory stimuli (Nager *et al.*, 2003). It is noteworthy that this ability has been developed in conductors through training. This suggests that at least the auditory abilities of language aptitude could benefit from training and need not necessarily be constant

over time. Of course it is also possible that even before training, those who would succeed as conductors might display this ability, in which case it could be argued that it is an ability that is not (solely) developed through training but (at least partly) present from the start.

Other auditory abilities include Temporal Tracking, which refers to the ability to recognise and remember the order in which particular sounds occur; maintaining and judging rhythm, and discrimination among sound patterns (Stankov & Horn, 1980, cited in Carroll, 1990). The idea that such auditory abilities might be important for language learning is not a new one. The ongoing belief that music and language processing may require some similar auditory abilities is reflected in the fact that this question was investigated as early as the 1940s (see *Sections 3.3.1, 3.3.2*), and indeed the quest to identify and perhaps even quantify to what extent language aptitude draws on musical or auditory abilities is the primary focus of the current research.

Carroll (1990) also decided that some factor of memory ability should be included in a test of language aptitude. He mentions free recall learning, memory for episodic events, verbal discrimination learning, memory span and tests of delayed memory as areas worthy of further research in relation to language aptitude. Current memory research focuses mainly on Working Memory, which is further discussed in *Section 2.7* below.

According to Carroll (1993), aptitude is separate from achievement, interest and motivation. This contrasts with Pimsleur's (1966) view which treats motivation and intelligence as integral components of aptitude. Where Pimsleur underlines the importance of intelligence, Carroll states that foreign language aptitude measures correlate differently with achievement in language learning than does intelligence. While he admits that intelligence may be the most weighty element in the relationships between many mental abilities (Carroll, 1993, p.27), he notes that "the possible importance of more specialised abilities cannot and should not be ignored". In addition, he refers to his own work (Carroll, 1981) in the area of foreign language aptitude, which he proposes, "strongly suggests that specialized abilities beyond general intelligence play an important role in learning a foreign language". Current

research suggests that Working Memory (WM) may be one of these specialised abilities which plays an important role in language learning. Just how specialised or domain-specific WM may be, or how it may relate to intelligence is examined in *Section 2.7.3* below. *Section 2.6* provides an overview of intelligence research in order to show how a theory of language aptitude may fit into an overall theory of intelligence.

According to Skehan (1986, 1989) the three important aspects of language aptitude are auditory ability, linguistic ability and memory ability. In many ways, these factors are similar to Carroll's factors, insofar as both models refer explicitly to auditory ability and memory ability. Skehan's auditory ability corresponds roughly to Carroll's phonetic coding ability. Skehan's linguistic ability encompasses both grammatical sensitivity and inductive language learning ability. In Skehan's view, memory ability is more complex than merely forming associations. This simply reflects an updated view of memory, according to which it involves both coding and retrieval skills.

Later Skehan (1998) proposed that the three components of aptitude could be viewed in terms of input, central processing and output skills. On this view, using Carroll's terminology, 'phonetic coding ability' relates to the input stage, 'grammatical sensitivity' and 'inductive language learning ability' constitute central processing abilities, while memory ability determines how new material will be stored and retrieved. In this way, memory determines output. Also memory is essential for output as it is now known that both learners and native speakers alike rely heavily on chunks of language and will generally use lexicalised chunks where possible rather than generating novel sequences (see Skehan, 2002).

To summarise, it may be said that language aptitude seems to consist primarily of memory ability, the ability to process sounds and the ability to manipulate language chunks. The research carried out for this thesis focuses on the first two of these components, i.e. the ability to deal with sounds and memory ability. The language aptitude tests used for this study are discussed in *Sections 4.2.2* and *4.3.2* below.

2.5.3 Other Approaches to Aptitude

Oxford (1990) suggests that learning styles and learning strategies may be components or at least potential correlates of language aptitude but admits that there are relatively few studies in this area and that results are inconsistent (see e.g. Bialystok, 1981; Politzer, 1983). However, learners' abilities to use language learning strategies may, in the future, emerge as significant components of aptitude.

A different approach to aptitude, which appears to be somewhat more promising, is that of Robinson (2001, 2002a). Drawing on the work of Snow (1987), Robinson presents the idea of 'aptitude complexes', according to which there is a hierarchical structure of abilities necessary for processing during SLA. Primary abilities such as pattern recognition and speed of processing combine to form sets of second order abilities (e.g. noticing the gap, memory for contingent speech, deep semantic processing) which support language learning. Groups of abilities are called aptitude complexes which "represent hypothesized combinations of aptitude variables that jointly influence learning in some particular situation" (Snow, 1994, p.9, cited in Robinson, 2001, p.372). One example of an aptitude complex is aptitude for explicit rule learning which consists of the secondary abilities of metalinguistic rule rehearsal (MRR) and memory for contingent text (MCT). A learner may be high or low in metalinguistic rule rehearsal ability and also high or low in memory for contingent text. This results in four possibilities, i.e. that the learner is high in both MRR and MCT, low in both, high in MRR and low in MCT or low in MRR and high in MCT. Thus, a learner who is high in both has a good ability for explicit rule learning but a learner who is low in memory for contingent text may simply not have the memory ability required to memorise the rule despite having good rule rehearsal abilities. The advantage of the aptitude complexes approach is that learners may be profiled and matched to suitable instructional methods.

Grigorenko, Sternberg & Ehrman (2000) propose that successful language learning depends on the ability to cope with novelty and ambiguity, so this should form part of a theory of aptitude. This view of aptitude comes from new findings from intelligence research. Rather than viewing intelligence as simply analytical ability, Sternberg (1985, 1997, 2002) extends his definition to include analytical, creative

and practical aspects (see *Section 2.6.3* below). Applying his work to the area of SLA has resulted in the Canal-F theory (Cognitive Ability for Novelty in Acquisition of Language – Foreign) which considers a learner's capacity to deal with novelty.

According to the Canal-F theory, the processes involved in SLA are (Grigorenko *et al.*, 2000, p.392):

1. Selective encoding: used to distinguish between more and less relevant information for one's purposes as this information arrives in the stream of incoming data.
2. Accidental encoding: used to encode background or secondary information and to grasp the background context of the information stream.
3. Selective comparison: used to determine the relevance of old information for current tasks.
4. Selective transfer: used to apply decoded or inferred rules to new contexts and tasks.
5. Selective combination: used to synthesise the disparate pieces of information that have been collected via selective and accidental encoding.

These five "knowledge acquisition" processes operate at four levels: lexical, morphological, semantic and syntactic. In addition, there are two modes of input and output: visual, which is the primary mode in reading and writing, and oral, which is used in speaking and listening. Finally, recall is either immediate, which occurs right after learning takes place, or delayed, which occurs long after learning takes place. Individual differences can arise in the knowledge acquisition processes, levels of operation, modes of input and output, or recall ability. The Canal-F theory has led to the development of a successful aptitude test which is described in *Section 2.5.5*.

That the processes involved in SLA should drive foreign language aptitude research is also advocated by Peter Skehan who examines SLA processing stages and the aptitude components necessary at each stage (Skehan, 2002). This is quite a new approach in aptitude research but appears promising as it highlights the areas of aptitude for which aptitude tests already exist, in addition to the SLA processes

which require specific types of aptitude for which no tests are currently available. It also helps to overcome the problem of early aptitude research where the purpose was simply to predict success in training programs rather than to develop an understanding of the types of ability needed to master a foreign language.

The work of Skehan and Robinson looks set to produce promising results for language learners. Skehan's examination of SLA results in a better understanding of the processes involved in learning a second language and the abilities required at each stage. If learners can then be tested for these abilities and classified according to their aptitude complexes, suitable instructional methods can be developed for the various types of learner.

2.5.4 Aptitude and Age

A study by Skehan (1980, 1982) conducted with British Armed Forces personnel revealed that successful language learners achieved their success in one of two ways (Skehan, 1998, p.193):

1. High linguistic-analytic abilities, whereby learners seemed to be treating the task as a linguistic puzzle-solving task.
2. Success based on memory, whereby learners considered language learning a task requiring commitment to memory of large amounts of material.

Interestingly, few learners were high in both verbal aptitude and memory. Generally, success came from just one of these sources. This may relate to the fact that different learners have different cognitive styles, with some types of learners more likely to use memorization as a learning strategy (Ehrman & Oxford, 1990). Another possibility is that the difference may be age-related.

Research carried out by Harley & Hart (1997) showed that the age at which adolescents started an immersion program determined which components of aptitude were associated with L2 success, i.e. whether a relationship held between L2 outcomes and analytical skills or between L2 outcomes and memory abilities. It was found that younger learners use memory skills to a greater extent while older learners use analytical skills. The findings of DeKeyser (2000) concur with these results as he shows that a high level of verbal analytical ability is essential for adult

learners to reach native speaker competence whereas verbal analytical ability is not a significant predictor of success for childhood SLA. In his study, very few adult immigrants scored within the range of child arrivals on a grammaticality judgement test, and those who did had high levels of verbal analytical ability (DeKeyser, 2000, p.499). Another similar study shows high language aptitude scores among adult learners of Swedish deemed native-like by native speaker judges while language aptitude scores among young learners and native speaker controls are normally distributed (Abrahamsson & Hyltenstam, 2004). Thus, it seems that language aptitude and specifically language analytic ability may become more important as the learner gets older.

2.5.5 Language Aptitude Tests

Two well known language aptitude tests are the Modern Language Aptitude test (MLAT) (Carroll & Sapon, 1955) and the Pimsleur Language Aptitude Battery (PLAB) (Pimsleur, 1966). The difference in opinion between Carroll and Pimsleur as to what constitutes aptitude is reflected in the difference between these two tests. The MLAT examines Number Learning, Phonetic Script, spelling clues, words in sentences and Paired Associates. Number Learning requires the subject to learn a set of numbers through aural input and then discriminate different combinations of these numbers. Phonetic Script requires the test-taker to learn a set of correspondences between speech sounds and phonetic symbols. In Spelling Clues, the test-taker must read words that are spelled as they are pronounced, rather than according to standard spelling conventions. They must then select from a list of words the one whose meaning is closest to the "disguised" word. The Words in Sentences test examines awareness of grammatical structure. A key word is marked in one sentence and the test-taker must select from another sentence a word that functions in the same way as the key word. Finally, in the Paired Associates test, the subject must quickly learn a set of vocabulary from another language and memorise their English meanings. Thus, the first three of Carroll's original four language aptitude components are measured by at least one of the MLAT subtests. The fourth (inductive language learning ability) is reflected only weakly in the MLAT. The MLAT was not updated to reflect Carroll's ideas on auditory ability and memory aspects of aptitude.

Pimsleur's Language Aptitude Battery consists of six parts:

- Grade Point Average
- Interest
- Vocabulary
- Language Analysis
- Sound Discrimination
- Sound-symbol correspondence

These six parts are designed to take into account the three components of language aptitude that exist according to Pimsleur (1962), i.e. verbal intelligence, motivation and auditory ability.

Despite the differences between the MLAT and the PLAB, Wesche (1981) has shown that both tests make a consistent and substantial contribution to the prediction of student achievement in a variety of adolescent and adult language training programmes. In general, according to Dörnyei and Skehan (2003, p.589), correlations of aptitude or motivation with language achievement range between 0.20 and 0.60, with a median value of approximately 0.40. Others have cited correlations as high as 0.73 between aptitude tests and measures of proficiency (Grigorenko *et al.*, 2000). However, a criticism levelled at the MLAT by researchers, including Dörnyei and Skehan, is that in designing the MLAT, understanding of the components of aptitude was sacrificed in order to improve predictive validity.

Grigorenko *et al.* (2000) attempt to overcome this criticism by presenting a new test based on the CANAL-F theory discussed above (*Section 2.5.3*). It is designed to be of predictive value but also to contribute to the understanding of what constitutes language aptitude. It is based on a simulation in which subjects learn elements of an artificial language. It comprises nine sections, five of which involve immediate recall and four of which were identical to these five sections except that they involve delayed recall. The five sections are (Grigorenko *et al.*, 2000, p.394):

1. Learning meanings of neologisms from context (immediate and delayed recall).

2. Understanding the meaning of passages (immediate and delayed recall).
3. Continuous paired-associate learning (immediate and delayed recall).
4. Sentential inference (immediate and delayed recall).
5. Language learning rules (only immediate recall questions).

A comparison between the CANAL-FT and the MLAT revealed a correlation of 0.75 ($p<0.001$) between the two. One major difference between the CANAL-FT and traditional tests is the use of dynamic testing in the CANAL-FT. This gives the test-taker the opportunity to learn at the time of testing. Where other tests emphasise that the administrator should give no assistance, dynamic testing investigates to what extent a subject can profit from teaching. It would seem beneficial to measure this ability as it may provide clues as to those students who would benefit from a course of instruction.

2.5.6 Relationship between Language Aptitude and Other Cognitive Skills

In his exploration of the relationship between language aptitude and intelligence, Cummins (1979) introduces a distinction between Basic Interpersonal Communication Skills (BICS) and Cognitive Academic Language Proficiency (CALP). BICS refers to the skills required for oral fluency and sociolinguistic appropriateness. Such skills are believed to develop naturally as a result of exposure to a language through communication. CALP proficiency consists of the linguistic knowledge and literacy skills required for academic work. Ekstrand (1977) found low-level correlations between intelligence and proficiency as measured on tests of listening comprehension and free oral production, but much higher correlations when proficiency was measured by tests of reading comprehension, dictation and free writing. The conclusion that can be drawn from this and other studies is that CALP is connected to intelligence. Genesee (1976) also found that intelligence is strongly related to the development of academic L2 French language skills such as reading, grammar and vocabulary but was largely unrelated to ratings of oral productive ability.

Not everyone is in agreement as to the existence of a special aptitude for language acquisition at all. Neufeld (1978) proposes that all humans are equipped to master basic language skills but that they vary in their ability to master higher-level skills. This variation is determined by intelligence. Oller and Perkins (1978) also dispute the existence of a special aptitude for language. They posit the existence of a general factor which accounts for most of the variance in a wide variety of language proficiency measures. This 'g' factor is the same as a general or 'g' factor of intelligence (see *Section 2.6.2* below).

However, it now seems unlikely that general intelligence and language aptitude are completely synonymous. Correlations between intelligence and aptitude range from low/moderate (e.g. R. C. Gardner & Lambert, 1965) to moderate/strong (e.g. Sasaki, 1996). A correlation of 0.53 ($p<0.001$) was found between the CANAL-FT and a test of crystallised intelligence, while a correlation of 0.58 ($p<0.001$) was found between CANAL-FT and fluid intelligence (see *Section 2.6.2* below for explanation of the terms crystallised and fluid intelligence). A construct validity analysis leads to the conclusion that "the CANAL-FT is a valid measure of FL aptitude, which, as expected, is related but not equivalent to both crystallized and fluid abilities" (Grigorenko *et al.*, 2000, p.399). So it seems that success on the CANAL-FT depends on two factors – an intelligence-related factor and a language-specific factor.

A study by Sasaki (1993) which looks at second language proficiency, foreign language aptitude, verbal intelligence and reasoning ability also suggests that second language proficiency is not identical to general cognitive ability or intelligence. However, these two factors are correlated ($r=0.65$). Her study reveals that less than half (42%) of the variance in second language proficiency could be explained by the general cognitive factor. The other 58% of the variance in second language proficiency must be accounted for by something other than general cognitive ability. She also considers other studies in which correlations were found between aptitude, intelligence and second language proficiency. She proposes that aptitude and intelligence may have appeared to be correlated with second language proficiency factors, not because they were all affected (or caused) by the same factor, but because the two different factors influencing them (General Cognitive

factor and General Second Language Proficiency factor) were correlated with each other (Sasaki, 1993, p.338). In the best fitting model, foreign language aptitude, verbal IQ and reasoning, all have high positive loadings on the general cognitive factor, while specific language tests had positive loadings on a second language proficiency factor. If indeed it is the case that aptitude is related more to general cognitive factors than language factors, it is difficult to explain the greater predictive power of language aptitude tests over general intelligence tests.

The relationship between language aptitude and intelligence requires further research before any definitive conclusions can be drawn. In order to allow for the possibility of a confounding effect of intelligence on language aptitude, subjects in the current study are tested for general fluid intelligence and correlations between intelligence and language aptitude are considered (see *Section 4.2.2.3*).

2.5.7 Importance of Aptitude in Different Contexts

One question posed by Dörnyei & Skehan (2003, p.591) asks if language aptitude always applies in a similar manner, without influence of learning context (e.g. foreign language versus second language context), learning methodology and L1 to L2 combination? It has been proposed that aptitude tests are methodology dependent and are only useful in formal learning contexts (Cook, 1996; Krashen, 1981). Larsen-Freeman and Long (1991, p.169) hypothesise that language aptitude tests may work well to predict success because of a concordance of tasks between the test and formal classroom study, and not because the test is measuring some innate linguistic ability. This view contrasts with that of Reves (1983) who showed that aptitude functions as an effective predictor both in second language (acquisition-rich environment) and foreign language (acquisition-poor) settings. Skehan (1989) suggests that aptitude is equally important in second language contexts because material is not presented to learners in such a structured way, and so they are forced to rely on their own aptitude to make sense of the language. Others, too, highlight the importance of aptitude in untutored learning situations, mentioning the fact that much research has shown that as more of the information-processing burden is put on the student, aptitude becomes a more important predictor of successful learning (DeKeyser, 2000). DeKeyser shows that aptitude is a predictor of ultimate attainment in L2, even after long periods of exposure in non-

tutored contexts. Hence, it now seems that aptitude is important in more than formal language classrooms.

2.5.8 Conclusion

Research in language aptitude has been relatively limited since the early efforts of John Carroll – efforts, which over forty years ago, yielded an aptitude test in the form of the MLAT with high predictive power. While to date this remains one of the most widely used language aptitude tests, new ideas have yielded promising results, e.g. in the form of the CANAL-FT (Grigorenko *et al.*, 2000). As the emphasis of the early language tests was on predictive power, the nature of language aptitude remains somewhat unclear although it seems that aptitude is relatively stable over time and is of use both in formal and informal learning contexts. Research in the area of Working Memory suggests that aptitude is closely related to Working Memory span. This hypothesis is considered in the section below on Working Memory (see *Section 2.7*).

To summarise the findings on the components of aptitude, it is useful to use Skehan's three way division of aptitude into auditory ability, memory ability and linguistic ability as it seems to adequately account for the various factors found in numerous different studies. To this perhaps should be added the ability to deal with novelty. While Skehan's model does not include intelligence, the contribution of intelligence to aptitude is disputed. Similarly, it does not include learner styles and strategies, but again there is as yet no conclusive evidence that these should be considered components of aptitude.

As language learners have different levels of skill in the different aptitude components, it has been shown that exploiting these differences can lead to improved results in achievement and attitude (Wesche, 1981). It appears that it may be helpful to group learners according to their scores on tests such as the Modern Language Aptitude Test (MLAT) or the Pimsleur Language Aptitude Battery (PLAB) or according to their aptitude complex profiles and use a method of instruction which corresponds to their particular aptitude strengths.

Twenty years ago, Skehan (1985, p.17) concluded that there were two main points to be made about language aptitude. His first point emphasised that language aptitude "is a hybrid, combining both a language processing ability as well as the capacity to handle decontextualised material". The fact that language aptitude is not a single construct is still important, in that while we may still be uncertain as to its exact components, modern research continues to show that it is indeed hybrid in nature (e.g. Robinson, 2001, 2002a; Skehan, 1998, 2002).

Skehan's second observation that "the analysis of aptitude presented here is general enough to be relevant not simply to formal learning situations, but also to more communicatively oriented classrooms as well as 'acquisition settings'", has been disputed but nonetheless would still seem to hold.

Foreign language aptitude is dealt with again in *Sections 4.2.2.2* and *4.3.2.2*, where details of the language aptitude tests used in the current study are presented. Hopefully, the discussion in this section has highlighted some of the difficulties inherent in aptitude research, e.g. the difficulty in choosing the components of language aptitude and the problems involved in attempting to measure them. It will be seen later that certain problems emerge in Phase II of the current study, where productive tests are used to measure foreign language aptitude (see *Sections 5.3.3.1* and *5.6*). It proves difficult to provide definitive answers to the questions which arise in this context; i.e. why do all components of the test not correlate with each other? However, given the discussion in the present section which highlights how much remains to be learned about foreign language aptitude, it is perhaps not surprising that a number of issues emerge in this thesis which require further research.

It was noted above that some researchers consider intelligence to be a component of aptitude. Given the close relationship between aptitude and intelligence, intelligence serves as the focus of the next section. As in the case of foreign language aptitude, much remains to be learned about intelligence. The following discussion is intended to be a brief overview of intelligence research.

2.6 Intelligence

It is important to consider what is currently meant by intelligence and to investigate how intelligence is related to other abilities. This is important for the current study as intelligence may be a confounding variable. One possibility is that this study could show a highly significant link between music and language. If it were simply the case that more intelligent individuals perform better in music and language tests than less intelligent individuals, this is much less interesting than if the relationship between music and language holds independent of general intelligence levels. Therefore, the aim of this section is to consider the nature of intelligence before embarking on any empirical investigation.

2.6.1 Introduction

While "intelligence" is hard to define, it appears that most people have a personal working understanding of what the term means. We can only assume from the abundance of terms available in everyday language to refer to varying levels of intelligence (e.g. "bright spark", "sharp", "genius", "thick as two short planks"), that the concept must be of significance in our daily lives.[8] Tests of intelligence are used to determine entry to particular schools and third-level educational institutions. They are also used by some companies for recruitment purposes.

Individual differences in intelligence seem to have been recognised since early times; e.g. the Chinese used tests of mental abilities as early as 2200B.C. to select talented individuals to serve as civil servants (Fox, 1981, cited in Sdorow & Rickabaugh, 2002). Plato, in his "Children of God", urged the Ancient Greeks to seek out gifted children and bring them swiftly into high levels of service to the country (cited in Crocker, 1999).

---

[8] It is, of course, dangerous to assume a direct relationship between the number of words, which exist in a language to describe a particular concept, and the existence of that concept in the community; e.g. from a disputable claim that there are significantly more words for snow in Inuit than in English, it was argued that Eskimo people, because of their language, have concepts that cannot be expressed in English (Whorf, 1940). This is unlikely to be the case because, even in English, there are many words for textures and quantities of snow. Even if there are lexical gaps in English, the concepts can be achieved by phrasal means. In his discussion of Eskimo words, Pullum (1991) points out the example of a printer who has many terms for fonts because they are important to his trade. The significance of the number of words for intelligence is that some spectrum of them exists for nearly all people. A consequence of this is that intelligence must be of significance to most people.

Plato can be read as having believed that important aspects of intelligence are the ability to learn and know facts, to extrapolate from what is known in order to acquire further knowledge, and to remember what is thus learned; e.g. Sternberg (1990, p.24) cites the *Republic* Book 5 (*Great Books of the Western World*, 1987, p.359) in which Socrates asks Glaucon:

> "When you spoke of a nature gifted or not gifted in any respect, did you mean to say that one man will acquire a thing easily, another with difficulty; a little learning will lead the one to discover a great deal; whereas the other, after much study and application, no sooner learns than he forgets ... ?".

The emphasis on intelligence as the ability to learn principles and group their consequences (rather than learn large sets of forgettable facts) is particularly interesting as it relates to the modern understanding of intelligence; e.g. dynamic testing, mentioned in *Section 2.5.5* above, is based on the premise that intelligence is the ability to learn. Autonomous learning (Holec, 1981; Little, 1991; Little *et al.*, 2003), another modern concept, also seems related to Plato's notion of the student extrapolating from one situation to another. Little (1991, p.4) defines autonomy as a capacity for:

> "detachment, critical reflection, decision-making, and independent action. It presupposes, but also entails, that the learner will develop a particular kind of psychological relation to the process and content of his learning. The capacity for autonomy will be displayed both in the ways the learner learns and in the way he or she transfers what has been learned to wider contexts."

Therefore, we see that critical reflection and the ability to transfer learning from one context to another are features of the autonomous learner. Very similar attributes seem to characterise Plato's intelligent person.

In trying to arrive at a definition of intelligence, Cooper (1999, p.6) suggests that: "Abilities, in the broad sense, are any behaviours that can be sensibly evaluated." He includes such abilities as typing, knowledge of steam locomotives, sprinting, map reading, swindling people and cooking in the list of abilities that can be sensibly evaluated. However, some of these abilities do not seem to represent the kinds of capabilities that are traditionally considered to be "intelligences". Typing for example would not generally be considered to be a distinct "intelligence", even though a person's typing capacity can be measured easily. In order to account for

this, Cooper proposes that abilities depend to differing degrees on specific knowledge, attainment or training, in addition to physical prowess, thought processes and emotional skills. Research on intelligence focuses primarily on those abilities which depend on thought processes rather than on specific knowledge, physical prowess or emotional skills. (However, for a different view which assigns a central role to the emotions in intelligent behaviour, see for example Goleman (1996) discussed in *Section 2.6.3* below).

Other definitions of intelligence include

> "the sum of all the knowledge and skill that we have acquired during our lifetime, and which we use to tackle the daily problems that we face" (Crocker, 1999, p.2),

while the psychologist David Wechsler proposed intelligence to be:

> "the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment"
>
> (Wechsler, 1958, p.7).

Crocker's definition emphasises the idea that intelligence is not fixed at birth but rather develops through interaction with the environment. This is firmly in line with Piaget's philosophy of child development (see e.g. Piaget, 1970, 1972).

Piaget proposes that our experiences change the way in which we think about the world, i.e. our experiences alter our schemas. Schemas are mental models of the characteristics of persons, objects, events, procedures or situations. Schemas are open to change through accommodation or assimilation. Assimilation occurs when new information is fit into our existing schemas while accommodation takes place when our schemas are revised to fit new information. Thus, we assimilate when new information does not contradict something which is already known, whereas accommodation requires us to reconsider and modify currently held beliefs in order to incorporate new information.

According to Piaget's teaching, a child's way of thinking is qualitatively different from that of an adult. It is only when the child has passed through the various stages of development that he has the capacity to think like an adult. Piaget's stages

of development are the Sensorimotor stage, Preoperational stage, Concrete Operational stage and the Formal Operational stage.

The Sensorimotor stage is claimed to take place between birth and two years, during which time the child moves from carrying out only reflexive, instinctual actions to the beginning of symbolic thought. The preoperational stage is said to occur between the ages of two and seven years. During this time the child begins to represent the world with words and images which reflect increased symbolic thinking. Between the ages of seven and eleven years, the child is said to enter the concrete operational stage. By now, the child can supposedly reason logically about concrete events and can mentally reverse information. The formal operational stage relates to abstract, idealistic and logical ways of thinking. It is thought possible that not everybody reaches the formal operational stage, particularly those from cultures which do not stress scientific thinking in their educational systems.

It has been suggested that reaching the formal operations stage may negatively affect SLA (Krashen, 1982), as it allows the learner to use abstract thought, which may not be as effective as the Language Acquisition Device available to younger learners. As noted in Chapter 1, the term Language Acquisition Device (LAD) was introduced by Chomsky (1964) to describe a hypothetical "black box" containing the fundamentals of language which would allow children to learn their native language despite imperfect and incomplete input. Krashen suggests that once the formal operations stage is reached, the LAD becomes unavailable. If indeed it were the case that after a certain maturational point, L2 learning is no longer subserved by the same mechanisms that are involved in child language acquisition, it would provide support for the Age Factor or "Critical Period Hypothesis" (Lenneberg, 1967; Penfield & Roberts, 1959), often used to account for the greater success of children over adults in second language learning. While the Age Factor is probably one of the individual differences which has received most attention in the literature, the Critical Period Hypothesis has now, to a large extent, been discredited (see e.g. Bongaerts *et al.*, 1995; Perani *et al.*, 1998; Singleton, 2001; Singleton, 2003, 2004). Research has shown that principles of Universal Grammar, not instantiated in the

L1, actually remain available to adult L2 learners (Marthohardjono & Flynn, 1995; Marthohardjono & Gair, 1993).

Another point of interest in relation to the Language Acquisition Device and the attainment of the formal operations stage, concerns the "Din in the Head" phenomenon. This refers to the phenomenon whereby people in a foreign country 'hear' the foreign language in their heads, after a period of time in contact with the language. The Din in the Head was first described by Barber (1980, pp.26*ff.*). She describes her experiences of travelling in Eastern Europe and explains how

> "by the third day ... the linguist in me was noticing a rising Din of Russian in my head: words, sounds, intonations, phrases, all swimming about in the voices of the people I talked with. ...The sounds in my head became so intense after 5 days that I found myself chewing on them, like so much linguistic cud, to the rhythm of my own footsteps as I walked the streets and museums. .... The constant rehearsal of these phrases of course was making it easier to speak things quickly; things popped out as prefabricated chunks."

Krashen (1983, p.41) further developed the idea of the Din, suggesting that it is a "result of stimulation of the Language Acquisition Device (LAD)". If indeed the din is a "result of stimulation of the Language Acquisition Device", and it occurs in both adults and children, the LAD must be accessible to both. Therefore, it is surely a contradiction to say that the LAD is not available after the formal operations stage has been reached.

The ideas mentioned above regarding the interaction of intelligence with the environment, and the modification of intelligence as new concepts are learned, are related to Sternberg's view of intelligence discussed below in *Section 2.6.3*. Another aspect of intelligence which is often discussed is its heritability; e.g. Horn and Cattell (1966) point to the heritability of fluid intelligence (see *Section 2.6.2* below for explanation of fluid intelligence). Numerous twin studies have shown that identical twins, even those separated at birth, have a very similar level of intelligence (see e.g. Plomin *et al.*, 1990; R. C. Wilson, 1983). Cooper (1999, p.88) concludes that "genetic makeup has the most potent influence on levels of general ability in western society (accounting for 50-60 percent of the variation)". On the

other hand, however, he does not deny the important role of environmental influences in determining intelligence.

2.6.2 Early Measurements of Intelligence

Although it is mentioned above that both the ancient Greek and early Chinese societies had views on intelligence, the early 1900s saw the beginning of formal intelligence testing taking place in Western society. Psychologist Alfred Binet, working in France, developed a test in 1905 to identify children with special educational needs (Binet & Simon, 1905a, 1905b). In order to determine what was normal for any particular age group, he examined large numbers of children and their capacity to perform different tasks at different ages. Children capable of performing tasks normally only completed by their older peers were deemed advanced, while children unable to perform tasks normal for their age group were described as retarded. Binet's tests were used to assign a 'mental age' to each child; e.g. a child who passed the 7-year-old test but failed the 8-year-old test had a mental age of 7, regardless of his/her chronological age.

William Stern (1912) proposed the ratio of mental age to chronological age to determine a child's level of intelligence (Kreppner, 1992, cited in Sdorow & Rickabaugh, 2002). Thus, a 10 year old with a mental age of 8 has an intelligence level of 0.8 (i.e. mental age/chronological age = 8/10 = 0.8), whereas a 10 year old with a mental age of 12 has an intelligence level of 1.2 (i.e. mental age/chronological age = 12/10 = 1.2). When Lewis Terman eliminated the decimal point by multiplying the ratio by 100, it became known as the Intelligence Quotient (IQ). The definition of IQ is therefore:
(mental age/chronological age) * 100.

This ratio IQ was later replaced by David Wechsler with a deviation IQ which compares a person's intelligence score with the mean score of his or her age peers, defined as the individual's actual score on a test relative to the expected average score obtained by people of the same age (Sdorow & Rickabaugh, 2002, p.29). This overcomes the fact that the mental age of adults does not continue to increase indefinitely. A consequence of this ceiling effect is that with the ratio IQ, adults could be considered below average intelligence simply because their chronological

age has increased more than their mental age. This problem does not arise with the deviation IQ.

At much the same time as Binet was working in France, Charles Spearman was investigating intelligence in Britain (Spearman, 1904). He was responsible for the development of the statistical technique of factor analysis. In factor analysis, results from a number of tests are examined and correlations between the different tests are calculated. Those tests which show consistently high correlations are grouped together to form a 'factor', as it is believed that some common underlying ability must account for the correlation. In conducting factor analysis of different tests, Spearman proposed that performance on all the tasks depends on the operation of a single underlying factor which he called 'g' or general intelligence.

There is now widespread support for the notion of some sort of general intelligence which underlies success in all cognitive tasks. Jensen (1997, p.122) notes that

> "since its discovery by Spearman in 1904, the 'g' factor has become so firmly established as a major psychological construct in terms of psychometric and factor analytic criteria that further research along these lines is unlikely to disconfirm the construct validity of 'g' or to add anything essentially new to our understanding of it."

However, at the time not all researchers supported Spearman's proposal of a general intelligence factor. Nor indeed today is the concept without its critics (see *Section 2.6.3* below). One of the dissidents, Louis Thurstone (1931, 1938, 1940) was working on intelligence research in the United States, at much the same time as Spearman was working in Britain. Using a different method of factor analysis from Spearman's, he identified seven factors which he called primary mental abilities. These were reasoning, word fluency, perceptual speed, verbal comprehension, spatial visualization, numerical calculation and associative memory. He concluded that scores on these tests did not correlate highly enough to assume the existence of a general factor of intelligence.

Cooper (1999) addresses possible reasons for the diverging findings of Spearman and Thurstone. One of the primary reasons relates to the different types of test used by the two researchers. Spearman used an overall measure of mathematical ability

in his analysis by adding together scores on addition, subtraction, multiplication and division problems. Similarly, he used an overall measure for verbal and other tests. Thurstone however, treated addition, subtraction, etc. as separate tests and his analysis showed that these tests formed a factor. Cooper suggests that the *factor* of numerical ability identified by Thurstone is almost identical to the *test* of mathematical ability used by Spearman (emphasis added). Thus, there may actually be no conflict between the two. Indeed, if Thurstone had done further factor analysis he may actually have found similar results to those of Spearman, i.e. further factors may have emerged such as a general factor underlying his primary factors.

Work in this direction was carried out by John Carroll (1993) who re-examined over 450 data sets, in order to determine the factors of intelligence. His findings support the existence of 'g' or general intelligence. Primary factors, e.g. speed of reasoning, memory span, grammatical sensitivity and spelling ability, exist at the first level of analysis which he calls Stratum I. In the studies examined, he found very many primary factors. Factor analysis of the primary factors reveals eight secondary factors (Stratum II), including Fluid Intelligence (gF) and Crystallised Intelligence (gC) which were originally proposed by Cattell (1971) and Horn and Cattell (1966).

Fluid intelligence and crystallised intelligence are often regarded as the most important secondary factors and certainly receive most attention in the literature, probably due to the fact that they are the two factors about which there is greatest consensus among researchers. Fluid intelligence corresponds roughly to non-verbal reasoning while crystallised intelligence corresponds roughly to verbal intelligence. More specifically, fluid intelligence reflects "thinking ability, memory capacity and speed of information processing" (Sdorow & Rickabaugh, 2002, p.315). It reflects an individual's capacity for novel thinking. The Raven Progressive Matrices test is a commonly used test of fluid intelligence which involves subjects examining novel patterns and selecting a smaller pattern to complete the overall pattern (Raven *et al.*, 1998). Fluid intelligence, according to Mackintosh (1998, p.267) represents

> "the biological potential for intelligence, which becomes manifest in gC, as people acquire knowledge and experience, absorbing the culture in which

> they live and profiting, or not as the case may be, from the educational system their society provides."

Thus, it can be seen from Mackintosh's definition that crystallised intelligence is considered to be close to "school intelligence", i.e. our ability to learn and remember facts. Our ability therefore to acquire skills and knowledge through education and experience depends on crystallised intelligence. Indeed differences in gC may actually reflect differences in the way people take advantage of their opportunities. Carroll (1993) shows that gC includes not only factors on verbal ability, reading comprehension and general information but also a factor of numerical ability.

Mackintosh (1998, p.266) suggests that

> "most factor analysts are agreed on one broad distinction – that between crystallized and fluid intelligence, gC and gF. To these broad factors, most would add further factors of spatial visualization (Gv) and perceptual speed (Gs), and many would add several more; e.g. an immediate memory or short-term acquisition and retrieval factor (Gm) and an auditory perception factor, which we could label Ga."

Other secondary factors identified by Carroll are General Learning and Memory; Broad Visual Perception; Broad Auditory Perception; Broad Retrieval Ability; Broad Cognitive Speediness and finally Processing Speed. The third level in Carroll's analysis is Stratum III at which a general factor of intelligence, or 'g', occurs. Deary (2001, p.15) notes that there is

> "no longer any substantial debate about the structure of human mental ability differences. Something like John Carroll's three-stratum model almost always appears from a collection of mental tests. A general factor emerges that accounts for about half of the individual differences among scores for a group of people, and there are group factors that are narrower abilities and then very specific factors below that."

Not all researchers agree with this conclusion however. Much modern day research aims to broaden the scope of intelligence research and to include qualities not traditionally recognised as intelligences. This is addressed in the following section.

2.6.3 Current Views on Intelligence

Increasingly it is believed that traditional views of intelligence are too narrow; e.g. Robert Sternberg (1985, 1990) has proposed a triarchic theory according to which intelligence consists of three components. These relate to the internal world of the individual (componential intelligence), the external world of the individual (experiential intelligence) and the adaptation of the individual's internal world to the requirements of the external environment (contextual intelligence). Componential intelligence examines the mental mechanisms underlying intelligent behaviour. These are internal to the individual. In this way, it is the component most closely related to traditional views of intelligence as it considers how information is processed by the individual, the strategies used in reasoning and how strategies are executed. In his later work, Sternberg refers to this as analytical intelligence (Sternberg, 2002).

For Sternberg, it is not enough to merely examine how individuals process information as he says that "there is more to intelligence than a set of information-processing components" (Sternberg, 1990, p.274). This leads him to consider the world outside the individual. In order to consider the other aspects of intelligence, he examines firstly experiential intelligence, which refers to the external world of the individual, i.e. experiential intelligence is "the use of these mental mechanisms in everyday life in order to attain an intelligent fit to the environment" (*ibid.*, p.268). Experiential intelligence relates to an individual's ability to use past experiences in solving a novel problem, e.g. a reasoning problem or in the design of an advertisement. Sternberg (1997, 2002) later describes this ability to apply experience in creative ways as creative intelligence.

Sternberg points to the research on fluid intelligence as providing support for the idea that the ability to deal with novelty is a good way of measuring intelligence. As mentioned above, fluid intelligence involves dealing with novel materials, e.g. solving pattern puzzles in the Raven Progressive Matrices test. In the psychometric tradition discussed above, tests of fluid intelligence are shown to have a particularly close relationship to 'g', general intelligence (see Carroll, 1993). Therefore, it appears to be the case that the ability to apply past experiences to a novel situation

is an important aspect of intelligence. While Sternberg supports the importance of the ability to deal with novelty as a component of intelligence, thus indirectly supporting gF, he rejects the notion of 'g', the general intelligence factor, on the grounds that his studies show only a very weak general factor. He suggests that the general factor is only relevant when a narrow range of abilities is measured (Sternberg, 2002).

The final component of Sternberg's theory is contextual intelligence (later called practical intelligence) which considers "how thought is intelligently translated into action in a variety of different contextual settings" (Sternberg, 1990, p.280). In other words, this component examines how the individual acts in different real-life situations. Individuals may adapt to their environment as it is or they may attempt to shape their environment so that it becomes more like they want it to be. If neither adaptation nor shaping produce the desired results, an individual may select a different environment. Sternberg suggests that it is through adaptation, shaping and selection that the components of intelligence become actualised in the real world.

An alternative, and perhaps better known attempt to broaden the definition of intelligence comes from the work of Howard Gardner (1983, 1993). Gardner has proposed a theory of multiple intelligences which recognises linguistic intelligence, logical-mathematical intelligence, musical, bodily-kinaesthetic, spatial, interpersonal and intrapersonal intelligences. In proposing the term 'multiple intelligences' rather than simply a single intelligence factor which comprises multiple components, Gardner attempts to accord equal importance to a wide range of skills, many of which are normally considered talents rather than intelligences. Musical ability, and bodily-kinaesthetic ability which involves the ability to use the body or portions of it in solving problems such as dancing, acting or performing surgery, are generally regarded as talents rather than abilities. Linguistic intelligence, i.e. the abilities required in language production and comprehension, and logical-mathematical intelligence, i.e. the abilities required in numerical computation and logical thinking, are undoubtedly the two areas most valued as intelligences in western society. While of course the use of terms such as intelligence and talents is simply an issue of terminology, Gardner's aim in using

the word intelligence is to acknowledge that these other abilities are of equal importance for functioning in society.

Two of his other intelligences - interpersonal and intrapersonal - are of great importance in everyday life. Interpersonal intelligence refers to the ability to understand other people, their moods and feelings. It is an essential ability for people such as teachers, counsellors and those in sales as they need to relate well to the individuals with whom they are dealing. Intrapersonal intelligence involves an individual's ability to understand his/her own emotions and feelings and to act in ways that further his/her goals and ambitions.

One criticism levelled at Gardner's array is that his criteria for identifying 'intelligences' would actually allow the inclusion of many more intelligences than he has actually proposed (see e.g. Brody, 1992; Sternberg, 1990). Indeed Gardner himself has said that his list of seven is not exhaustive and has recently added naturalistic and spiritual intelligences as additional candidates for consideration (H. Gardner, 1999). Two of Gardner's criteria for identifying an intelligence are that firstly, the ability should depend on identifiable brain regions and that secondly, there should be large differences between individuals. However, this logic would allow the ability to identify faces to qualify as an intelligence, given that certain brain regions are responsible for visual recognition and also that some people are much better at this task than others, i.e. there are large differences between individuals on this task. Most theories of intelligence would not wish to consider the ability to identify faces as an intelligence, leading researchers such as Sternberg (1990) to conclude that Gardner's theory is over-inclusive and too vague.

Despite the criticisms, Gardner's theory has certainly been widely cited in the popular press and received much interest from the general public. It has made an important contribution to the discussion of the nature of intelligence as it emphasises the narrowness of traditional approaches. The theory of multiple intelligences is addressed again in Chapter 3 in order to consider how musical intelligence is related to linguistic intelligence.

Other approaches to broadening the definition of intelligence come from researchers working in the area of emotional intelligence (e.g. Goleman, 1996; Mayer & Salovey, 1993). Goleman (1996, p.33) cites the case of an American student who, always having received straight As on exams, stabbed a teacher following the receipt of 'only' 80% on a quiz and asks "How could someone of such obvious intelligence do something so irrational – so downright dumb?". Throughout his work he presents numerous examples of cases where academic intelligence seems unrelated to everyday life whereas emotions are essential for functioning in everyday situations. He suggests that IQ and emotional intelligence are separate concepts and that people may be high or low in one or other or both. He quotes psychologist Martin Seligman who emphasises the importance of reasonable talent in combination with the ability to keep going in the face of defeat as the factors leading to success (Goleman, 1991).

As IQ has been extensively researched elsewhere in the literature, Goleman concerns himself with

> "abilities such as being able to motivate oneself and persist in the face of frustrations; to control impulse and delay gratification; to regulate one's moods and keep distress from swamping the ability to think; to empathize and to hope" (Goleman, 1996, p.34),

which have received much less attention. His work certainly highlights the importance of emotions in successful living and underlines the fact that emotional intelligence may be worthy of greater investigation than it currently receives.

2.6.4 Conclusion

The aim of this section was to present an overview of the main theories of intelligence which have influenced educators and psychologists in the recent past. In addition, some of the newer theories of intelligence, which have attempted to widen the traditional boundaries of intelligence research, have been introduced. The research conducted for the current study considers general fluid intelligence as a possible confounding factor when investigating the relationship between music and foreign language aptitude. The Raven Progressive Matrices test, used as a measure of gF intelligence in the current study, is discussed in more detail in *Section 4.2.2.3*. The discussion above shows that gF intelligence attracts

widespread support from researchers in the field, and, was therefore decided to be a useful measure in the current study. General fluid intelligence is considered in this research, rather than using more broad definitions of intelligence, because in many cases, a consensus has not yet been reached on newer approaches to intelligence. For the purposes of this research, the existence of some sort of 'g' is assumed; i.e. given the amount of support in the literature, it seems reasonable to suppose that some sort of general intelligence factor is likely to exist. The aim of this research is to consider to what extent a relationship exists between music and language aptitude, over and above that which can be accounted for by 'g'. If a relationship is found between aptitude in music and language and, as Sternberg has suggested, no 'g' factor exists which could partly account for this relationship, it simply means that the relationship arises from commonalities in music and language, rather than being influenced by a general intelligence factor. If 'g' should later be discredited, it would simply add support to the finding that there is a special relationship between music and language aptitude, as there is no longer any way to account for the relationship through recourse to a general intelligence factor.

This introduction to intelligence research was also useful in order to better appreciate theories of language aptitude, particularly as some of the most influential researchers in the aptitude domain also investigated intelligence quite extensively, e.g. John Carroll. Moreover, as the following section examines Working Memory, which has also been linked to intelligence, it is quite useful to know something of the research which has taken place in this field to date.

2.7 Working Memory

This section introduces and discusses the term Working Memory. Firstly, different ways of analysing the components of memory are presented (*Section 2.7.1*). This is followed by a discussion of how Working Memory impacts on language learning and language use (*Section 2.7.2*). The relationship between Working Memory and intelligence is then examined (*Section 2.7.3*), before turning to a discussion of some potential components of Working Memory which have not been extensively studied to date (*Section 2.7.4*).

The concept of Working Memory is of particular importance for this thesis as the language aptitude tests used in Phase II of the current study exploit the relationship between Working Memory and language aptitude. As discussed below, Working Memory ability (or at least phonological loop capacity) can be measured by asking subjects to repeat foreign language words. A subject's ability to do this task is significantly related to his/her language aptitude. The tests used in Phase II of this study are based on the premise that tests of phonological loop capacity are significantly related to language aptitude.

2.7.1 Introduction

The term memory

> "has to do with the capacities human beings have to retain information, to recall it when needed, and to recognize its familiarity when they later see or hear it again" (Wingfield & Byrnes, 1981, p.4).

Memory has been traditionally divided into sensory memory, short-term memory (STM) and long-term memory (LTM). Sensory memory is an 'echo' of just received visual and auditory information. This information is held in sensory memory in an unanalysed form just long enough for its potential encoding.

The distinction between short-term memory and long-term memory is attributed to Atkinson & Shiffrin (1968). Short-term memory is capacity limited and can only hold information for a short duration. Many attempts have been made to determine this capacity, starting with Miller (1956) who introduced the term 'chunk' to describe the unit of short-term memory. He proposed that short-term memory could hold 7±2 chunks. Later attempts to measure the capacity of short-term memory have focussed on digit span and word span tests, which are now believed to be good measures of STM capacity. These are measures of the maximum length of sequence of digits or words that an individual can correctly recall. It is likely that span differs significantly among languages, but is generally between 5 and 10 items (Nell, 2000). According to Ardila (2003, 1994), the number of phonemes and digits that can be repeated after a single presentation depends upon the specific language, whereas the number of "semantic units" that can be processed is likely to be equivalent across languages. Baddeley, Thomson, & Buchanan (1975) estimate that

the temporal capacity of the phonological store[9] is about two seconds (see below for discussion of the phonological store). The two second duration is universal across languages. As length of digits differs across languages, the average number of digits which can be repeated during a two second interval will vary from one language to another.

Long-term memory on the other hand is not known to be subject to capacity constraints. It has seemingly unlimited capacity and duration.[10] Retrieval from long-term memory is constrained by practicalities; e.g. a speaker cannot take an unlimited time to retrieve the name of his interlocutor while he is speaking to him.

Baddeley & Hitch (1974) propose that it is necessary to introduce a further component of short-term memory, i.e. a type of memory which involves both storage and processing operations. They suggest that this is necessary as a result of their findings that many patients with short-term memory problems performed within the normal range on complex cognitive tasks which seemed to require short-term memory abilities. If their short-term memory was indeed impaired, how they did manage everyday tasks so successfully? This gave rise to the notion of Working Memory (WM).

According to the definition of Baddeley & Hitch (1974), Working Memory comprises a central executive and at least two other components, often called slave systems. These are the phonological loop and the visuospatial sketchpad. The central executive is limited in capacity and deals with focussing attention and controlling behaviour. The phonological loop is a temporary verbal-acoustic storage system which itself contains a storage component and a rehearsal component. The rehearsal component serves to refresh decaying phonological traces. The visuospatial sketchpad for its part is a parallel system for visual and spatial information. Thus, whereas the phonological loop deals with verbal and acoustic information, the visuospatial sketchpad is responsible for processing visual and spatial information. More recently, Baddeley (2000) has posited an additional

[9] The phonological store is the storage component of the phonological loop, one of the components of Working Memory. It stores verbal-acoustic information and is discussed in detail below.
[10] This is not to suggest that forgetting cannot take place and indeed be permanent.

component – the episodic buffer, which is capable of binding together information from a number of different sources into chunks or episodes. It is called a buffer as it provides a way of combining information from different modalities into a single multi-faceted code.

The different components of Working Memory are relatively independent, as seen from the fact that they can be independently impaired. Ardila (2003) notes that Working Memory storage processes show greater dependence on the posterior cortex, while rehearsal processes exhibit greater dependence on the prefrontal cortex.

Baddeley's theory highlights the separation of Working Memory from long-term memory and indeed Baddeley & Logie (1999, p.31) claim that viewing WM as an activated portion of LTM is probably an unhelpful oversimplification. There is, however, by no means unanimous acceptance of this in the literature. Many other authors; e.g. Cowan (1999) and Engle, Kane & Tuholski (1999a), propose that WM is a component of short-term memory, which is in turn a subset of long-term memory. According to this view, Working Memory is the subset of long-term memory information that is in the current focus of attention; e.g. Cowan (1999, p.62) notes that all three of the memory components (activation, focus of attention and awareness, and long-term memory) contribute to Working Memory. He suggests that WM refers to

> "cognitive processes that retain information in an unusually accessible state, suitable for carrying out any task with a mental component. The task may be language comprehension or production, problem solving, decision making, or other thought."

Similarly, Engle *et al.* (1999a, p.102) define Working Memory as

> "a system consisting of those long-term memory traces active above threshold, the procedures and skills necessary to achieve and maintain that activation and limited-capacity, controlled attention."

Evidence for this assertion that Working Memory is an activated part of long-term memory comes from experiments on priming which show that information held in long-term memory can influence subjects' decisions unconsciously. This is an

interesting debate but its outcome does not have an impact on the research pursued for this dissertation.

2.7.2 Phonological Loop and Language Acquisition

In any case, despite the controversy over the relationship between Working Memory and long-term memory, there is now little doubt but that the phonological loop plays an important role in spoken language acquisition (Baddeley *et al.*, 1998; Gathercole & Baddeley, 1989; Papagno & Vallar, 1995; Service, 1992), see *Section 2.7.4* below for discussion of WM and sign languages.

As mentioned above, the phonological loop is one of the slave systems in Working Memory. It deals with verbal-acoustic information and consists of a storage component and a rehearsal component. Phonological traces are stored temporarily in the storage component. The rehearsal mechanism acts to keep the traces 'alive' for a short period of time. Rehearsal allows the traces to be stored for longer than would be possible if no rehearsal took place; e.g. consider memorising a shopping list while doing shopping. Memorisation of the list is easier if the items on the list are rehearsed either subvocally or audibly.

Baddeley *et al.* (1998) propose that the primary purpose for which the phonological loop evolved is to store unfamiliar sound patterns while more permanent records are being constructed. This is an essential ability in language acquisition. Thus, the primary purpose of the phonological loop may be to facilitate language learning. Based on the results of an investigation into how native-language word-decoding ability predicted individual differences in native and second-language learning, Meschyan & Hernandez (2002) suggest that second-language learning is founded on native-language phonological-orthographic ability.

Gathercole & Baddeley (1989) attribute the poor performance of specific language impairment (SLI) children compared to normal children, to the storage component of the phonological loop. On the other hand, superior phonological loop capacity may be what distinguishes exceptional second language learners from poor learners. Papagno & Vallar (1995) compared polyglot and nonpolyglot university students and found that they performed indistinguishably on tests of nonverbal ability,

visuospatial short-term memory span and visuospatial learning. However, the polyglots performed significantly better on two phonological memory tests.

The phonological loop may also play a role in the acquisition of syntax. King & Just (1991) suggest that individual differences in syntactic processing are governed in part by the amount of Working Memory capacity available for language comprehension processes. Williams (1999) too, acknowledges the role of phonological Working Memory in the acquisition of syntax, arriving at the conclusion that

> "there is at least some sense in which knowledge of grammatical rules emerges out of memory for input and that individual differences in memory ability that are apparent even in the earliest stages of exposure have consequences for ultimate levels of learning" (*ibid.*, p.22).

Caplan & Waters (1999) however, suggest that the verbal Working Memory system specialised for assigning the syntactic structure of a sentence and for using that structure in determining sentence meaning is distinct from the Working Memory system that underlies the use of sentence meaning to accomplish further functions.

The following sections (*2.7.2.1, 2.7.2.2*) deal in greater depth with the role of the phonological loop in language learning, in particular how the capacity of the phonological loop may influence the learning process, and also how the two components of the phonological loop contribute to its overall importance in language learning.

2.7.2.1 Capacity of Phonological Loop

The ability to repeat nonwords has been shown to be an excellent predictor of later vocabulary knowledge in a first or second language (Gathercole *et al.*, 1991; Papagno & Vallar, 1995; Service, 1992). Such a test investigates subjects' ability to repeat nonsense words and is claimed to be a relatively pure measure of phonological loop capacity. Even with nonverbal ability partialed out, correlations between nonword repetition and vocabulary knowledge range from 0.31 to 0.50 ($p<0.05$) (Gathercole & Adams, 1993; Gathercole *et al.*, 1992). The task of repeating nonwords and learning L2 vocabulary are surprisingly similar, as both require a focus on new sounds.

Also, the ability to repeat digits, a test known as digit span, is positively correlated with vocabulary knowledge. Correlations between digit span and vocabulary measures, with nonverbal ability partialed out, range from 0.22 to 0.46 ($p<0.05$) (Gathercole & Adams, 1994; Gathercole *et al.*, 1997). This task also tests the capacity of the phonological loop. The advantage of the nonword task over the digit span task is, according to Baddeley (2003), that the former is closer to the situation facing language learners as it involves an unfamiliar sequence of phonemes.

These verbal short-term memory tasks are contrasted with reading span tasks which are designed to be a test of both storage and processing capacity (Daneman & Carpenter, 1980). Reading span tasks generally require subjects to read a number of sentences and remember the last word in each sentence. Remembering the final words is a storage task, while comprehension of the text is a processing task. Turner & Engle (1989) developed a similar test where the subject solved a string of arithmetic operations and then read aloud a word that followed the string. After a series of operations, the subject recalled the words. Thus, these tests examine overall Working Memory, i.e. storage and processing capacity, not merely storage capacity. Significant relationships between reading span (or other similar span tasks) and many cognitive tasks including reading comprehension (Daneman & Carpenter, 1980, 1983), language comprehension (King & Just, 1991) and complex learning (Kyllonen & Stephens, 1990) have been reported.

Many authors propose that span tasks are better predictors of other cognitive functions than the simple digit span test mentioned above, for example, Harrington & Sawyer (1992) found that reading span measures better predicted L2 reading skill than did passive, digit span measures. Similar findings were reported in Geva & Ryan (1993), who found a closer relationship between L2 proficiency and measures of L2 Working Memory than between L2 proficiency and passive measures of short-term storage (Robinson, 2003). However, it is worth emphasising that verbal STM tasks and WM capacity tasks are fundamentally different, in that the STM tasks measure storage only whereas the WM capacity tasks measure both storage and processing. As storage is particularly important for language learning, this may account for the positive correlations between storage tasks and later language

learning success. Clearly, language learning also requires general cognitive abilities such as the capacity to focus attention. Hence, positive correlations are also found between language learning success and the more general tests of WM capacity.

2.7.2.2 Phonological Store or Rehearsal System

As outlined above, the phonological loop comprises a passive temporary storage system and an active subvocal rehearsal system. Baddeley, Gathercole and Papagno (1998) have proposed that it is mainly the phonological store which is of primary importance in L1 acquisition, as they suggest that children do not employ rehearsal mechanisms until the age of 7, by which time L1 acquisition is well advanced (N. Cowan & Kail, 1996; Gathercole & Hitch, 1993). However, literature from other disciplines, such as developmental psychology, actually refers to the rehearsal mechanisms employed by children. This section aims to reconcile the assertion of Baddeley et al. with the findings which indicate the use of rehearsal mechanisms in children.

A different type of rehearsal mechanism to the one posited by Baddeley and colleagues is the Din in the Head phenomenon (see *Section 2.6.1* above), mentioned in the literature on SLA. It was noted above that the Din is a type of involuntary rehearsal which occurs when people spend time in contact with a foreign language. Research into the phenomenon, specifically the work of Jennifer Sevilla (1996), shows that the type of rehearsal taking place in the Din may occur in children as young as five. Krashen (1983, p.41) suggests that the Din occurs even in very young children and is manifested as "language in the crib". To the best of my knowledge, a comparison between the Din in the Head rehearsal and the rehearsal governed by the phonological loop, has not been made in the literature. The main difference between the two seems to be the time elapsed between presentation of the stimulus and rehearsal taking place. In phonological loop rehearsal, the rehearsal takes place immediately following presentation, in order to preserve the phonological traces in Working Memory for longer than would otherwise be possible. The material to be rehearsed comes from the phonological store, which is a type of short-term memory. The Din in the Head however, rehearses material from long term memory. Rehearsal takes place after a longer period of time has

elapsed. It was noted above (*Section 2.6.1*) that the Din took place after three days in contact with the foreign language.

Piaget refers to the rehearsal mechanisms of young children, a concept which he calls *langage égocentrique* (Piaget, 1923/1955, cited in Murphey, 1990, p.54). *Langage égocentrique* consists of three types of involuntary repetition employed by children. These are:

1. Echoic repetitions, in which the child seems to repeat for the pleasure of speaking without any concern for an addressee nor for words with meaning.
2. Monologues, which may accompany or replace action, often simply called "thinking out loud" among adults.
3. Collective monologues, in which children seem to be speaking together but in reality they are only under the illusion that they are being understood while paying little attention to others.

Vygotsky proposed that at around school age, egocentric language becomes inner speech, i.e. children realise that they can think words without saying them (Vygotsky, 1962). It is possible that this inner speech may be related to the Din in the Head phenomenon. According to the theories of Piaget and Vygotsky, the rehearsal mechanisms of young children may involve audible speech. As children get older, the audible speech rehearsal may turn into inner speech, i.e. it is possible that this inner speech could be related to the subvocal rehearsal referred to by Baddeley and his colleagues.

This could be one way to reconcile the fact that some researchers posit rehearsal mechanisms in young children, while Baddeley et al. suggest that rehearsal mechanisms are not used until the age of 7, i.e. it could be the case that audible rehearsal mechanisms are found in young children whereas subvocal rehearsal does not begin until later. While rehearsal may be important in L1 acquisition, it is possible that it is not the same subvocal rehearsal which is employed by adults in L2 acquisition and later acquisition of L1 lexis. The subvocal rehearsal employed by adults has been shown to be of significance in L2 learning (N. Ellis & Beaton, 1993; Papagno *et al.*, 1991).

Baddeley, Gathercole and Pagagno (1998, p.167) suggest that the phonological store is in fact the fundamental mechanism linking phonological memory and vocabulary acquisition. They suggest that the importance of the nonword repetition task in predicting later language success provides evidence supporting the hypothesis that phonological storage is of primary importance in L2 acquisition. Because the task involves repetition of a single item immediately after presentation, it is unlikely that rehearsal is involved to any great extent. Baddeley *et al.* (1998, p.168) suggest that

> "performance on the task is constrained by the quality of the phonological representation of the just-spoken unfamiliar item. In other words, nonword repetition provides a measure of the phonological store, not phonological rehearsal."

The phonological store is of particular importance in the early stages of language acquisition before there are representations of the language in long-term memory. This is because people try as far as possible to use their existing language knowledge when learning new vocabulary. Evidence in support of this comes from patients with brain injury who no longer have recourse to the phonological store, although their long-term memory is unimpaired. They have been shown to be able to learn pairs of words in their native language but not pairs of words where one of the pair is in an unknown foreign language (Baddeley *et al.*, 1988). A short-term memory patient, identified as P.V., is unable to learn associations between unfamiliar verbal material. However, she retains the ability to form associations between meaningful items that are already known in her native language. This is believed to be due to the fact that subjects form semantic associations between words that are familiar. P.V. can draw on her long-term memory and existing language knowledge. However, in the case of an unknown second language, subjects do not have long-term memory representations and cannot form semantic associations. Thus, they are dependent on the phonological loop (Baddeley, 2003; Dale & Baddeley, 1966). In cases where the phonological loop has been damaged as a result of brain injury, subjects no longer have recourse to it and are unable to learn associations between pairs of novel words.

To conclude, it can be said that researchers in the area of Working Memory attach great importance to the phonological store in vocabulary acquisition. However, short-term rehearsal mechanisms have also been shown to be important. Results from studies in SLA suggest that longer term rehearsal, i.e. the Din in the Head, is also an important and useful phenomenon for the second language learner. This phenomenon is also a rehearsal mechanism but seems to rehearse much larger chunks of language than those believed to be rehearsed in the phonological loop. In addition, the language rehearsed in the Din seems to come from long term memory as it usually occurs after a few hours of input in the language. The phonological loop rehearses short stretches of language immediately after it is input. Thus, perhaps efficient short term storage and long term rehearsal are among the primary mechanisms employed by the language learner.

It is likely to be the phonological store which is of primary importance to the subjects in the current study. In Phase II, subjects are asked to repeat short words and sentences in languages which are unknown to them (see *Section 4.3* below). As repetition immediately follows presentation, subjects do not have time to rehearse the material. Rather, they need to store the stimulus and repeat it immediately.

2.7.3 Working Memory & Intelligence

As discussed in *Section 2.6.2* above, the term 'g' was introduced by Spearman (1927) to describe general mental ability. It was the factor common to all the variables he measured such as Classics, French, English, Math, Pitch and Music. A student of Spearman's, Raymond Cattell then proposed the existence of two kinds of intelligence which are dominated by 'g', i.e. fluid intelligence (gF) and crystallised intelligence (gC) (Cattell, 1943). It was also noted in *Section 2.6.2* that there is now widespread support for the existence of 'g'; e.g. Carroll (1993, p.624) proposes that there is abundant evidence for a factor of general intelligence. Carroll also states that the 'g' factor is likely to be correlated with measures of speed of information processing and capacity of Working Memory.

The relationship between Working Memory and intelligence remains somewhat controversial but it now seems increasingly likely that individual differences in performance on tests of intelligence are associated with differences in the efficiency

or capacity of Working Memory (e.g. Mackintosh, 1998). Numerous researchers suggest that Working Memory is closely related to 'g' (e.g. Colom *et al.*, 2004; Jensen, 1998) and in particular, it appears to be closely linked to gF intelligence (R.W. Engle *et al.*, 1999b; Kyllonen & Christal, 1990).

The relationship of Working Memory to intelligence is important for this thesis. If it is simply the case that Working Memory is identical to intelligence and also that Working Memory is very similar/identical to language aptitude, the logical conclusion would be that language aptitude is very similar/identical to intelligence. However, the picture presented here seems more complex.

The following section examines to what extent Working Memory is domain-general or domain-specific. This is important when considering how Working Memory relates to language and musical aptitude. If WM is entirely domain-specific, it cannot be used to account for links between music and language aptitude. On the other hand, if WM is entirely general, it may be used to account for any link found between music and language aptitude.

Many studies highlight the domain-general aspects of Working Memory, focusing on the correlations between Working Memory and numerous other cognitive tasks. Kane *et al.* (2004) for example, emphasise the domain-general aspects of WM and they propose that WM capacity reflects primarily a domain-general, attentional construct. They present four categories of data that suggest the generality of WM capacity across verbal and visuospatial domains (*ibid.*, p.192):

- Verbal Working Memory span can sometimes predict spatial ability, and spatial WM span can sometimes predict verbal ability, with cross-domain correlations as high as those within domains.
- Cross-domain correlations among WM span tasks are higher than those among STM span tasks, suggesting that STM reflects more domain-specific skills, strategies, and storage abilities than does Working Memory capacity.
- Individual differences in domain-specific ability can be substantially reduced by accounting for WM span measured in a different domain.

This point may require further clarification. Now, if WM were domain-specific we should expect that an individual's WM span might vary from one domain to another; e.g. a person could have high verbal WM span but low visuospatial WM span. If this were to be the case, it is likely that only WM span in the particular domain under investigation would account for individual differences in ability; e.g. in accounting for differences in verbal ability between individuals it would only be useful to look at verbal WM span and not visuospatial WM span. However, according to the proposal by Kane *et al.* this is not the case, and in fact WM span measured in any other domain (e.g. visuospatial) may actually also contribute to individual differences in a particular domain (say verbal abilities); this they use as evidence for the generality of WM.

- Studies using latent-variable procedures find that constructs composed of multiple verbal and spatial WM span tasks are identical or share 65% or more of their variance.

Other authors, however, point to the domain-specific aspects of WM, highlighting the fact that correlations between the different domains, e.g. verbal and spatial, are often quite low compared to associations within the same domain (e.g. Jurden, 1995; Shah & Miyake, 1996). A study by Mackintosh & Bennett (2003) investigated the relationship between gC (crystallised intelligence), Gv (spatial intelligence) and gF (general fluid intelligence) and tests of WM components (verbal, spatial and numerical). They found a correlation between verbal WM and the vocabulary test (a measure of gC), and a correlation between spatial WM and the mental rotation test (a measure of Gv), but only a weak relationship between these two domains. Thus, they found that the correlations between the tests reflected the content of the tests (reasoning, verbal and spatial), rather than the structure of the tests (ability vs. WM). Hence, they conclude that WM may be partly general but that it is at least in part domain-specific, with three of these domains corresponding to gC, Gv and gF.

Despite these apparent contradictions, it is likely that the central executive is domain-general and quite closely related to gF intelligence. The slave systems such as the phonological loop and visuospatial sketchpad appear to be the more domain-specific components of WM. In some cases it may be a problem of terminology, as

Kane *et al.* (2004) say that their use of the term "WM capacity" refers to the domain-general executive component of the Working Memory system, whereas Baddeley's (2003, 1974) use of the term Working Memory refers to the central executive and the slave systems. Engle *et al.* (1999a), like Kane *et al.*, also consider WM capacity to be the capabilities of the limited capacity attention mechanism, which is the central executive in the model of Baddeley and Hitch (1974).

Kane *et al.* (2004) indicate that Working Memory tasks largely reflect a domain-general factor, whereas short-term memory tasks, based on the same stimuli as the WM tasks, are much more domain-specific. Their study therefore, actually supports the idea of a domain-general central executive and more domain-specific slave systems and they conclude that

> "the findings support a domain-general view of WM capacity, in which executive-attention processes drive the broad predictive utility of WM span measures, and domain-specific storage and rehearsal processes relate more strongly to domain-specific aspects of complex cognition"
> (Kane *et al.*, 2004, p.189).

A recent meta-analysis by Grubb & McDaniel (2000) shows that a closer relationship exists between Working Memory and fluid intelligence than between short-term memory and fluid intelligence. They consider word and digit span tasks, letter recall tasks, and visual and auditory memory tests, to be short-term memory tests. Working Memory tests include operation-word or digit recall tasks, sentence-word or digit recall tasks, ordered letter tasks and iconic memory tasks. Their results show a population correlation of 0.38 between Working Memory and gF intelligence and 0.19 between short-term memory and gF intelligence. They also examine crystallised intelligence (gC) but find a smaller correlation between WM and gC ($r=0.22$) than that between WM and gF ($r=0.38$). There is no strong evidence of short-term memory being more highly correlated with either gF or gC, which contrasts with Working Memory which is more closely related to gF.

Given the only moderate correlation (0.38) between WM and gF, Grubb & McDaniel (2000) deny the suggestion that WM is the same as gF intelligence. While the two are certainly related, it is far from a perfect correlation. Even given

the fact that this may be an underestimation of the population correlation, the authors conclude that the best that can be said is that

> "we have provided reasonable evidence that WM is moderately related to fluid intelligence and that STM has much lower relationships with fluid intelligence."
> (*ibid.*, p.15)

It makes intuitive sense that a closer relationship should exist between Working Memory and fluid intelligence than between short-term memory and fluid intelligence, given the Cowan (1999) definition of WM as:
WM = STM + controlled attention
It has been established that short-term memory is relatively domain-specific, while Working Memory, or at least the central executive component, is relatively domain-general. General fluid intelligence is general and should therefore be more closely related to the more domain-general component of Working Memory, i.e. controlled attention or the central executive. Given the confusion over the WM definition, which in some instances actually refers only to the controlled attention component, clearly gF relates more closely to WM than to STM. Direct support for the idea that the capability for controlled processing (i.e. a central executive functioning) is the critical element common to the Working Memory tasks, tests of gF intelligence and tests of higher-order functions such as reading comprehension, is presented in Tuholski (1997).

2.7.4 Other Working Memory Components

One problem with Baddeley's model of Working Memory is that it focuses particularly on two storage types but neglects others such as the storage of nonverbal sounds and tactile sensations. Clearly not all memories are either phonological or visuospatial; humans can have memories of smell, taste and feelings. This leads to the suggestion that other slave systems may exist in addition to the phonological loop and visuospatial sketchpad. Cowan (1988) aims to address this issue. He points out that at least some aspects of representations are similar across modalities; e.g. in each modality (including at least vision, audition and tactile senses), the ability to compare two slightly different stimuli declines rapidly as a function of the time between the two stimuli across about 10 to 20 seconds.

Also, in each modality, the greatest interference comes from additional similar stimuli in that modality.

One possible additional slave system would control the processing of musical stimuli. Such a system has received relatively little attention but has been considered by Logie & Edworthy (1986) and Berz (1995). It is, as yet, uncertain how closely such a system might be linked to the phonological loop. Evidence suggesting that they would not be completely separate comes from the study by Logie & Edworthy (1986) which shows overlap between a homophonic judgement task and two types of musical task. This implies that the verbal task involves some of the same mechanisms as the non-verbal auditory task. A spatial task did not interfere with the musical task, and so the verbal and musical tasks appear to be more closely related than the spatial and musical tasks.

Using a performance decrement on a primary task when combined with a specific secondary task to indicate some functional overlap between the cognitive mechanisms involved in the two tasks is a widely used technique. Interference has often been used as evidence of association; e.g. suppression of subvocal rehearsal has been shown to interfere with learning vocabulary (Papagno *et al.*, 1991). Thus, subvocal rehearsal has been positively implicated in vocabulary learning. Similarly, if a language task interferes with a music task, this suggests that they must draw on at least some of the same resources.

It would therefore seem premature to propose absolute separation of a musical loop and the phonological loop, although it is equally clear that they are not identical. Evidence against the identity of the two loops comes from the unattended music effect, i.e. it has been shown that unattended vocal music interferes with language to a greater extent than unattended instrumental music. If there was a singular acoustic store, unattended instrumental music would cause the same disruptions on verbal performance as would unattended speech or unattended vocal music but this is not the case (Berz, 1995). Thus, Berz (*ibid.*) concludes that the actual attachment of a music loop is open to some question even though the existence of the loop does seem fairly clear. While he suggests that the existence of a separate loop seems

more justified, given the results of Logie & Edworthy (1986), it seems preferable to consider some attachment between the two.

It is interesting to consider how sign languages employ Working Memory. Sign languages, like spoken languages, exhibit phonological and syntactical organisation, given the definition of phonology as the "patterning of the formational units" of a natural language (Coulter & Anderson, 1993, p.5). As a result, the question arises as to whether sign languages exploit the phonological loop and rehearsal mechanisms used by spoken languages or whether, given their reliance on visual signs, they use the visuospatial sketchpad. It now seems likely that

> "the architecture of Working Memory is shaped both by the properties of language structure and by the constraints imposed by sensorimotor modality."
>
> (M. Wilson & Emmorey, 1997b, p.121)

Thus, the fact that sign languages exhibit some of the properties of spoken languages means that they may in fact use some of the same Working Memory principles. However, there are also certain differences in the WM system used by sign and spoken languages, arising out of the differences in modalities. Both empirical research and neural imagery investigations contribute to the understanding of the overlap between signed and spoken languages.

Regarding the neural architecture involved in sign language, it has been noted that the activation pattern for sign language bears some similarity to the neural activity during nonverbal visuospatial tasks (Rönnberg *et al.*, 2004). Overall results show specific activation patterns for Swedish sign language which are different from audiovisual Swedish speech, except in Broca's area where there are similarities across modalities for all tasks and in one other area (anterior left inferior frontal lobe) where similarities occur for semantic retrieval. Therefore, from the point of view of brain regions, there are both similarities and differences between sign and spoken languages.

One of the differences between spoken and sign languages is in the way in which serial order information is expressed, with the auditory modality being heavily dependant on linearity. Sounds are produced consecutively whereas signs can be

produced concurrently. The visual modality employs spatial representations which do not depend to the same extent on linearity. It seems that serial order information in American Sign Language (ASL) is stored in a format that does not have a preferred directionality, with the result that native users of ASL are found to be as good at backward recall of digits as forward recall in a digit span test (M. Wilson *et al.*, 1997). Spoken language users find backward recall much more difficult that forward recall.

Experience with American Sign Language as a native language appears to positively affect visuospatial WM, with native users of ASL outperforming native speakers of English and deaf non-native speakers of ASL on a test requiring memorisation of a sequence of spatial locations (*ibid.*). In addition, both deaf and hearing signers have an enhanced ability to generate relatively complex images and to detect mirror image reversals, as compared to hearing speakers of English (Emmorey *et al.*, 1993). Deaf signers are disrupted on serial recall of lists of ASL signs, by pseudosigns or moving shapes, whereas hearing subjects recalling lists of printed English words are not (M. Wilson & Emmorey, 2003). Thus, clearly ASL has a much closer relationship with non-linguistic visuospatial memory than spoken language has.

While it may be true that experience with spoken language plays a role in developing non-linguistic auditory WM, it is almost impossible to verify this empirically in the way that sign language and visuospatial memory have been investigated. This is because it is virtually impossible to find hearing subjects without spoken language. It is more likely to be possible to investigate empirically the proposal that experience with non-linguistic sounds helps develop linguistic auditory memory. Indeed this has been investigated and is discussed in Chapter 3. There it is shown that adults who received musical training in childhood have better verbal memories than adults without musical training in childhood (Chan *et al.*, 1998).

Despite the differences between auditory and sign languages, there are also certain similarities between the two modalities. A significant phonological similarity effect (worse recall of similar signs), a manual articulatory suppression effect and a sign-

length effect have been documented among deaf users of ASL (M. Wilson & Emmorey, 1997a, 1997b, 1998). These are in line with the phonological similarity effect, articulatory suppression and word-length effect in hearing subjects. Thus, it can be concluded that "the existence of a phonological loop for sign is highly similar in structure to the phonological loop for speech" (M. Wilson & Emmorey, 1997b, p.127). They propose also that the "phonological loop is a mechanism that develops in response to appropriate linguistic input, regardless of the modality of that input" (*ibid.*).

2.7.5 Summary of Discussion of Working Memory

Working Memory is generally viewed as a system comprising a central executive which controls attention and processing, and two slave systems; a phonological loop and a visuospatial sketchpad. The majority of research has focussed on these components. However, the evidence presented here from preliminary research into possible additional components suggests that extra slave systems may be worthy of further consideration. To date little is known about memory for music or auditory stimuli other than language but it is possible that a slave system for musical processing may exist. Such a loop may use some of the same mechanisms as the phonological loop for language. Evidence from research into sign language suggests that the phonological loop and rehearsal mechanisms are important in language processing even across different modalities. If aspects of the phonological loop are used in processing sign language, which is certainly at least somewhat different from spoken language, it seems plausible that the phonological loop could also be involved in processing musical sounds, on the basis that music is another rule-based auditory system, only somewhat further removed from spoken language than sign language.

Another area of memory research which is receiving increasing attention relates to the link between WM and intelligence. It now seems that while WM capacity and gF intelligence have much in common, they are not identical.

Working Memory is important for this thesis, as the productive language aptitude tests used in Phase II of the current study are based on the premise that Working

Memory ability is a good indicator of foreign language aptitude. These tests are introduced and discussed in *Section 4.3.2.2.*

2.8 Conclusion

The aim of this chapter was to present a detailed account of some of the ways in which learners differ. Emphasis was placed on language aptitude, intelligence and Working Memory as these factors prove significant to the remainder of this thesis. Of course numerous additional differences arise between learners and have been extensively studied. These include age (e.g. Bongaerts et al., 1995; Marthohardjono & Flynn, 1995; Muñoz, 2003; Singleton, 2003, 2004), anxiety levels (e.g. Kitano, 2001), motivation (e.g. Dörnyei, 1994, 2002; Garcia & Pintrich, 1995; R. C. Gardner, 2001; R. C. Gardner & Lambert, 1965; MacIntyre, 2002), learning styles and strategies employed by learners (Chapelle, 1995; Ehrman & Leaver, 2003; Oxford, 1990; Şakar, 2003). Other individual differences have been posited to affect the language learning process but have not been so extensively studied. These include language disability (e.g. Dinklage, 1971; Ganschow & Sparks, 1986; Grigorenko, 2002; Sparks *et al.*, 1989), sex (Eisenstein, 1982; Farhady, 1982) and birth order (Pine, 1995).

Typical classroom situations group learners according to age or ability and then assume that all students can learn in the same way, although some may make faster or slower progress. Heterogeneity of learners is all but ignored. However, individual difference research shows that learning styles and preferred strategies vary across individuals. In addition, personality traits, aptitude, age, intelligence, Working Memory capacity and motivation vary from one learner to another. To complicate matters even further, these variables do not act in isolation. A change in the strategies employed by a learner may result in a change in levels of motivation. As a learner grows older, the importance of aptitude may change. When teaching methods take account of these differences between individuals, learners make faster progress and experience greater success.

It often appears to be the case that the more that is learned about individual differences, the more complex the situation becomes. Each of the areas examined

in this chapter is worthy of much greater attention than it can receive here and already forms the subject of innumerable research papers and books. Every day, as new results become available, new light is shed on the questions examined here. This is particularly true in the cases of intelligence and Working Memory research, as new developments in technology make possible brain imaging, which was heretofore unimaginable. While mapping functions to parts of the brain does not in itself improve our understanding of the functions, it can show us which processes take place in similar brain regions, suggesting that these processes may have something in common. Also, an understanding of how processes break down in cases of brain injury can provide useful information even about normal functioning.

While this chapter could not discuss in great detail the IDs considered, it is hoped that adequate background is given in order to demonstrate the complexities of the issues involved. It is hoped that the reader can now appreciate that it is very difficult to give one definition of language aptitude or to define precisely what is meant by intelligence or even Working Memory. The following chapter investigates musical ability and training as one specific individual learner difference.

Chapter 3 Music, Language and the Brain

3.1 Introduction

There is now little doubt but that the cognitive processes involved in the perception and performance of music interact with other cognitive processes in ways which are not yet fully understood. The view that music may have some effect on other mental processes is becomingly increasingly widespread but the ways in which it may have its effects are still unclear.

In 1983, Gardner (1983) first proposed his theory of multiple intelligences where he named seven distinct areas which he believes constitute intelligences. As noted in Chapter 2, these are musical intelligence, bodily-kinaesthetic, logical-mathematical, linguistic, spatial, interpersonal and intrapersonal. An intelligence entails the ability to solve problems or fashion products that are of consequence in a particular cultural setting or community. The various intelligences are for a large part independent of one another.

Varying degrees of different intelligences are clearly required for different tasks; e.g. dancing requires skills in bodily-kinaesthetic, musical, interpersonal and spatial intelligences. Thus, an excellent dancer is likely to show high levels of each of these intelligences. However, it is not the case that these particular intelligences always co-vary with each other. Hence, the claim can be made that the intelligences are relatively independent. Rauscher and Zupan (2000) cite results of a study which provides evidence in favour of the independence of intelligences (Rauscher, 1999). They note that musical aptitude and spatial reasoning scores of 3 and 4 year olds were found not to be correlated in pretests.

However, it is worth noting that in 1993, Gardner stated that there is no theoretical reason why two or more intelligences could not overlap or correlate with one another more highly than with others (H. Gardner, 1993, p.41) and later he asserts that "music may be a privileged organizer of cognitive processes, especially among young people" (H. Gardner, 1997, p.9).

As yet, it remains unclear to what extent training in one area may affect another (see *Sections 3.3* and *3.4* below for consideration of the effects of musical training on other cognitive abilities). It seems likely nonetheless, that the various intelligences are correlated in different ways and to varying degrees. Thus, it appears that while the different intelligences may be mainly independent, experience in one domain may affect another.

Recent research points towards the likelihood of a relationship between musical activity and motivational and emotional states. Kallinen (2002), for example, shows that background music can influence how people evaluate news content, with men evaluating news most positively in the presence of slow music. Researchers in the domain of advertising have also long been interested in how music influences message recall and emotions (see e.g. Hahn & Hwang, 1999).

Musical activity has also been related to spatial reasoning (e.g. Rauscher *et al.*, 1993; Rauscher *et al.*, 1995), to increased creativity (Weinberger, 1998b) and to linguistic intelligence (e.g. Douglas & Willatts, 1994). When considering the possibility of a relationship between music and other mental tasks, there are two general issues which must be addressed. These are, firstly, the conditions under which music gives rise to enhanced mental performance and secondly, the mechanisms by which music can improve mental performance, i.e. can/does music play a role in improving mental functioning and if so, how does it do it? Research in neurology attempts to shed some light on the second of these questions while disciplines such as education, linguistics and music-therapy investigate the first. The relationships between various intelligences, focusing primarily on the correlation between music and language, will be investigated in this chapter.

3.2 Music and Spatial Reasoning

The relationship between music and spatial-reasoning hit the headlines in the popular press with the discovery of the so-called "Mozart-effect". Suddenly people believed that listening to music would make them more intelligent. However, the effects in question are much more complex and the term has now been generalised to refer to both the short term effects of the music of Mozart on spatial reasoning

and to the long term benefits of music instruction on other areas. The Mozart effect seems to relate specifically to spatial-temporal reasoning and less so to spatial recognition. Spatial-temporal tasks require spatial imagery but also some kind of temporal ordering of objects whereas spatial recognition simply requires that an individual recognise similarities between objects in order to classify them.

Rauscher *et al.* (1993) originally showed that listening to a Mozart piano sonata produced significant short-term enhancement of spatial-temporal reasoning in college students. A relaxation tape or silence did not cause the effect. Rauscher *et al.* (1995) attempted to replicate these findings and showed that listening to ten minutes of Mozart's Sonata for 2 pianos, K.448, resulted in an increase of 8 to 9 points on the spatial IQ subtest of the Stanford-Binet Intelligence Scale (Thorndike *et al.*, 1986). This effect lasted only 10 – 15 minutes and was not caused by listening to repetitive music such as dance, or speech (taped story). Short-term memory was not enhanced, in so far as there was no difference between the experimental group who had heard the music and the control group who had had ten minutes of silence prior to the test, when asked to remember a list of sixteen items.

3.2.1 Explaining the "Mozart Effect"

One possible explanation for the "Mozart effect", first proposed by Leng and Shaw (1991), relates to the spatial-temporal firing patterns associated with the networks of trions in the cortical column. They point to the special role of music among higher brain functions and suggest that music may play a role in enhancing the cortex's ability to develop these inherent firing patterns, thus improving other higher brain functions. While this is undoubtedly the most commonly cited explanation in the literature, a possible alternative is suggested by Parsons, Hodges and Fox (1998).

According to this theory, the rhythmic elements of music which are processed in the cerebellum are responsible for the enhancement of spatial tasks that also require cerebellar function. Gruhn and Rauscher (2002, p.453*f.*) describe an experiment carried out by Parsons and colleagues in which subjects performed two spatial-temporal tasks following one of five conditions: auditory exposure to rhythm without melody (a popular song base line), auditory exposure to melody without rhythm (a melody presented in a steady beat), visual exposure to rhythm (a

pulsating square on a computer screen), auditory exposure to a continuous tone, or silence. Enhanced performance on the spatial-temporal tasks was found following only the auditory and visual rhythmic conditions, leading to the conclusion that "the enhancement of spatial temporal tasks is due to rhythm, regardless of the modality of presentation." Interestingly, the particular contribution of rhythmic skills to other abilities resurfaces in *Section 3.3*, in relation to children with Specific Language Impairment.

The reasons proposed by Rauscher *et al.* (1995, p.47) to account for the enhancement of spatial reasoning by music are as follows:

- Listening to music helps 'organise' the cortical firing patterns so that they do not wash out for other pattern development functions, in particular, the right hemisphere processes of spatial-temporal task performance.
- Music acts as an 'exercise' for exciting and priming the common repertoire and sequential flow of the cortical firing patterns responsible for higher brain functions.
- The cortical symmetry operations among the inherent patterns are enhanced and facilitated by music.

One very interesting study attempting to account for the "Mozart effect" used electroencephalogram (EEG) recordings to investigate the effects of listening to speech on a paper folding and cutting (PF&C) task compared to the effects of listening to Mozart on the same task (Sarnthein *et al.*, 1997). It was found that the patterns of cortical activity induced by music seemed to carry over to the subsequent tasks whereas this did not happen for the speech condition. The PF&C task alone showed activity patterns in the two brain regions engaged in processing spatial temporal sequences (parietal cortex and prefrontal cortex). When doing the PF&C after listening to music, the left temporo-parietal cortex was additionally activated; thus music induced involvement of left and right temporal lobes, as opposed to only right involvement in PFC without music (*ibid.*, p.113). In addition, the music seems to have caused increased co-operativity between different prefrontal sites which provides evidence as to the mechanism underlying increased task performance: instead of activating several sites independently, synchronization

between different cortical sites was increased. Listening to text did not have the same carry-over effects.

3.2.2 Replication of the "Mozart Effect" Experiments

Numerous investigators, with varying degrees of success, have attempted to replicate the Mozart effect (e.g. Carstens *et al.*, 1995; Hallam, 2000; Steele *et al.*, 1999). One of these, Carstens, Huskins and Hounshell (1995), used the Revised Minnesota Paper Form Board Test (Likert & Quasha, 1948) rather than the Stanford-Binet scale used in the first experiment. They were unsuccessful in their attempt to replicate the earlier results. They put this down to the type of visuo-spatial factor under investigation in the two tests. The Stanford-Binet tasks consist of a pattern-analysis test, a multiple choice matrices test and a multiple choice paper-folding and cutting test. The Minnesota Paper Form Board Test presents the subject with a picture of a 2-dimensional figure cut into two or more parts. The subject's task is to select the display which shows how the figure would look were the pieces put together. These problems require mental rotations of the 2-dimensional figures. Using the terminology of Gustafsson (1984), Carstens *et al.* propose that the Minnesota test corresponds to the Spatial Orientation component of visualization, whereas the Stanford-Binet test corresponds to Gustaffson's Visualization factor. According to Gustaffson there are at least three visuospatial first-order factors (Visualization, Spatial Orientation and Flexibility of Closure) subsumed by a more general visualization factor.

It is interesting to see from the experiment of Carstens *et al.* that the relationship between music and reasoning ability may be even more specific than that originally envisaged, i.e. music may be related to Spatial Orientation but not at all, or to a lesser extent, to Visualization. Indeed Rauscher herself later notes that the "Mozart effect" seems only to relate to spatial-temporal tasks and not to other types of spatial tasks (Rauscher & Shaw, 1998). Other reasons put forward for failure to replicate the "Mozart effect" in certain experiments are as follows (Rauscher, 2000):

- Task validity: The effect has been shown by researchers using spatial-temporal tasks but researchers using other types of visuospatial tasks have

failed to replicate. This is the reason given above to account for the results of Carstens *et al.* (1995).

- Expectancy Effects: Experimenters' beliefs about the outcome of a study can affect its actual outcome. In cases where experimenters have positive beliefs, it is much more likely that the study will produce positive results. This is borne out in an experiment by Rauscher and Sparr (2000) which shows that subjects in the high expectancy group scored significantly higher than those in the low expectancy group, even though experimenters had identical instructions and scripts.

This issue is very often addressed by those working in the area of medical research, where the outcomes of clinical trials can result in major changes in hospital policies or the treatment of patients. Trials are very often 'double-blind', which means that neither participants nor experimenters know whether individual subjects belong to the control group or to the experimental group. One example of this is a study which examined the effects of multivitamin supplements on quality of life in people aged 65 or over (Avenell *et al.*, 2005). Participants were randomly allocated to either the multivitamin group or a placebo[11] group. Participants did not know which group they belonged to, nor did the administrator of the tablets know to which group participants belonged. In this way, the outcome of the study cannot be affected by experimenter beliefs. In the case of the Mozart effect experiments, it is much more difficult to blind participants and experimenters, as it is quite clear whether participants belong to the music group or a different group. It is possible however, to blind participants and experimenters to the hypothesis predicting greater success for the music group.

- Instructions to Participants: Greater effects are found in cases where participants are explicitly told to listen carefully to the music compared to situations where the music is played without subjects being told to focus on it. Rauscher (2000, p.3) suggests that

> "attention may either permit a more complete neural activation of the networks involved in processing music and spatial information, or it may increase arousal. In either case, it seems unlikely that a Mozart

[11] A placebo is an inactive substance used as a control in an experiment to determine the effectiveness of a drug. The use of a placebo allows medical researchers to see if it is simply the act of taking medication which causes a patient to feel better, or if it is actually the active ingredient which causes the improvement.

effect would be found for participants who did not actively process the music."

- Item Difficulty: In cases where the spatial-temporal task is too easy, subjects may use automatic processes which are not affected by the music. According to Block and Grosfield (2000), there is a positive relationship between effect size and item difficulty.
- Practice Effects: Practice effects may occur in those studies using a pre-test post-test design. Rauscher (2000) points out that if subjects are pre-tested and post-tested, a ceiling effect may arise which obscures any post-test differences.

A recent meta-analysis concludes that it now seems safe to claim that the "Mozart effect is a moderate effect and it is robust" (Hetland, 2000b, p.33), given the fact that it has been replicated 29 times in 13 independent laboratories (Block & Grosfield, 2000; Hetland, 2000b). It seems that it is not simply a case of the music changing arousal states of subjects, as Rauscher *et al.* (1993) examined pulse rates and found no difference in pulse between those listening to Mozart and those listening to a relaxation tape or silence.

On the other hand, Schellenberg (2003) cites findings on the association between positive mood states and cognition. Positive moods increase levels of dopamine which is believed to lead to an improvement on a variety of cognitive tasks. He suggests that listening to music is just one way to induce positive affect, thereby increasing levels of dopamine which in turn enhances performance on cognitive tasks.

3.2.3 Extension of Findings

If the existence of the "Mozart effect" is indeed accepted, it may also be worth considering if only the music of Mozart can cause this effect. Ivanov and Geake (2003) present results of a "Mozart effect" in a natural setting among school children. A similar effect was also found in one of their groups who listened to Bach during testing. These results, along with those of other studies (e.g. Nantais & Schellenberg, 1999; Rideout *et al.*, 1998), show that the "Mozart effect" may not be specific to the music of Mozart, as similar effects have been found while listening to

Schubert or a contemporary Yanni composition (proposed to be similar to the Mozart sonata in terms of tempo, structure, melodic and harmonic consonance and predictability). It may be the case that regularly repeating sequences in the music cause the brain to react positively. Hence, other compositions featuring this property would also generate the effect. Rauscher and Shaw (1998, p.839) suggest that

> "complexly structured music, regardless of style or period, may enhance spatial-temporal task performance more readily than repetitious music."

Given that the scope of the term "Mozart effect" has been widened to include the long term benefits of music training on spatial reasoning, it is also necessary to consider this aspect of the phenomenon. Following the suggestion made by Leng and Shaw (1991) that music training of young children should produce long-term enhancement of spatial-temporal reasoning, given that their cortices are highly plastic, an experiment was undertaken to investigate this possibility (Rauscher *et al.*, 1997). It was shown that preschool children who had received keyboard lessons over a two year period made significant improvements on spatial-temporal tests. Children who had received singing lessons or computer classes did not make similar improvements. The effects are considered to be long term as they last for at least a day. This contrasts with the short-term benefits of listening to Mozart mentioned above which lasted for only ten minutes. Rauscher *et al.* (*ibid.*) suggest that the results of this study provide evidence that music training produces long-term modifications in underlying neural circuitry in regions not primarily concerned with music. This issue is addressed again in *Section 3.4.2* below.

A study, similar to that of Rauscher *et al.* (1997), conducted in 2000, produces converging results (Rauscher, 2002; Rauscher & Zupan, 2000). In this study, sixty-two children attending kindergarten were assigned to one of two conditions, group keyboard lessons or no music. Members of the keyboard group were found to score significantly higher than those in the control group on spatial-temporal tasks after 4 months of lessons. When tested again 8 months after lessons began, the keyboard group had made additional improvement over the non-music group. In both studies, the improvements were found only in spatial-temporal tasks and not in spatial recognition or pictorial memory. Costa-Giomi (1997, 1999) also shows that 9 year

olds who were provided with two years of keyboard lessons significantly outperform children without such lessons on spatial tasks.

In addition to considering the short term "Mozart effect" in the meta-analysis mentioned above, Hetland (2000a, p.41-42) also looks at the longer-term effects of music instruction, concluding that the effect is "remarkably consistent" and could be

> "generalized to similar populations of preschool- and elementary-school-aged children, while they are engaged in similar kinds of active music programs, with or without keyboard instruments, taught in groups or individual lessons. ... It is a solid finding."

3.2.4 Conclusion

While the 'Mozart-effect' discussed here applies specifically to enhancement of spatial-temporal skills, it seems that the long term benefits of music instruction may be more far reaching, even developing short term memory (see *Section 3.3.3*). In acknowledging that simply listening to music may only affect short-term spatial-temporal skills, it is useful to look at evidence of long-term music training on other abilities, such as language skills. Schellenberg (2003, p.430) proposes that the short-term 'Mozart-effect' is

> "small and unreliable. ... By contrast, the effect of music lessons on non-musical aspects of cognitive development is still an open question."

He sees the long-term effects as being more significant than the short-term effects. The longer term benefits of musical training on other cognitive abilities are examined in the following sections.

3.3 Music and Linguistic Intelligence

Music is not only related to spatial intelligence. It has also been shown to relate to another of Gardner's intelligences – linguistic intelligence. Many studies have investigated this relationship, looking at both first and second language ability as well as at breakdown in the normal course of events.

Music therapy may be employed to help children with speech and language impairment (SLI). SLI children may have sufficient speech sounds and vocabulary

but may stop expressing themselves fully through speech. Sutton (1995) found parallels between SLI children's progress in music and their progress in language. She found that as children began to build their music into phrases and structures, so they also began to express themselves with their voices and to construct simple sentences. This is interesting in light of the discussion in Chapter 1, where it was noted that researchers have suggested that the same parsing model may be used to parse both music and language structures. It was also pointed out that music and language are rule-based systems and that grammar rules can predict the most psychologically plausible segmentation of each. The research discussed in Chapter 1 is of a more theoretical nature so it is noteworthy that further support for overlap of structure in music and language should emerge from empirical work with SLI children. The question of structure in music and language is addressed again in *Section 3.3.3*, where the 'Shared Syntactic Integration Resource Hypothesis' is considered.

Sutton's proposal that function and structure in music and language may be related led her to carry out a study of how SLI children perceive and pattern sound. She hypothesised that SLI children may have difficulty hearing, processing and assimilating groups of sounds that children with mainstream speech did not. She tested her hypothesis using rhythm patterns which the children had to imitate. Early indications suggest that her hypotheses are indeed borne out.

Therapy using music (Melodic Intonation Therapy) has also been used to good effect with certain aphasic patients (Benson *et al.*, 1994). It has been beneficial for patients with poor production skills but relatively intact comprehension abilities following left-hemisphere brain injury. It is based on the premise that intonation and singing abilities may be preserved in the right hemisphere despite the damage to the left-hemisphere. Word sequences are incorporated into a song. After some time the melody is phased out until the patient can speak without singing. Therefore, the intact right hemisphere learns the phrases with the music and develops more language production skills to compensate for the left-hemisphere deficit. It has been shown that patients having undergone this treatment may participate successfully in short conversations.

Of course it is not only in the case of injury or developmental delay that associations have been found between music and language. The following sections consider findings related to normal language acquisition.

3.3.1 Music and First Language Acquisition

Interest in the idea that music and language abilities may be related is not new. The old adage of a "musical ear" being helpful in language learning inspired research as far back as the 1930s (e.g. Dexter & Omwake, 1934). Though it was considered to be an old wives' tale by some, others set about examining the issue in a scientific fashion.

Blickenstaff (1963) carried out a review of studies which examine how musical talents may be related to first language acquisition. Firstly, he discusses factor analytic studies which examine L1 competence and how it relates to musical ability as measured by the Seashore Measures of Musical Talents. Factor analysis was described in *Section 2.6.2* above. It is a statistical technique which groups tests into factors which share some common underlying ability.

The Seashore Measures consist of six subtests: Pitch, Loudness, Rhythm, Time, Timbre and Tonal Memory. The test of pitch requires the subject to determine in a pair of tones whether the second tone is higher or lower in pitch than the first. The loudness test requires the subject to determine whether the second tone is stronger or weaker than the first. For rhythm, the subject indicates whether the two patterns in each pair are the same or different. The time test requires the subject to determine whether the second tone is longer or shorter than the first. The subject is required to judge whether the tones are the same or different in tone quality, for the test of timbre. For tonal memory, there are 30 pairs of tonal sequences consisting of four-, five-, and six-tone groups. Within each pair, one tone in the second sequence differs in pitch from its corresponding tone in the first sequence. The subject must identify which note it is by the number of its position in the sequence. The Seashore test battery is not unlike the Bentley Measures of Musical Aptitude used in the current study. A more detailed analysis of the Bentley measures is presented later in *Section 4.2.2*.

One of the studies cited by Blickenstaff is that of Holmes (1954) who examined the relationship between subtests of the Kwalwasser-Dykema Music test (Kwalwasser & Dykema, 1930) and L1 spelling ability. The Kwalwasser-Dykema Music test consists of eight tests, i.e. tonal memory, quality discrimination, intensity discrimination, tonal movement, time discrimination, rhythm discrimination, pitch discrimination and melodic test. Six of these tests were significant in accounting for variance in spelling ability. Holmes found that the subtests from the Kwalwasser-Dykema Music test accounted for the following percentages of variance in spelling ability on a sample of 227 high school students: tonal movement 7.3%; pitch 6.5%; tonal memory 1.7%; intensity 1.4%; rhythm 1.3% and melodic taste 0.4%. At college level (N=193), the Seashore subtests were also found to account for variance in spelling ability even with the effects of intelligence partialled out. Tonal memory accounted for 5.0% and pitch 3.1%. Spelling ability was measured throughout by specially constructed tests having high reliabilities. While these percentages are clearly small, they are nonetheless noteworthy and suggest that auditory abilities may play an important role in L1 spelling ability. This issue re-emerges in *Section 3.3.3* in the context of children having difficulty with L1 reading ability. Interestingly, reading ability was also considered by the early researchers.

Blickenstaff cites two investigations of English reading ability and how it correlates to musical aptitude test results. Ewers (1950) found significant correlations between reading measures (on the Iowa Silent Reading Test and Gray's Oral Reading Paragraphs) and many auditory measures (Seashore Measures of Musical Talents and other specially constructed tests). The Iowa test correlated with a specially constructed test of pitch discrimination for very short stimuli (r=0.70), with the Seashore loudness test (r=0.23) and with another specially constructed test of musical rhythm (r=0.18); the Gray test correlated with the Seashore tonal memory test (r=0.55), with the Seashore loudness test (r=0.25), with the Seashore pitch test (r=0.20) and with the musical rhythm test (r=0.19).

Similarly, Wheeler and Wheeler (1954) found that the Seashore pitch test correlated with a specially constructed test of auditory discrimination for English sounds and with the Metropolitan Achievement Tests of silent reading skills. The auditory

discrimination test further correlated with the tests of reading skills, with coefficients ranging between 0.25 and 0.35. Pitch discrimination was not significantly correlated with general intelligence. While exact significance levels are not given for these correlations, Blickenstaff simply says that they were "significant".

He concludes that a small but consistent relationship exists between abilities in the auditory and musical domain and various native language skills at a number of age levels (Blickenstaff, 1963, p.361). The relationship appears to be independent of general intelligence in studies controlling for this variable.

3.3.2 Music and Second Language Acquisition

Concerning the relationship between L2 ability and musical aptitude, Blickenstaff suggests that the foreign language studies reported in the literature are both too few and too contradictory to warrant any sweeping generalisations (*ibid.*, p.362). However, he cites a study by Karlin (1942) which succeeded in isolating a number of factors on which tests of the musical elements loaded appreciably (.30 to .50). Although second language acquisition was not under investigation, three of the factors mentioned seem logically related to L2 learning ability. These factors are:

1. The ability to understand the native tongue under conditions of masking.
2. Formation of auditory span – the ability to fuse perceived parts of an auditory stimulus into an integrated whole.
3. Memory Span – the ability to remember an auditory span.

The Seashore tests of tonal memory and rhythm appeared in the factor "ability to understand the native tongue under conditions of masking", tests of loudness and time in the factor "formation of auditory span", and tonal memory, rhythm and timbre appeared in the memory factor. It is interesting that these three factors isolated by Karlin are very similar or even identical to those appearing as components of language aptitude in later research (see *Section 2.5.2* above). The ability to understand speech under conditions of masking was explicitly added by John Carroll as a component of language aptitude, while the ability to fuse perceived parts of an auditory stimulus into an integrated whole seems not unrelated

to the question of field dependence/independence[12], although the importance of this distinction has never been fully resolved in the literature on SLA. Given this early suggestion of a link between these language factors and the Seashore measures, it certainly seems worthwhile to consider how language aptitude as defined today might relate to musical aptitude.

Other evidence pointing to a relationship between L2 ability and musical skills comes from Pimsleur, Stockwell and Comrey (1962), who offer substantial evidence that pitch discrimination is related consistently and rather independently to the auditory comprehension of French and probably Spanish. Timbre discrimination also appears to bear some relationship to the auditory comprehension of French. While their findings show that the most important factors in college foreign language learning seem to be verbal IQ and interest (motivation), they conclude that reasoning, word fluency and pitch discrimination also contribute. Indeed the Seashore Timbre test and Chinese Pitch test accounted for 1.3% and 1.5% respectively of the variance in French auditory comprehension.

Another study dealing with the relationship between music and second language ability was that undertaken by Dexter and Omwake (1934). They found that the Seashore tonal memory test correlated with neither the criterion nor the intelligence variable. On the other hand, the Seashore pitch test correlated significantly ($r=0.30$, intelligence partialled out) with the criterion of French accent ratings. It is interesting that this is almost identical to the correlation found between the Bentley Measures of Musical Aptitude and language aptitude scores, when intelligence is partialled out, in the current study (see *Section 5.2.5*). Here, the partial correlation between music and language aptitude is $r=0.28$, $p<0.01$.

The primary aims of the study by Dexter and Omwake were to investigate the existence of a relationship between the ability to discriminate pitch and the ability to pronounce well in French and perhaps other languages. Secondly, the authors wished to establish the existence of a relationship between the number of years of

[12] "In a field-dependent mode of perceiving, perception of the surrounding field, and parts of the field are experienced as 'fused'. In a field-independent mode of perceiving, parts of the field are experienced as discrete from organized ground" (Witkin et al., 1971, cited in R. Ellis, 1994, p.500).

French studies and ability in pitch discrimination. As noted above, they used two subtests of the Seashore tests – Pitch Discrimination and Tonal Memory. Ratings for pronunciation and accent were obtained from the French, German and Spanish departments.

The results of Dexter and Omwake's study show that individuals with poor pitch discrimination tended not to continue to study French to a high level. None of the 29 in the lowest quartile in pitch discrimination and only 28% of those in the next lowest quartile had taken more than two years of French in college, i.e. only 12% of students below the group median in pitch discrimination had taken more than two years of French. However, a student may have good ability in pitch discrimination and still not study French.

With respect to rating in accent or pronunciation and the ability to discriminate pitch, the conclusion that is drawn is that a student is unlikely to have low ability in pitch and a good French accent. However, a student may have high ability in pitch and still have a low rating in accent. Only five subjects of a total of 48 were below the median in pitch discrimination and above average in French. Of these five, only one was in the lowest quartile. To summarise, it could be said that in general low pitch rating leads to low accent rating but the converse is not true, i.e. high pitch rating does not automatically imply high accent rating. It is also shown that intelligence affects accent rating but not ability to discriminate pitch.

In Dexter and Omwake's study, small groups of students studied German and Spanish (approx 20 in each group) so any findings must be treated with caution. It was shown that in Spanish, the ability to discriminate pitch contributes more towards a good accent than does intelligence. However, in German, intelligence seems to contribute more towards accent than does ability to discriminate pitch. The authors do not comment on why this might be the case but one possibility might relate to the 'sing-song' musical nature of the Romance languages. To many linguists and non-linguists alike, Spanish 'sounds' more musical than German. Perhaps this intuitive idea is indeed borne out in the data.

Concerning the relationship between personality and French pronunciation, it was shown that there is a slight tendency for those rated low in introversion to be rated low in accent, and for those high in introversion to be rated high in accent. These findings are somewhat counter-intuitive as it would be expected that more extraverted learners might seek out more opportunities to communicate with native speakers, which in turn would be expected to help them improve their accents. However, findings in this area still remain ambiguous, as researchers have not yet managed to show conclusively the effects of introversion or extraversion on language learning. Some studies show that introversion is advantageous for the language learner (e.g. Dewaele, 2001), while others show an advantage for extraverted learners (e.g. Rossier, 1976). The general conclusion is that further research is needed on this point (see Gass & Selinker, 1994; Skehan, 1989). In any case, the effects of introversion on the language learning process are not directly relevant to the current study.

One study which shows a striking link between musical ability and linguistic ability is that of Eterno (1961). He shows that 90% of eighth graders who played a musical instrument for a year or more scored above average in a pronunciation test, while only 10% scored a rating of average. Not one person who played a musical instrument scored below average.

The primary aim of his study was to compare test score results in musical aptitude and foreign language (Spanish) pronunciation. One hundred and fifty 6th, 7th and 8th graders from a New York grammar school were tested. The musical aptitude test used was the Conn Musical Aptitude Test which consists of tests in rhythm, tempo, pitch, melody, chords, vision and mathematics. The modern language test was a subjective one and involved 6th graders with no previous foreign language experience, and 7th and 8th graders who had had previous foreign language teaching. The 6th graders were given an intensive pronunciation course. After the course each student was tested orally for his pronunciation. Approximately 60% of 6th graders had similar scores in both pronunciation and musical aptitude.

Seventh graders had had six months of foreign language training and were given the same tests as the sixth graders. In line with the results of the 6th graders, 60% of 7th

graders showed similar scores in both music and pronunciation. Eighth graders had had Spanish classes for one and a half years and again there was a high rate of agreement between pronunciation and musical aptitude scores (77% of students had similar scores in both). While the author does not explicitly state what counts as 'similar scores', it appears to be when a student was awarded the same rating (i.e. Above average, Average or Below average) in both language and music. Scoring was carried out as follows: a score of 4 or 5, out of a possible 5, on the Conn Test was classified as Above average, 3 is Average and 2 is Below Average. The total number of words in the Spanish test was 25. A subject was considered above average if he correctly pronounced 20 to 25 words, average if 16 to 19 were correct, and below average when 15 or less were correct. One problem with this study is that there is no control for intelligence so it could simply be the case that more intelligent subjects performed better on tests of music and language.

Eterno (1961, p.169) concludes that there appears to be a direct relationship between musical aptitude and foreign language pronunciation and also that the musical aptitude becomes a more significant factor as instruction in the modern language increases. He goes on to do a comparison of pronunciation to musical instrument training. Results are presented in Tables 1, 2 and 3. The finding of primary importance in all cases is that a greater percentage of students who play an instrument for a year or more receive a rating of above average, compared to the non-musically trained group. Also, when looking at those receiving a below average score, there tends to be a much greater percentage of students who do not play an instrument than those who do. Again, it is possible that there are confounding factors which impact on these results. If musical training takes place outside of school hours, it is likely that it is facilitated by parental support and subject to financial constraints. Parents who support music lessons may, in general, be greater supporters of the arts, and may also encourage foreign language learning to a greater extent than parents with no interest in the arts. While Eterno does not address these issues directly, he is aware that the study is

> "not scientific proof of anything, but merely an indication that a relationship between foreign language pronunciation and musical aptitude and/or musical training may be an existent factor to be contended with for future study" (*ibid.*, p.170).

6th Graders

Pronunciation Test Results	Play a Musical Instrument (1 year or more)	Do not play a musical instrument
Above average	57.2%	52.6%
Average	33.3%	23.7%
Below Average	9.5%	23.7%

Table 1. Effects of Musical Training on Pronunciation Test Scores of 6th Graders

7th Graders

Pronunciation Test Results	Play a Musical Instrument (1 year or more)	Do not play a musical instrument
Above average	71.4%	48.7%
Average	17.2%	35.9%
Below Average	11.4%	15.4%

Table 2. Effects of Musical Training on Pronunciation Test Scores of 7th Graders

8th Graders

Pronunciation Test Results	Play a Musical Instrument (1 year or more)	Do not play a musical instrument
Above average	90%	64.3%
Average	10%	32.1%
Below Average	0%	3.6%

Table 3. Effects of Musical Training on Pronunciation Test Scores of 8th Graders

Leutenegger, Mueller and Wershow (1965) undertook a study of the auditory factors in foreign language acquisition. This study is interesting because it examines both the effects of music aptitude on language learning ability and also the effects of language learning on music aptitude. The specific aims of this study were as follows:

- Measure various auditory factors of University of Florida students enrolled in the beginning courses in French and Spanish.

- Ascertain whether a sex difference exists in any of the auditory aspects measured by the Seashore Measures of Musical Talents.
- Ascertain whether any of the Seashore scores, plus various intelligence and aptitude factors can enable the prediction of ease or difficulty of mastering French or Spanish, i.e. this is an investigation of how musical aptitude affects language learning.
- Ascertain whether scores on the Seashore auditory measures improve significantly after completion of a semester's study of French of Spanish, i.e. this is an investigation of whether language learning affects musical aptitude.
- Investigate Seashore and other possible factors in French and Spanish course drop-outs.
- Ascertain whether significant differences exist between French and Spanish students on any of the above factors.

Tests administered were the Seashore Measures of Musical Talents, the School and College Ability Test, the Committee on Diagnostic Reading – Survey Test, the ETS English Expression Test. Language achievement scores consisted of an average of the number of errors on daily laboratory tests. Results do not reveal any powerful relationships between the Seashore subtests and the foreign language achievement scores. When calculating multiple regression equations for each language separately by sex, it was discovered that for the prediction of female French students, the only independent variable to emerge as significant was Tonal Memory. The authors note that the emergence of Tonal Memory as a predictive criteria would seem to lend support to John Carroll's advocacy of "Phonetic Coding" (see *Section 2.5.2*) as a required ability in learning a foreign language.

It is interesting to note that in the study by Leutenegger *et al*, the only independent variables to emerge as significant were Tonal Memory (for the prediction of results of female students of French) and Total Reading Score (for male students of Spanish). However, Dexter and Omwake (1934), discussed above, found that the test of Tonal Memory made no apparent contribution to accent. Thus, it may be the

case that tonal memory is not related to accent specifically, but is related to language learning more generally.

In summarising the findings on pitch in the various studies, it may be said that findings relating to pitch seem to be somewhat more conclusive than those relating to Tonal Memory. Pimsleur *et al.* (1962) found that Chinese Pitch and Seashore Timbre tests were among the top five most useful tests in predicting auditory comprehension results. Dexter and Omwake (1934) also found positive correlations between Seashore pitch and accent.

A lot of these studies undertaken during the 1950s, 60s and 70s were attempting to find ways to predict success in language learning and in particular to specify the traits which an individual needs to possess to be successful in language learning, e.g. the ability to discriminate pitch. One recent study in this area was conducted along the same lines as many of these earlier studies discussed above (Tucker, 2000). Tucker examined foreign language aptitude in native speakers of English and native speakers of Japanese using the Modern Language Aptitude Test (Carroll & Sapon, 1955) and a translation into Japanese. He also measured musical aptitude using the Seashore Measures of Musical Talent. Finally, he considered second language proficiency by asking the native speakers of English (NSE) for a self-assessment of their foreign language ability. Japanese native speakers (NSJ) also completed a self-assessment of their English ability, in addition to the Wepman auditory discrimination test. Results show a significant correlation between the MLAT Paired Associates test (*see Section 2.5.5*), or its Japanese equivalent, and the Seashore Tonal Memory test ($r=0.32$, $p<0.01$). The MLAT Paired Associates test also correlates significantly with rhythm ($r=0.25$, $p<0.02$). These correlations are considered small to moderate by Cohen (1988), see *Section 5.2.2* below.

Tucker's results also show that the Seashore rhythm test is significant for predicting self-assessment of language ability in the group as a whole, i.e. for predicting proficiency rather than aptitude. When separated into native speakers of English compared to native speakers of Japanese, the MLAT Paired Associates test is more important for predicting self-assessment scores in the native English speaking group while rhythm is predictive for the Japanese group. Tucker concludes that

"...measures of music aptitude and language aptitude are related by common variables inherent in these subtests" (*ibid.*, p.57).

It appears that the ability for rote memorization underlies both language and musical aptitude. Clearly the MLAT Paired Associate subtest requires memorization as it examines Kurdish-English equivalents after a four-minute delay. The music tests also require memorization ability and Tucker suggests that the Tonal Memory test requires greater memorization ability than the Rhythm test, thus accounting for the greater correlation between Tonal Memory and MLAT than between Rhythm and MLAT. However, given the small to moderate magnitude of the correlations I would suggest that it is important not to over-emphasise the importance of the small difference between the two ($r=0.25$ versus $r=0.32$). Tucker does not state whether or not the difference between the two is of statistical significance. In any case, both the rhythm and the tonal memory test require memorisation ability, which is also necessary for the MLAT subtests.

Limitations in Tucker's study relate to the relatively small number of participants in the English group - 34 native speakers of English and 58 native speakers of Japanese. Moreover, it is necessary to question the self-assessment reports as measures of foreign language proficiency. The native speakers of English were simply asked to rate as excellent, good, average or poor, their foreign learning ability under the following headings: Listening; Speaking; Reading; Writing; Memorizing and Motivation. They were asked to think of language in general rather than a specific language. This is clearly not an ideal way of measuring language proficiency. Nonetheless, it is interesting that the results of the study highlight tonal memory as important in the same way as the earlier studies mentioned above.

In many ways, the idea that musical ability may have positive effects on language learning ability, or vice versa, is related to the possibility of overlapping intelligences. As mentioned in *Section 3.1*, Gardner himself says that it is as yet not known how much interaction there is between the different intelligences. The focus of much current research is the possibility of interaction between different facets of intelligence and possible overlap in various cognitive processes. It can be seen, therefore, that while the focus of the earlier studies discussed in this chapter is

somewhat different in that they were searching for components of language aptitude, their outcomes can have much to offer modern research. In some cases, results of more recent studies even mirror those found in earlier work.

3.3.3 Other Recent Research

Douglas and Willatts (1994) show an association between rhythmic ability and reading. Their study was inspired by the knowledge that phonological awareness seems to play an important role in learning to read. They also drew on earlier findings which related phonological awareness to musical sounds; e.g. Bradley and Bryant (1983) showed that children who were better able to discriminate sounds were better readers. Wisbey (1980) suggested that working with musical sounds might help in learning to read. The study by Douglas and Willatts shows that rhythm is significantly related to reading ability and to a lesser degree, to spelling ability. Reading and spelling are not significantly correlated with pitch. It is noteworthy that this finding is at odds with the results of Holmes (1954) cited in *Section 3.3.1* above, who showed that pitch accounted for a certain amount of variance in L1 spelling ability. One possible reason for the difference relates to the age of subjects – whereas Douglas and Willatts worked with young children aged between 7 and 8, Holmes examined high school students. This difference in age groups may contribute to the difference in findings.

Douglas and Willatts also undertook a pilot intervention study in which children with reading difficulties were given training in music. A control group undertook exercises in non-musical activities such as discussion. Significantly, reading scores for the intervention group increased from the initial to the final test, but scores for the control group did not. In order to account for the relationship between rhythm and reading, the authors propose that both may be processed by the same area of the left hemisphere (see *Section 3.4* for further consideration of this question). This is related to the suggestion by Parsons *et al.* (1998), mentioned above in *Section 3.2.1*, according to which location of processing may account for the link between rhythm and spatial reasoning. Of course, simply having a common location of processing does not imply similarity or identity of representation. Indeed, this theme is well developed by Aniruddh Patel (2003), where he tries to reconcile apparently

contradictory evidence from neuropsychology and neuroimaging, as elaborated below.

Neuropsychology presents cases of dissociations between musical and linguistic syntactic processing whereas neuroimaging studies point to overlap between the two. Patel suggests that the representations in musical and linguistic syntax may be quite different, with language using grammatical categories which have no direct analogs in music. This was seen earlier in Chapter 1, where tree structures for language were seen to include labels for constituents. Neither these labels, nor any kind of similar labels, could be applied to musical structures. Patel suggests that there may be domain-specific representations, with different representations for music and language. However, the two varieties of syntax may share some common processes; e.g. structural integration, which involves mentally connecting each incoming element to the previous elements, is necessary when either music or language is encountered. Patel (2003, p.678) calls this the "shared syntactic integration resource hypothesis" (SSRH) and states that

> "overlap in the syntactic processing of language and music can thus be conceived of as overlap in the neural areas and operations which provide the resources for syntactic integration."

Additional evidence in favour of the separation of representations and processing comes from a study showing that Broca's area is important for the phonological processing stage of speech perception, while a region in the inferior parietal cortex acts as the "phonological store" (Frackowiak, 1994). Here again we see the separation of the processing and the storage of representations, with processing taking place in a different location to storage. These findings show the importance of precision when talking about music and language in the brain. It is not sufficient to say that Broca's area is involved in both. It is necessary to give detailed consideration to the exact processes which may overlap and contrast these with functions which are separate. Further discussion of how the brain deals with language and music is presented in *Section 3.4.4* below.

Grogan (1995) carried out a longitudinal study investigating the relationship between cognitive abilities at age four and reading ability at age seven. He found

that auditory sequential memory scores at age four accounted for 13% of the variance in reading scores at age seven. This is related to the findings of studies discussed in *Section 2.7.2.1* above, which show that Working Memory span in young children is an excellent predictor of later language abilities. In fact, it is not only auditory sequential memory which impacts on reading ability. The ability to encode serial order even in the visual domain is important in reading, and difficulties in this area can also lead to problems in spelling and reading (Romani *et al.*, 1999). As a result of his findings, Grogan suggests that it may be the case that children's reading will benefit from training in verbal short-term memory skills, as children with poor short-term memory may have problems in reading.

With Grogan's results in mind, and those of others who have also posited a link between short-term memory and reading, or between short-term memory and general language ability, it would be interesting to investigate the possibility that it is not only verbal short-term memory which is related to reading skills. Perhaps any short-term memory training, e.g. memory for rhythm patterns, could be equally helpful.

Evidence to support this comes from Chan, Ho and Cheung (1998) who found that musical training at a young age causes an improvement in short-term verbal memory in adulthood. They showed that adults with music training have better verbal memory than adults without such training. The adults in this study received musical training before the age of 12 and this resulted in better short-term memory for spoken words. According to Chan *et al.*, the reason for this concerns the change in cortical organisation arising from musical training. As a result of this change, the left temporal area in musicians may have a better developed cognitive function than the right temporal lobe. As verbal memory is mediated mainly by the left temporal lobe, it is improved through musical training which takes place in childhood. Thus, where Grogan advocates training in verbal short-term memory skills to improve reading skills, it seems that verbal short-term memory itself may actually benefit from musical training.

Explicit findings to this effect are presented by Jakobson *et al.* (2003, p.308) who propose that music training strengthens temporal order processing ability. They

define auditory temporal processing skills as "the skills that allow us to make fine discriminations between rapidly changing acoustic events". Temporal order processing skills are developed by musical training. The findings of their study show enhanced verbal memory performance in musicians but they conclude that it is simply a by-product of the music instruction. Hence, where it appears that music directly affects reading or language skills, it may actually be the case that musical training enhances memory or temporal order processing ability which in turn enhances language skills. This issue is important for Chapter 5 where the role of memory in music and language aptitude is considered at length.

Stevenson (1999) examined the association between musical and language aptitude. Significantly, in the light of discoveries by Douglas and Willatts above, she also found a correlation between rhythmic ability and language ability, more specifically the ability to reproduce words in a foreign language. There was also a moderate correlation between the ability to sing back melodies and the ability to reproduce foreign language words but this correlation was not as high as that between rhythmic ability and word imitation ability. In her study a distinction is not made between musical training and innate musical aptitude. Given that the subjects were primary school students, they may have learned to copy rhythm patterns as part of the musical curriculum in schools. In this case, the explanation for her findings may be the same as in the study by Jakobson *et al.* (2003), i.e. rhythmic training developed memory which in turn facilitated the copying of the foreign language words.

Overy (1998) asks the question, "Can music really improve the mind?". She then goes on to cite several studies which seem to provide convincing evidence that the answer to this question is 'yes'. A study by Spychiger (1993) in which children in 50 primary schools in Switzerland received extra music lessons in place of other school subjects over the course of three years, resulted in these children keeping up with their peers in all areas of the school curriculum and performing slightly better than their peers in language and reading. Similarly, Costa-Giomi (1997) (as mentioned above in *Section 3.2.3*) showed the benefits of piano instruction, i.e. two years of piano instruction significantly improved the verbal, quantitative and spatial abilities of 10 – 11 year olds compared with controls.

Lowe (1998) discusses a second language teaching programme which involved the integration of French teaching with music teaching. She proposes that second language education is particularly suitable for an interdisciplinary approach as certain commonalities may exist between the processing, structure and properties of language and music. It is interesting to compare this with a project undertaken by Karimer (1984) twelve years earlier which showed that Asian students of English learned to distinguish between minimal pairs more effectively when the sounds were presented contextually in various songs and chants rather than when the sounds were given in word lists of minimal pairs. Karimer used rhythm and music as a technique for developing short-term memory for phonological learning.

It may be significant that some of Karimer's subjects are Lao Hmong and belong to a society where music plays a pivotal role. Their courtship ritual involves the man singing an original love song to the object of his affection as they throw a ball back and forth to each other. The woman must respond by imitating the man's intonation patterns exactly and by initiating her own. These people are clearly accustomed to paying much attention to musical details. It would therefore seem plausible that their ability to discriminate between similar sounds is highly developed. They have much experience of listening to music and making judgements. As a result, it is perhaps not surprising that music is useful in helping them to discriminate between minimal pairs. Hence, it is perhaps premature to generalise and say that music would be of equal benefit in teaching about L2 minimal pairs in a culture where music does not play such a central role.

From the studies mentioned above, there can be little doubt about the fact that there has been much interesting research into the positive effects of music on cognitive development in areas other than music. However, there are those who highlight the inherent difficulties in this line of research; e.g. Lamont (1998) points out that the process whereby children take up music lessons is influenced by many factors including the motivation of the child, parental support, teacher perceptions of children's abilities, in addition to perseverance and financial constraints. These factors have their own relations with cognitive achievements.

Waters (1998) notes that there are many types of musical activity including listening and performing. Also there are many types of mental process (e.g. logical reasoning, attention, emotion, etc.). In order to get a complete account of how music affects mental functioning, it would be necessary to consider as many cognitive tasks as possible. It is not enough to say that "music" facilitates "cognitive processing", rather, Waters suggests, it must be asked what is the "active ingredient" in music? Also which cognitive tasks are enhanced? It may be that higher level processes shared by music production and perception enhance cognitive functioning, or it may be the case that something specific to either production or perception is behind the effects. Indeed the effects may even be indirect, i.e. music may affect something else (e.g. mood or arousal state) which in turn affects cognitive performance.

A quotation from Lamont (1998, p.203) sums up in a concise manner the current state of this research:

> "In summary, neither 'music' nor 'the mind' have been consensually defined and it thus seems prudent to concur with Sharp's conclusion that at present the scientific case is not proven (Sharp, 1998, p.17)."

Lamont also raises the issue of specific intelligences or the modularity of the mind theory (Fodor, 1983; H. Gardner, 1983). If it is the case that different intelligences are dealt with in different independent modules, why should the development of one module facilitate the development of another? She proposes that the different skills and experiences subsumed under "music" in the research could be considered to span a range of domains, and as such it is clear that any individual skill could facilitate the other skills in the domain. While it is necessary to give this matter further consideration in order to reconcile the modular view of the mind with the idea of interaction between music and other cognitive processes, Gardner's (1997) suggestion that "music may be a privileged organizer of cognitive processes especially among young people", provides a useful starting point for any discussion of how different intelligences may be autonomous to differing degrees. It may even be the case that music has a special status. In any case, as discussed in Chapter 1, the absolute encapsulation of the language module no longer seems fully plausible. Waters (1998) points out the necessity of investigating the acute effects versus the

chronic effects, i.e. how does one "dose" of music affect subjects' performance compared to multiple "doses" over many months or years?

Ultimately, the aim would undoubtedly be to eventually be able to answer all these questions. For the moment, it is in the area of neurology that most progress seems to have been made. Research has dealt with the physical differences between the brains of musicians and non-musicians. Instruction in music has been shown to cause some physical changes. Another branch of research has looked at the evidence in favour of common processing of certain aspects of music and language. These issues are examined in the following section.

3.4 The Brain, Music and Cognition

3.4.1 Introduction

Over a hundred years ago, it was discovered that certain types of speech and language problems were related to damage in specific areas of the brain (Broca, 1861; Wernicke, 1874). This led to the proposal that the left hemisphere was largely involved in language and analytic processes while the right hemisphere was seen as important for processing global information (e.g. patterns). Music was believed to be a right-hemisphere function, while language was a left-hemisphere function. The picture now appears to be far more complex than this, with much complementary functioning by both hemispheres taking place in the processing of both linguistic and musical material (see Maess *et al.*, 2001).

Despite the tradition of associating language with the left hemisphere, the right hemisphere now seems to be important for linguistic prosody and intonation (Blonder *et al.*, 1991; Weintraub *et al.*, 1981). It may also be important for processing non-literal language such as metaphor. The right frontal cortex has also been implicated in language processing, with Zatorre (2001) proposing that it deals not only with musical pitch but also with linguistic pitch processing. However, it has also been suggested that there may be a left-hemisphere bias for the processing of categorical pitch information in both language and music (Patel & Peretz, 1997).

Springer and Deutsch (1998, p.221) account for this apparent divergence of opinion by resorting to the "global" versus "local" distinction. They suggest that

> "although the right-hemisphere was involved with pitch, especially where melodic contour was important, the left played a role where local cues were significant (e.g. with respect to the structure of musical intervals)."

Thus, both right and left-hemispheres appear to play a role in pitch processing although differences between musical and linguistic pitch categories mean that the two forms of pitch may be processed separately within the left-hemisphere. The primary difference between the two is that musical pitch is based on scales which are absent in language.

The functions of the right-hemisphere are no more fully understood than those of its left counterpart. While the right-hemisphere was traditionally considered to deal with musical components, patients with right-hemispheric damage have been shown to have preserved rhythmic production in music (Murayama *et al.*, 2004), suggesting that the right hemisphere cannot be primarily responsible for rhythm. Hence, it is concluded that the left hemisphere is involved with rhythm processing (Milner, 1962; Prior *et al.*, 1990). Other evidence for left-hemisphere involvement in musical ability comes from the fact that left-hemisphere damage impacted on the skills of composer Maurice Ravel, who, after a left-hemisphere stroke suffered substantial loss in his ability to label notes, recognise written music and even play the piano.

Handedness may even be related to individual differences in lateralization; e.g. lateralization of language functions is apparently related to handedness. Research shows that 95% of right-handed people have language localised to the left-hemisphere, while 70% of left-handers show the same pattern. Within the remaining 30% of left-handers, most show some evidence of bilateral speech representation (Loring *et al.*, 1990). However, it is not simply the case that right handers and left handers with left brain language are identical. Springer and Deutsch (1998, p.131) suggest that in the case of brain injury, left-handers may make greater recovery of language functions than right-handers as they may have "the other hemisphere in reserve to a greater extent than right-handers".

It may even be the case that lateralization is related to gender as there is some evidence to suggest that language and spatial abilities are more bilaterally controlled in females than in males (Lansdell, 1962). Thus, it is evident that the organization of language and music in the brain is a complex matter that is as yet not fully understood. The conclusion of Springer and Deutsch (1998, p.221) is that

> "the data on music and the hemispheres suggest that, just as all of the components of language do not appear to be equally lateralized to the left hemisphere, all aspects of musical skills do not reside exclusively in the right hemisphere. Those aspects of musical processing that require judgements about duration, temporal order, sequencing, and rhythm differentially involve the left hemisphere, whereas the right hemisphere is differentially involved when judgements about tonal memory, timbre, melody recognition and intensity are required."

This conclusion highlights the fact that it is in the area of rhythm and temporal processing that language and music are most likely to share neural resources. This may account for the findings of Douglas and Willatts mentioned above which show that it is rhythmic ability which is the area of musical processing most likely to impact on language skills.

Increasing interest in the overlap of language and music continues to shed greater light on the issue of its organization and on how, even from one individual to another, this organization may differ depending, for example, on whether the individual has had musical training or not. The question of musical training is considered in the following section.

3.4.2 Anatomical Differences Arising From Musical Training

With respect to the existence of any anatomical differences between musicians and non-musicians, Schlaug, Jancke, Huang and Steinmetz (1995) showed that people with perfect pitch have a larger left vs. right planum temporale[13] than non-musicians. It seems therefore, that perfect pitch may have a cortical anatomical base. Bever and Chiarello (1974) found that non-musicians depend on the right hemisphere for discriminating melodic sequences, whereas musicians use the left

[13] The planum temporale refers to a portion of the auditory temporal cortex. It is part of Wernicke's area and is important for linguistic function in humans.

hemisphere for this task. From this, they conclude that musicians may process melodies in a more analytic "language-like" manner than non-musicians. However, it must be noted that attempts to replicate this experiment were not always successful (e.g. Zatorre, 1979). If indeed it is the case that musicians and non-musicians process melodies differently, Weinberger (1995) proposes that a transition from right hemisphere to left hemisphere processing of music might be a fundamental aspect of the transition from musically naïve to musically adept.

Schlaug, Jancke, Huang and Steinmetz (1994) found that musicians, and in particular those who had begun their training before the age of seven, had thicker corpus callosums[14] than non-musicians. The strengthening of the corpus callosum provides physical evidence of the relationship between music and cognitive development. Elbert, Pantev, Wienbruch, Rockstrub and Taub (1995) found additional evidence that music training may cause physical effects. They note that the cortical representation of the fingers of the left hand is larger in the right primary somatosensory cortex of string players than of non-string players. The earlier the subjects started playing a stringed instrument, the greater the effect. A more general discovery by Pantev, Oostenveld, Engellen, Ross, Robers and Hoke (1998) is that no matter what instrument is played and even without perfect pitch, the cortical representation is 25% larger in musicians than in non-musicians. From this it can be seen that musical training has positive effects on the brain. However, it is not clear whether only musical instruction would have this effect or if perhaps other training which requires the fingers to move in patterns, e.g. typing, could have equivalent consequences.

3.4.4 Music and Language in the Brain

As mentioned above (*Section 3.2*) in relation to Rauscher *et al.* (1995), specific neural firing patterns exist for different types of activity. It is noted in *Section 3.2.1* that it may be the case that common patterns are involved in music and spatial reasoning tasks; e.g. Leng and Shaw (1991) developed a cortical model according to which certain neural firing patterns organised in a complex spatial-temporal code over large regions of the cortex are exploited by both types of task.

[14] The corpus callosum is connective tissue which connects the right hemisphere to the left.

Gruhn (1998) notes that these characteristic activation patterns which occur in different brain areas in response to learning depend on the way in which students were taught. According to Gruhn (1997) formal musical representations cause a more stable and efficient representation than other teaching modes. Spychiger (1998) cites the fact that frequently stimulated links between brain areas create their own memories. These memories are the carriers of associations between brain areas and she suggests that they may explain the "transfer-effects" of music into other domains. It is interesting that the methods used in teaching music may actually affect how much transfer takes place into other cognitive domains and is something which would require careful consideration by educators hoping for extra-musical benefits of musical training.

Some studies have shown a certain overlap in the processing of music and language in the brain; for example, Maess *et al.* (2001) present results of an experiment which show that Broca's area is involved in the processing of musical syntax. They suggest that this area is involved in processing much more information than was previously believed. While they propose that the left pars opercularis[15] is more involved in the processing of language syntax and that the right pars opercularis is more involved in processing musical syntax, it is important to point out though that both hemispheres are considerably activated by both music and language. According to Maess *et al.*, their study provides evidence of a strong relationship between the processing of language and music and they conclude that this relationship might at least partly account for influences of musical training on verbal abilities (see e.g. Chan *et al.*, 1998; Douglas & Willatts, 1994).

Similarly, Besson and Schön (2003) present evidence in favour of the view that general cognitive principles are involved when aspects of syntactic processing in language are compared with aspects of harmonic processing in music. Also temporal processing involves qualitatively similar processes in language and music, as similar brain areas were activated by violations in both language and music. However, they point towards domain specificity of semantic processing following an experiment which shows that lyrics and tunes seem to be processed in an

[15] The pars opercularis occupies the posterior part of the inferior frontal gyrus. The left pars opercularis is part of Broca's area.

independent fashion. This experiment investigated the brain potentials which occur when subjects listen to opera music. Subjects listened to various excerpts, where the final word of the excerpt was either

A. semantically congruous and sung in tune

B. semantically incongruous and sung in tune

C. semantically congruous and sung out of tune, or

D. semantically incongruous and sung out of tune.

Effects in the double incongruity condition, i.e. D, were not significantly different from the sum of the effects observed in each of the other conditions. The authors propose therefore, that this is strong evidence in favour of the independence of the computations involved in processing the semantic aspects of language and the harmonic aspects of music.

Other evidence of overlap between musical and linguistic processing comes from Patel *et al.* (1998), who show that in musically educated adults, linguistic and musical structural incongruities elicit brain potentials[16] that are statistically indistinguishable in a specified latency range. Thus, whereas the P600 event-related brain potential was originally believed to be language specific, it now emerges that it is also the result of structural incongruities in music. In the study of Patel *et al.*, structural incongruities in language involved the violation of principles of phrase structure, i.e. phrases were either easy, difficult or impossible to integrate with the prior context. Examples are as follows (*ibid.*, p.719):

A. Some of the senators had promoted *an old idea of justice.* (Easy to integrate phrase)

B. Some of the senators endorsed promoted *an old idea of justice.* (Difficult)

C. Some of the senators endorsed the promoted *an old idea of justice.* (Impossible)

In order to generate musical incongruities, a target chord in a musical phrase was either within the key of the phrase or out of key. If out of key, the chord could come from a nearby key or a distant key, as determined by the Circle of Fifths. Interestingly, other researchers have shown that incongruities in numerical sequences elicit the same brain potential (Núñez-Peña & Honrubia-Serrano, 2004).

[16] Evoked Potentials are a measure of the brain's electrical activity in response to sensory stimuli, obtained by placing electrodes on the surface of the scalp (or more rarely, inside the head), repeatedly administering a stimulus, and then using a computer to average the results, ("Brain facts", p.52) http://apu.sfn.org/content/Publications/BrainFacts/brainfacts.pdf (last verified 13.06.05).

Hence, they conclude that a similar neurophysiological process could be required for the processing of violations in numerical sequences and in linguistic sequences.

Thus, it now seems that the process of dealing with syntactic violations may be the same whatever the domain, i.e. whether numeric, linguistic or musical stimuli are involved. This relates back to Patel's proposal (*Section 3.3.3*) which argues for the dissociation of representations and processes. Here we see that the process of dealing with syntactic violations is the same whatever the representations, be they music, language or numbers. While the representations of musical, linguistic or numerical sequences may be different, the processes operating on them may have much in common.

Patel and Peretz (1997, p.191) point out the importance of taking music to be a "confluence of interacting cognitive processes" rather than an "indivisible whole". This issue is relevant to the following chapter which discusses the musical aptitude test used in Phase I of the current study. Where some researchers suggest that musical aptitude is a single ability, the Bentley test allows examination of the components of aptitude. This seems to be in line with the proposal of Patel and Peretz that music is not a single ability. Similarly, language aptitude has been shown to consist of various different abilities, such as phonetic coding ability and grammatical sensitivity, see *Section 2.5.2* above.

Given that music is a "confluence of interacting cognitive processes", it is then possible that some aspects engage domain-specific processes whereas other aspects may share processes with language. Even within what is known as language, there are a range of processes which interact to varying degrees. Thus, while we consider pitch, intonation, rhythm, syntax and semantics to be separate aspects of language, research has shown that prosody and semantics may interact (Astésano *et al.*, 2004). While different brain potentials are associated with semantic mismatch (N400) and prosodic mismatch (P800), the P800 elicited by prosodic mismatch was larger when the sentences were semantically incongruous than congruous, suggesting some interaction between the two processes.

Various studies investigate different aspects of musical structure and their relationship to linguistic structures (see Patel & Peretz, 1997). Aspects under consideration include melody (melodic contour, pitch, tonality), rhythm (tempo, grouping[17], metre) and song. Results suggest an association between performance on musical contour and linguistic intonation tasks. Similarly, results from grouping tasks seem to show that common mechanisms may exist for grouping in language and music. A study by Fries and Swihart (1990) offers support to the existence of a common mechanism governing metrical organisation in language and music. On the other hand, it seems that the processing of tonality is specific to music.

The issue of rhythm in language becomes important in *Section 4.3.1* below, where it is necessary to use a language which is rhythmically different from English, in order to test subjects' ability to imitate rhythm in an unknown language. Classifying language rhythm is by no means without controversy, and the following brief discussion exemplifies the difficulties inherent in classifying languages as belonging to one rhythmic grouping or another.

One common distinction is between stress-timed and syllable-timed languages. However, even the division of languages into stress-timed and syllable-timed is somewhat controversial. While Abercrombie (1967, p.97) states that "as far as is known every language in the world is spoken with one kind of rhythm or the other", it is not always clear which category a language fits into; e.g. Pike (1945) and Hockett (1958) classify Spanish as syllable-timed, whereas Pointon (1995) believes that Spanish does not fit conveniently into either category. More recently, Grabe and Low (2002) found Spanish to be strongly syllable-timed whereas a study by Miller (1984) showed that phoneticians classify Spanish as clearly stress-timed. Thus, Dauer's (1983, 1987) claim that languages are more or less stress-based or Grabe and Low's (2002) proposal that there is a weak categorical distinction between stress and syllable-timing with considerable overlap between the two groups now seems more plausible. Miller (1984) suggests that variations in the way

[17] Individual elements tend to be grouped into larger units, e.g. phrases in music. Also important for speech as boundaries between groups of words have been shown to be marked by pre-boundary lengthening.

language rhythms are perceived may occur as a result of different speakers, different styles of discourse and different tempos, among other factors.

Despite difficulties of classification, it is clear that rhythm is a very important aspect of language; e.g. numerous studies examine how the rhythm of one's native language influences speech processing (e.g. Otake *et al.*, 1993). Thus, speakers of French, English and Japanese seem to use different segmentation procedures based on the rhythmic units of their native language. In addition, the segmentation procedures used by adults appear to be determined by their native language rather than by the language to which they are listening (see e.g. Cutler *et al.*, 1986; Otake *et al.*, 1993). Infants can discriminate between pairs of languages on the basis of rhythm as long as the native language or one of its variants is presented, but not when both languages are equally unfamiliar (Nazzi *et al.*, 2000).

Rhythm is one area which shows evidence of common processing in music and language. Others include grouping, metre and contour. Processes, which at this stage seem not to be shared between the two domains, include tonality and pitch (although pitch is somewhat controversial. Given the presence of scales in musical pitch, it is likely that there must be at least a certain degree of separation between the two). It is as yet too early to draw any conclusions about song as research results are inconclusive. Tempo also awaits further investigation.

Patel and Peretz (1997) highlight the falsity of the argument which uses cases of amusia[18] without aphasia[19] (and vice versa) as evidence of no cognitive overlap between music and language. This argument, they claim, does not give due consideration to the subcomponents of music and language. Secondly, cases of aphasia without amusia are generally found in exceptional individuals such as conductors and composers. Finally, aphasia does not include all disorders of language.

[18] An acquired disorder of music perception or production following brain damage. Related to severe deficiencies in processing pitch variation as well as impairments in music memory, recognition, singing and ability to tap in time.
[19] Disorders of language.

3.5 Conclusion

Weinberger (1998a) says that

> "Music has benefits to intellectual development that transcend music itself...the learning and performing of music is very likely to be of direct neurological benefit."

From the evidence presented in this chapter, this would seem to be the case. The correlation between music and the other intelligences seems to be an area worthy of further research, as results to date look extremely promising. It may well turn out to be the case that "Music" and "Language" are not entirely independent mental faculties but rather that these are labels for complex sets of processes some of which are shared and some of which are not. It is now accepted that the notion of right-hemisphere specialisation for non-verbal auditory functions stemmed from the idea that the cerebral hemispheres are specialised for dealing with entire functions. The possibility that components of these functions, rather than the entire function, might be lateralised differently was overlooked and now seems very likely.

As Patel and Peretz (1997) point out, studying language and music in parallel offers a chance to understand human auditory communication and cognition in a broader perspective than is possible by studying either domain alone. Practical and immediate implications of a correlation between music and language are in the fields of education and speech and language therapy. Of course it would be inadvisable to encourage all schools to given children more music at the expense of English and Maths as this might well lead to a plummeting of standards in these areas. Nonetheless, it may be possible to exploit the potential of music as was done by Spychiger (1993) so that instruction in reading for example, becomes more effective by utilising the extra-musical benefits of musical instruction.

It now seems likely that the improvement in memory caused by musical training may be one contributory factor in accounting for the positive effects of musical training on language learning. This increased memory ability then facilitates language learning. However, it would also be interesting to consider if high levels of innate musical aptitude facilitate successful language learning, i.e. in the case where musical training did not occur, thus did not play a facilitatory role, does

musical aptitude play a role in language acquisition? Perhaps those with high musical aptitude simply have better short-term memory abilities to begin with.

The question of musical aptitude and its relationship with foreign language aptitude forms the central focus of this thesis and is investigated empirically in the following chapters. The central purpose of this dissertation is to offer and apply rigorous psycholinguistic methodology to the quantification of the relationship between music aptitude and foreign language aptitude, independent of general intelligence. Tests of music and foreign language aptitude are employed, in addition to a test of non-verbal intelligence. The relationship between scores on the music aptitude tests and scores on the language aptitude tests are compared. Intelligence is taken into account. Languages such as Czech, Korean and Spanish, which are unfamiliar to subjects, are employed in order to examine their ability to hear sound contrasts and rhythmic differences which are present in these languages but not in English. Subjects are also tested on various aspects of music, such as the ability to discriminate pitch and the ability to imitate rhythms. A detailed methodology is presented in Chapter 4.

Chapter 4 Methodology

4.1 Introduction

A number of studies were presented in Chapter 3 which looked at language learning and musical ability. Some of the more recent studies focused on the effects of musical training on language learning (e.g. Chan *et al.*, 1998; Costa-Giomi, 1999; Spychiger, 1993), while many of the earlier studies investigated musical aptitude as a component of language aptitude (e.g. Blickenstaff, 1963; Eterno, 1961). Many of these studies produced positive results, yet they are too few in number to draw any definitive conclusions about the contribution of musical aptitude to language aptitude.

Given the apparent relationship between the music and language domains, and, in particular, the increasing body of evidence coming from the area of neurology suggesting some commonalities between music and language processing, it seems important to pursue somewhat further the question of association between music and language aptitude. Since the studies carried out in the 1960s and 1970s, there have been many developments in research; e.g. the contribution of Working Memory to language aptitude has been highlighted. It would be interesting to conduct another study into the music-language aptitude relationship which takes into account some of these new developments.

Another area of concern is that many of the existing studies neglected the issue of intelligence and surely this must be examined as a possible confounding variable. In addition, the majority of early studies used the Seashore Measures of Musical Aptitude along with components of the Modern Language Aptitude Test (MLAT). Now that progress has been made in considering what constitutes language aptitude, it is time to revisit the issue of how it relates to musical aptitude.

The current study aims to address the issues raised above. It includes three phases – Phase I investigates receptive skills in music and language, accepting a traditional view of language aptitude. Thus, the aim is primarily to see if results of earlier studies are replicated. Moreover subjects are tested in general fluid intelligence

(gF) so that this may be factored in. It will be seen later in Chapter 5 that a moderate relationship holds between music and language aptitude, and that this relationship holds even when non-verbal intelligence is taken into account. Phase II takes another view of language aptitude, i.e. that the ability to repeat back verbal sequences is a good indicator of phonological short-term memory, which in turn is an excellent predictor of language learning success (see *Section 2.7.2.1* above). As Phase II focuses on productive language skills, it was decided to also focus on productive skills in music. A moderate and significant relationship is found between productive language and musical aptitude. Phase III examines the results of Phase I in conjunction with those from Phase II, i.e. the aim is to consider to what extent productive and receptive abilities overlap. Results of Phase III are less clear cut, revealing that productive skills in music are related to receptive skills in music but that such a relationship does not hold between productive and receptive language aptitude.

This chapter describes the design, scoring and administration of the test battery used in each phase of the study. In addition, details of the sample are provided. The questions of inter-rater reliability, intra-rater reliability and reliability of scale are addressed. Actual results of the study, and some discussion of the consequences of these results, are presented in Chapter 5.

4.2 Phase I - Receptive Skills

4.2.1 Introduction

As discussed in *Section 2.5* above, a learner's level of language aptitude was traditionally viewed as predicting the speed with which a person could learn a foreign language. High value was placed on the ability to discriminate sounds, form memory links between words and the ability to induce rules from linguistic input. Intuitively, and based on the discussion in the previous chapter of how musical skills also use memory, it would seem to be in the areas of memory and sound discrimination that language and music are most likely to coincide.

While it may now be proposed that the Modern Language Aptitude Test is somewhat outdated, as it was most suitable for predicting success in audiolingual

settings, there is no doubt but that it has been shown consistently to be a useful predictor of success in language learning. Therefore, it can be proposed that it is still useful for examining sound discrimination and memory abilities. Despite progress in teaching methods, it is likely that these basic abilities are still important in language learning, although they almost certainly do not constitute all that is necessary to master a language.

Given that sound discrimination and memory are still likely to be important for language learning, in addition to the fact that these seem to be the best potential candidates for overlap between music and language, Phase I of the current study uses the traditional view of language and musical aptitude. Musical aptitude is seen as a group of separate abilities such as pitch discrimination, a sense of rhythm and a sense of harmony. This view of musical aptitude as consisting of individual abilities is in line with researchers such as Bentley (1966b) and Seashore (1939, 1960). It is also compatible with the view of Patel and Peretz (1997) presented in Chapter 3. Language aptitude is understood as consisting of Carroll's (1962) four factors – phonetic coding, grammatical sensitivity, memory ability and inductive language learning.

A standardised musical aptitude test was available for the current study. The Modern Language Aptitude Test itself was unfortunately not available. However, a version of the test had been developed previously in the Centre for Language and Communication Studies, Trinity College with a view to being distributed among first year language students. This modified MLAT contains sections relating to Number Learning, Minimal Pair Discrimination and a Grammar sensitivity test in English (L1). These tests are discussed in detail below.

4.2.2 Design of Test Battery

4.2.2.1 Music

The musical aptitude test used is the Bentley Measures of Musical Abilities (Bentley, 1966a). This is a standardised test designed in the 1960s to measure basic music skills in children, i.e. those skills which children have inherited or acquired incidentally and not through specific musical training. The test consists of four

parts – pitch, tunes, chords and rhythm (Bentley, 1966b). The statistical reliability of the test is 0.84 (*ibid.*).

The pitch test requires subjects to distinguish between different pitch sounds. Given a pair of tones, the subject must say if the second is higher, lower or the same as the first. There are twenty pairs. Answers are either 'S' for same, if the second sound of the pair is the same as the first; 'U' for up, if the second sound moves up, i.e. higher in pitch than the first; or 'D' for down, if the second sound moves down. Some of the intervals used are those normally used in Occidental music, i.e. semitones, whole-tones, thirds, fifths, etc. Others are micro-intervals, i.e. pitch differences smaller than a semitone. The intervals used are given in Appendix 4.

The test of tonal memory is based on the assumption that appreciation of a melody is impossible without the ability to recall, in detail, sounds that have already been heard. Ten items of paired comparison are heard, each half of each item being a five-note tune. In the second half of each item, one note is changed by either a whole tone or a semitone. Subjects must say which note is changed. All notes are equal in length and there are no dynamic accents. The items are given in Appendix 4.

The test of chords consists of ten two-note chords, eight three-note chords and two four-note chords. The subject must say how many notes are heard at once. Bentley (1966b) argues that the ability to analyse chords is not fundamental to melody but he suggests that it is a highly desirable ability. Again, details of the chords used are given in Appendix 4.

The rhythmic memory test consists of ten items of paired comparisons, each half of each item being a four-pulse rhythmic figure. In the second half of each item, one pulse may be changed. The subject must state whether a change has occurred or not. The subject answers 'S' for same if no change occurs. In the case where a change has occurred, the subject must note the pulse of the note which is changed.

Rhythmic memory is measured separately from tonal memory as Bentley suggests that there is only a weak correlation between the ability to appreciate relations of

pitch and relations of rhythm. It is interesting that although this test dates back to the 1960s, these sentiments of separation of rhythm and pitch are completely in line with the current findings from neurology presented in *Section 3.4.4* which suggest separation of pitch and rhythmic processing.

A distinction which often arises in any discussion of rhythm is that between rhythm and metre. Metre refers to the division of notes into evenly distributed note values and is indicated at the beginning of a composition by a metre signature, e.g. 3/4, 6/8, 4/4. Therefore, in 3/4 time, there are 3 crotchet (quarter note) beats in one bar, while in 6/8 time there are 6 quaver (eighth) notes in one bar. Rhythm, on the other hand, refers to stressing the notes according to the musical sense of the phrase (see e.g. Copland, 1957). The difference between rhythm and metre can be seen clearly in the traditional Irish reel and hornpipe. Both may be written in 4/4 time, i.e. the metre is 4/4, yet they have distinctive rhythms. The rhythm is defined in terms of the underlying structure of the metre; e.g. it would be very difficult to have a waltz rhythm written in 4/4 time. Rhythm may be defined as "the temporal and accentual patterning of sound" (Patel & Peretz, 1997, p.202). Tempo, grouping and metre are considered to be subsidiary concepts of rhythm. Patel and Peretz (*ibid.*) suggest that metre and grouping are separate, though interacting, aspects of rhythm.

It seems fair to say that this test is a test of rhythm rather than metre as only 4/4 time is used, i.e. subjects are not asked to distinguish between different metres.

4.2.2.2 Language

The modified Modern Language Aptitude Test contains the sections Sound Discrimination (50 test questions) and Numerical (15 questions) plus an English Grammar test (35 questions). This test is therefore quite time-consuming. In order to ensure co-operation of subjects and to prevent subject fatigue, it was decided to use 16 of the Sound Discrimination pairs and 20 Grammar test questions. Some of the original Grammar pairs were found to be ambiguous or particularly difficult and as such better omitted. Therefore, in conjunction with a senior linguist, the author selected sentence pairs to be included. It is important to remember that the revised MLAT was designed for university students whereas the current study employs secondary school students.

An additional problem with the revised MLAT was that no answers were provided. Accordingly, a native speaker of Czech was recruited to provide answers for the Numerical Test and Sound Discrimination Test. The author provided answers for the Grammar test. As the primary aim of this study is to investigate parallels in language and music, the author believed that it might be useful to consider pitch discrimination in language. Hence, the Chinese test described below was added to the test battery.

Unfortunately as this is not a standardised test, no reliability scores are available for the language test.

Numerical Test

The Numerical test requires subjects to learn a set of numbers in Czech through aural input and then to discriminate different combinations of these numbers. A native speaker of Czech gives number pairs, i.e. English followed by Czech for the digits 2 to 6. The numbers 20, 30, 40, 50, 60 are then given along with their English counterparts. Finally the subject hears numbers 200, 300, 400, 500, 600. The test involves the foreign numbers being jumbled up. When the subject hears a number on the tape, he must write down the digit. There are 15 items in this test. A transcript of the instructions, along with the test items, is given in Appendix 4.

Sound Discrimination

The Sounds test involves the subject deciding which of two words is repeated, given aural input. This tests the subject's ability to discriminate between sounds which are not phonemic contrasts in English. Sixteen items are recorded by a native speaker of Czech. Each item consists of three words. The third word is a repetition of either the first or second word. The subject ticks Box A if the first word is repeated. Otherwise, it is the second word which is repeated so the subject ticks Box B. Instructions and tests items are given in Appendix 4.

Grammatical Sensitivity

The grammatical sensitivity test aims to examine grammatical sensitivity in the native language, i.e. English. Subjects are given a key word in one sentence. They

are then asked to read a second sentence and select the word in that sentence which has the same function as the key word in the first. There are twenty such items; e.g. Given sentences:

- His friend bought **her** a new car.
- Why won't **(A) you** tell **(B) me (C) something (D) yourself**?

the subject must decide which of the words in bold in Sentence 2 fulfils the same grammatical function as the word in bold in Sentence 1. In this case 'me' in Sentence 2 has the same grammatical function as 'her' in Sentence 1, as both act as indirect objects.

Similarly, in the following sentence pair,

- The **basic** rules of the game are not hard to learn.
- **(A) She** had great **(B) difficulty** in finding the **(C) proper** exit she should go **(D) through**.

'Basic' in Sentence 1 is an adjective so the subject should look for an adjective in Sentence 2. Of the possible choices, the only adjective is 'proper'. Hence, 'proper' is the answer. The test is based on the idea that it is not necessary to know the grammatical labels for the parts of speech. Rather the subject should be able to see how certain words have certain functions. The test sentences are provided in Appendix 4.

Chinese

The Chinese test was included in the battery as a result of the author's hypothesis that the capacity to discriminate words which differed only in pitch might be a useful ability in language learning. A native speaker of Mandarin Chinese recorded four words differing only in tone along with their English equivalents (object, house, five, without). Subjects listened to these four pairs four times, before hearing ten Chinese sentences, each containing one of the learned words. Subjects were asked to write the English equivalent of the Chinese word which they had heard in the sentence. A transcript of the instructions, along with the test items, is presented in Appendix 4.

4.2.2.3 Non-verbal Intelligence (gF)

Subjects also completed the Raven standard progressive matrices test (Raven *et al.*, 1998) which tests non-verbal intelligence. The test was designed to measure the eductive component of 'g' as defined in Spearman's theory of cognitive ability.

Eductive ability is

> "the ability to forge new insights, the ability to discern meaning in confusion, the ability to perceive, and the ability to identify relationships. Since perception is primarily a conceptual process, the essential feature of eductive ability is the ability to generate new, largely non-verbal concepts which make it possible to think clearly" (Raven *et al.*, 1998, p.1).

It is proposed that the test is

> "designed to provide a reliable estimate of a person's capacity to think clearly when allowed to work steadily and undisturbed at his or her own speed" (*ibid.*, p.2).

The test originally appeared in the 1930s but has undergone numerous revisions in the interim. Standardisations were carried out on British, American, German, Irish and Chinese populations.

The test is made up of five sets of diagrammatic puzzles exhibiting serial change in two dimensions simultaneously. Each puzzle has a missing part, which the subject has to find among the options provided. Each of the five sets contains twelve problems. Hence, there are 60 questions in the test. The problems become progressively more difficult.

Raven *et al.* (1998) cite numerous studies which find that internal consistency is somewhere in the range of 0.97 to 1.00. Short-term test-retest reliability is said to be around 0.90, reducing to 0.80 at longer intervals. Test-retest reliability is satisfactory for periods of up to a year.

This test is included in the test battery in order to investigate the possibility of both music and language being related to non-verbal intelligence. It seems preferable to measure non-verbal intelligence than verbal intelligence as it would obviously be

expected that language aptitude should be more related to verbal intelligence than musical aptitude. There seems to be a greater possibility that non-verbal intelligence could be equally related to musical aptitude as to language aptitude.

4.2.3 Scoring of Tests

A scoring card is distributed with the Bentley measures and music tests were scored using this card. One mark is given for each correct answer and a composite music score is obtained by adding the scores for Pitch, Tunes, Chords and Rhythm.

Similarly in the language tests, subjects were awarded one point for a correct answer, zero for an incorrect answer. A composite score is calculated by adding the scores for Sounds, Chinese, Numerical and Grammar.

The Raven Standard Progressive Matrices Test was scored in accordance with smoothed British Norms for the Group Test of Children (Raven *et al.*, 1998).

All scoring was done by the author. As scoring was completely objective, it was not deemed necessary to involve any other raters.

4.2.4 Subjects

A group of 20 secondary school students initially took the tests as a pilot group in order to investigate any problems that might arise. Motivation to complete the tests proved to be an issue and in many cases students were not interested in completing the tests independently. The school manager indicated that the participating group had poor results in academic assessments and were generally of low motivation. It was felt that this issue could be overcome by selecting schools which were known to feed students into third level education, as even if not all students follow this track, all students share a similar secondary school atmosphere which values motivation and effort. While actual achievement on the tests is not of major importance for this study, motivation and effort are essential. In order to accurately reflect the relationship between music and language skills, it is important that subjects undertake both tests with equal effort. Otherwise the pilot phase was successful. Data from students in the pilot phase are not included in the overall analysis.

In order to recruit subjects, it was decided to contact post-primary schools as it was likely that they could provide large numbers of subjects who could take the tests in groups. Schools which were relatively convenient to access were chosen from the Department of Education list of second level schools. They were contacted by telephone and invited to take part in the study. It is important to note that schools were not chosen at random from all the schools in the country. Rather, one school familiar to the author in a rural area was contacted, as were several urban schools convenient to access by public transport.

This is clearly a convenient sample rather than a random sample. To select a random sample, it would have been necessary to include all post-primary school students in the country and select randomly from this population. For the sake of convenience, not all students in the country could be included. It was necessary to take a number of students from each school in order to save time travelling around a very large number of schools. There is of course a possibility that this sample is not representative of the population at large. In addition, as many individuals were from the same school, independence of individuals cannot be guaranteed.[20] However, as discussed below, even in this sample there is a mix of males and females drawn from co-educational and single sex schools in both urban and rural areas. Secondary schools and vocational schools are both represented. If indeed aptitude for learning languages can be measured by tests of memory and grammar, it is unlikely that these abilities are different in urban and rural dwellers. Therefore, it is hoped that results can be generalised to apply to all post-primary school students in Ireland.

Approximately half of the subjects came from a co-educational rural school (n=75). All rural candidates came from the same school. The rest of the sample came from urban schools (n=71). Of the urban subjects, 58 were in single sex schools (3 different schools - one all-girls school, two all-boys schools) and 13 were from a co-educational school. In total there were 80 males and 66 females. See Figure 4 for urban/rural distribution among males and females.

[20] In some cases, more than one member of a family may have participated in the study. If music and language aptitude are even partly innately determined, siblings may display similar levels of aptitude. In this way, siblings are not independent, but dependent, individuals.

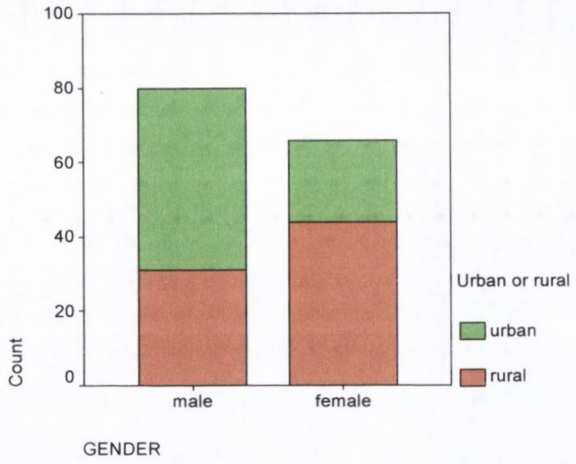


Figure 4. Gender and origin of subjects in Phase I

If the tests genuinely measure innate language and musical aptitude, it is unlikely that the method of choosing the sample should unduly influence results. However, it is possible that to a limited extent there may be bias in the sample, arising from the fact that the sample is not random and that independence of individuals cannot be guaranteed.

Subjects who participated were selected by the school principals, usually on the basis of time-tabling constraints; e.g. one school was enthusiastic about a transition year[21] class participating, as their school days were not as busy as students in the ordinary second level programme. While there was a risk of school managers picking a "good" class in order to ensure that the investigator was impressed by the standard of the school, this does not seem to have been the case, as it will be seen in Chapter 5 that results show a good range and mix of abilities. Subjects ranged in age from twelve years and one month to sixteen years and six months.

4.2.5 Administration of Tests

Testing took place in seven different class groups from five different schools. Three classes were from one school. Class groups were as follows: four first-year classes, one second-year class, one third-year class and one transition-year class.

[21] Transition year is an option which students may take after three years in post-primary school. Rather than following the Leaving Certificate curriculum, for one year, students undertake different activities not normally studied in school, e.g. driving, cookery, entrepreneurial skills.

The Bentley test and the audio components of the language test were played on a Philips tape recorder (model AZ1015) at a comfortable volume.

In the majority of classes (5/7), testing took place on two different days, with music and language tests on one day and the Raven Progressive Matrices test on a subsequent occasion. In the other two classes, students took all three tests on one day, with a short break between the second and third test. While ideally all subjects should have experienced identical testing conditions, with all subjects completing all tests over two days, this was not possible for logistical reasons. This is noted as one of the limitations of the study as it is a source of potential bias[22]. In this case, it is possible that subjects completing the three tests in one day suffered from fatigue which was not suffered by the subjects who completed the tests over a two day period. The latter may therefore have achieved higher scores. Owing to logistical constraints such as timetabling, it was not possible to eliminate this potential source of bias.

In schools where testing took place over two days, a number of students were absent on one occasion. Accordingly, their tests were omitted. In total, 149 completed tests were available for analysis.

4.2.6 Analysis

Phase I resulted in primary variables Language (Receptive), Music (Receptive) and Raven. Secondary variables are Pitch, Tunes, Chords, Rhythm, Chinese, Sounds, Numerical and Grammar. The analysis stage examines firstly the contribution of each subtest (secondary variable) to the overall score on the test of which it is a component, in order to see that each subtest makes a reasonable contribution to the overall test; e.g. the contribution of Pitch to Music (Receptive) is examined, as is the contribution of Sounds to Language (Receptive).

Secondly, the correlations between the three main variables (Language (Receptive), Music (Receptive) and Raven) are considered. The correlations between the three

[22] An unbiased measurement is one which is close to the true value, i.e. the sample value is close to the true value for the population (see e.g. Swinscow, 1997).

main variables are clearly the most important aspect of this study as the primary aim is to investigate the relationship between music and language abilities overall.

An examination of the correlations between the secondary variables and the other primary variables is also conducted; e.g. the correlation between the secondary variables in music (Pitch, Tunes, Chords and Rhythm) and the primary variable Language (Receptive) and the correlation between the secondary variables in Language (Chinese, Numerical, Sounds and Grammar) and the primary variable Music (Receptive). These correlations are of secondary importance but may reveal something which would be worthy of further research.

An analysis is also conducted of how gender influences scores and whether or not correlations between music and language vary in groups with different levels of non-verbal intelligence. Statistical techniques such as partial correlation and regression are used to examine the data. These methods of analysis are explained in Chapter 5 along with the results of the analyses.

4.3. Phase II - Productive Skills

4.3.1 Introduction

The majority of studies investigating correlations between music and language aptitude focus on receptive skills in the two domains. It was therefore decided that productive skills were worthy of investigation as to date little work has been done in this area (although a preliminary investigation conducted by Stevenson (1999) cited above in *Section 3.3.3* produced interesting results in this area). As noted in Chapter 2, the ability to reproduce words in an unknown language is believed to be a good test of Working Memory ability, which in turn is a good indicator of language aptitude. This method of testing language aptitude has now become fairly commonplace so can be considered to be reasonably valid. It was therefore decided to employ such a test in order to examine productive language aptitude in this study.

Results from Phase I were available by the design stage of Phase II and following the discovery of significant correlations between rhythm in music and language test

scores, it was decided to focus on rhythmic skills in music for this stage of the study.

It was also decided to investigate subjects' ability to imitate rhythm in language and so it was deemed necessary to have language stimuli in a language which is rhythmically different from English. In order to see if subjects could successfully imitate rhythm, they needed to attempt longer sequences than simply words. To this end, it was necessary to provide short sentences which subjects could attempt to imitate. Individual words have been used previously and have been shown to be adequate for testing Working Memory ability, whereas something longer was deemed likely to be necessary for investigating rhythm.

While French may seem like an obvious choice for the test language in the current study given its classification as a syllable-timed language, the majority of Irish secondary school students study French. Hence, it was not suitable, as the aim was to select a language unknown to the subjects. Given the fact that Korean is also generally considered to be rhythmically different to English (see e.g. J.-M. Kim *et al.*, 2004) and the availability of a native speaker to record the stimuli, Korean was selected as a suitable language for the purposes of this experiment.

It was noted in *Section 3.4.4* above that classifying language rhythm is often quite controversial. One approach to measuring rhythm which does not involve interstress intervals or syllable durations is the Pairwise Variability Index (e.g. Low & Grabe, 1995; Low *et al.*, 2001). The Pairwise Variability Index (PVI) is calculated using durational variability in successive acoustic-phonetic intervals. The advantage of this approach is that it is an objective measure of language rhythm. Durational variability is greater in stress-timed languages than in syllable-timed languages. Hence, PVI values for English speech are expected to be greater than those for Korean speech, for example. Korean was not however, to the best of my knowledge, specifically investigated by Grabe and her colleagues. It was decided to employ the PVI measure in the current study, as it is an objective measure, and could be compared to the scores awarded by native speakers. It would be interesting to see the degree of similarity between the two.

To date, there appears to be little work done on how perception of speech by native speakers is influenced by PVI values. However, Hutchinson (1973) found that accent judgements by native English-speaking listeners were related to the ratio of stressed to unstressed vowel duration in the English spoken by adult native speakers of Spanish. The Spanish learners who increased the duration ratio to the greatest extent were judged to pronounce English better than those who produced stressed and unstressed vowels with the ratios typical of Spanish. This would seem to be related to the PVI measure which examines ratios of vowels to consonants.

Initially, 15 words and 15 short sentences were recorded in Korean and imitated by a pilot group of subjects. The sentences were segmented and PVI measures calculated. The resulting PVI measures did not seem dramatically different from values for British English reported in Grabe and Low (2002). However, as Grabe and Low did not consider the Korean language, it was difficult to know if the lack of difference was a true effect, or if an error had occurred in the segmentation process. Further examination of the literature revealed that Korean has been considered stress-timed by some authors (e.g. Park, 1990). Therefore, if Korean may be considered stress-timed, it is likely that it might not appear very different rhythmically from British English. A further complication is that subjects in the current study use Hiberno-English rather than British English. Reference values were not available for Hiberno-English.

In order to eliminate as much ambiguity as possible, it was decided to use a language which had been investigated by Grabe and her colleagues and which had been shown to be rhythmically different to English. One such language is Spanish. Therefore, it was decided to use Spanish for the sentence stimuli. As the individual words were not going to be assessed for rhythm, but simply used in order to determine subjects' ability to pronounce them, the Korean words were suitable for inclusion in the test, as they were from a language unknown to the subjects and their rhythmic properties were irrelevant.

Therefore, the current study investigates firstly, whether native English speakers with no knowledge of Korean can imitate Korean words, as judged by native speakers of the language. Secondly, the study investigates whether native English

speakers with no knowledge of Spanish can accurately imitate the rhythm of Spanish. Rhythm is judged by native speakers and by calculating a Pairwise Variability Index. One aim is to see if English speakers imitating Spanish get PVI values similar to that of a native Spanish speaker. Perhaps, on the other hand, native English speakers would get values closer to that of English speech. A third possibility is that PVI values of English speakers imitating Spanish would lie between those of English and Spanish native speakers.

Another question of interest is whether or not the PVI value assigned to the subject correlates in any way with the score given for rhythm by the native speaker; i.e. does the native speaker base his/her judgements on how syllable-timed/stress-timed the utterances sound?

4.3.2 Design of Test Battery

4.3.2.1 Music

A music test was designed based on 16 rhythm patterns taken from a book of Irish folksongs (Bradshaw, 1988). There are 4 patterns in 4/4 time, 4 patterns in 3/4 time, 4 patterns in 6/8 time and 4 patterns in 2/4 time. Patterns were chosen by the author with the assistance of the Head of the Musicianship Faculty in the Royal Irish Academy of Music. Patterns were chosen so as to be representative of standards required for Grade 3 examinations or less as it was felt that any more than this was too difficult for musically naïve subjects. Each pattern was preceded by a pulse to determine tempo, i.e. the 4/4 patterns were preceded by 4 beats; the 2/4 patterns were preceded by 2 beats; the 3/4 patterns were preceded by 3 beats and the 6/8 patterns were preceded by 2 (dotted crotchet) beats. The patterns were generated in Anvil Studio[23] (Version 2003.11.12). A different instrument was used for the introductory beats than the pattern to be imitated.

A screen shot of the program is presented in Figure 5. The time signature of the piece is selected, in the case of this example, 2/4 time. The grid is divided into semiquaver divisions and notes are then placed into the grid; e.g. a quaver covers two boxes as it is twice the length of a semiquaver. A crotchet covers four boxes as

[23] Available at www.anvilstudio.com (last verified 09.02.2005)

it is four times the length of a semiquaver. Figure 5 also shows the two introductory beats played on one instrument (mute triangle) and the pattern itself played on another (low wood block). The file is saved as a MIDI sequence file.

In order to put the files on audio CD it was necessary to have them in .wav format. This conversion was done using a free trial version of the program Audio Compositor Version 4.5.20030512[24].

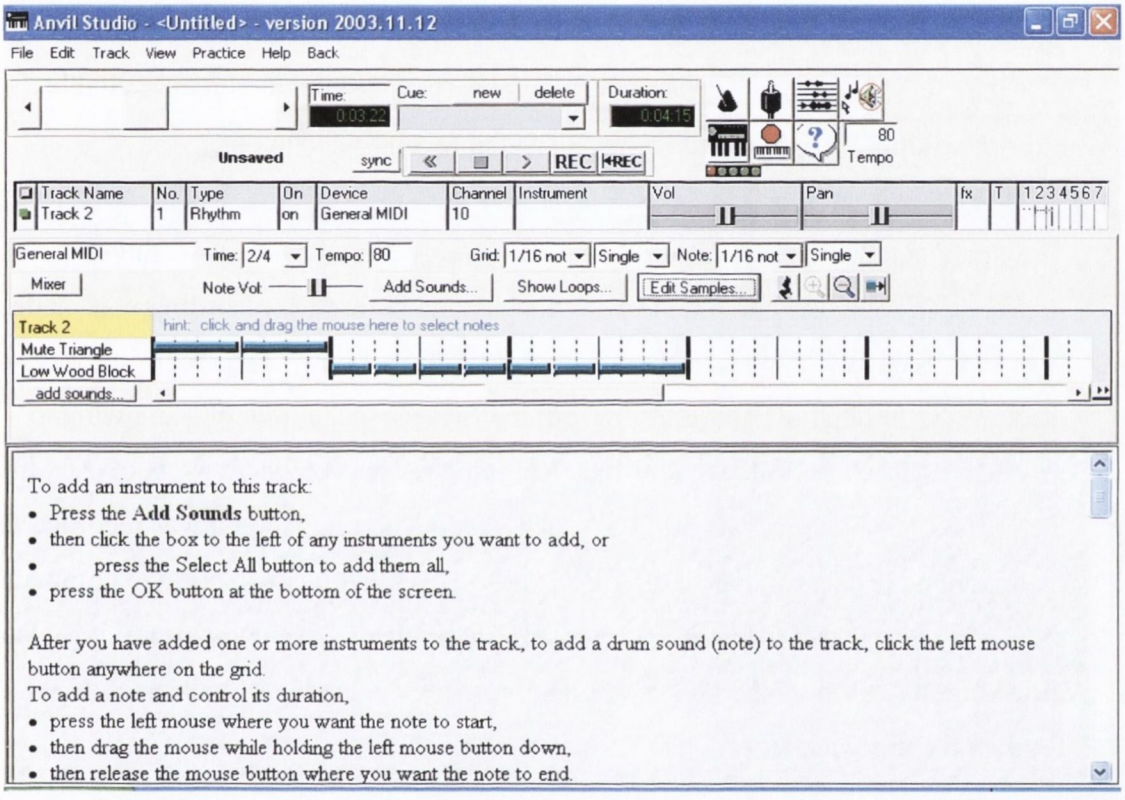


Figure 5. Screenshot from Anvil Studio

4.3.2.2 Language

Spanish

A native speaker of Spanish was recorded reading a passage from which sixteen short sentences were later extracted digitally. The speaker was given time to read the passage before the recording was made. The passage chosen was taken from a beginner's Spanish textbook. This was chosen rather than the standard text for phonetics research ("The North Wind and the Sun"), as sentences in this passage were short whereas those in "The North Wind and the Sun" were likely to be too

[24] Was available at http://audiocompositor.home.att.net. No longer available at this URL.

long for naïve listeners with no knowledge of Spanish to remember. The recording was made in a sound-treated booth in the Phonetics Laboratory, Trinity College.

One sentence was used as an example, leaving fifteen sentences in the test. The fifteen test sentences were recorded onto CD. The sentences are given in Appendix 5. Each sentence was repeated twice, preceded by the number of the sentence. A short gap was left between the repetitions of the sentence, and a longer pause to allow subjects time to repeat the sentence was inserted before each new sentence. Sentence length was as follows: one 5-syllable, one 7-syllable, three 8-syllable, three 9-syllable, five 10-syllable, one 11-syllable, one 12-syllable.

The Spanish text was translated into English and recorded by a native English speaker. The speaker had time to read the passage before the recording was made. It was then made in a quiet room in Trinity College Library. The recording was not made in the Phonetics Laboratory, where the first recording took place, owing to practical difficulties with availability. The same fifteen sentences as were used in the test described above were digitally extracted. The English text was recorded for comparison purposes in calculating PVI values. While it is interesting to compare the PVI values found for Spanish against those published in the literature, it is also interesting to compare the PVI measures found here against PVI measures of English for the same text.

Korean

A Korean test was devised in a similar fashion. A native speaker was invited to prepare a word list of short words. She then recorded the words. Fifteen words were recorded, one to be used as an example and 14 to be used in the test. The words are given in Appendix 5. Of the fourteen Korean words, there were ten 2-syllable words, two 1-syllable words and two 3-syllable words. Piloting of the test with native speakers revealed that one of the words was ambiguous and native speakers needed to hear it a number of times in order to establish which of two words it actually was. Thus, this word is not included in the test for the purposes of analysis, with the result that 13 words were included in the test.

Piloting and Control Group

In order to establish that the tests were of adequately good audio quality, it was decided to pilot them on native speakers of Korean and Spanish. Five native speakers of each language took part. It was felt that this should also reveal any major problems with the tests. Thus, the primary aim of the piloting stage was to detect any problems with the audio quality of the recordings or any cases where pronunciation was not clear even to native speakers. This proved useful as one of the Korean words was proven to be ambiguous and was removed from the test.

The native speaker productions were not scored by rater 1 in either Korean or Spanish (see below). However, when it was decided to recruit a second rater for each language, it was decided to ask them to score the native speaker productions as well as the non-native speaker productions. The responses of the native speakers provided control sentences to be inserted into the audio CDs being scored by the native speaker raters. The raters were not told that some of the responses came from native speakers so in this way it was possible to see the kind of scores that would be awarded to native speakers. This provided a control group against which the scores of the non-native speakers could be compared. It was also a useful check on the quality of ratings awarded by the rater, since if native speakers received low scores, it would be necessary to ask if the rater was paying adequate attention to the stimulus.

In the case of both Spanish and Korean, it was Rater 2 who scored the native speaker productions. The native speaker productions were scored using the same system of scoring as that used for the other subjects. This is described below in *Section 4.3.3.*

The maximum possible score in Spanish is 150. The mean score in Spanish for the 5 native speakers is 149.4, with a standard deviation of 0.89. Thus, it can be seen that the Spanish native speakers were awarded almost perfect scores. The native speaker judge recognised that these speakers imitated Spanish rhythm and pronunciation almost perfectly.

In Korean, the maximum possible score is 65. The mean score for the five native speakers of Korean is 61.2 with a standard deviation of 2.17. This mean score is somewhat low given that it was in fact native speaker productions which were being scored. Below it will be seen that the Korean Rater 2 was not as consistent in her scoring as Rater 1. The fact that she did not award perfect scores to native speakers is further evidence that she may have had some difficulty with the marking scheme. On the other hand, it is possible that these speakers spoke with a different accent which may have influenced the rater. In any case *Section 4.5.1* below shows that the two raters in general showed good agreement and, as Rater 1 was more consistent, her scores are used in the analysis of the data.

4.3.3 Scoring of Tests

4.3.3.1 Music

The music patterns were corrected by the Head of the Musicianship Faculty in the Royal Irish Academy of Music (RIAM). Patterns 1 to 4 are in 2/4 time, patterns 5 to 8 are in 3/4 time, patterns 9 to 12 are in 4/4 time and patterns 13 to 16 are in 6/8 time. The marking scheme for patterns 1 to 12 allowed 1 mark for each crotchet, ½ mark for each quaver and ¼ mark for each semiquaver. The marking scheme for patterns 13 to 16 allowed one mark for each quaver and three marks for each dotted crotchet. The instructions given to the rater were as follows:

1. *No penalty for including introductory beats but no marks given for introduction either, i.e. if pattern is not correct no extra marks can be given for including introduction.*

2. *2/4 pattern (numbers 1, 2, 3, 4) marked out of 4 – one mark for each crotchet, ½ mark for a quaver, ¼ mark for each semiquaver.*

3. *3/4 patterns (numbers 5,6,7,8) marked out of 6 – one mark for each crotchet, ½ mark for each quaver, ¼ mark for each semiquaver.*

4. *4/4 patterns (numbers 9, 10, 11, 12) marked out of 8 - one mark for each crotchet, ½ mark for each quaver, ¼ mark for each semiquaver.*

5. *6/8 patterns (numbers 13, 14, 15, 16) marked out of 12 – one mark for each quaver, 3 marks for a dotted crotchet.*

Clearly, this means that different bars are weighted differently with a bar in 4/4 time receiving twice as many marks as a bar in 2/4 time, i.e. a 2 bar 4/4 pattern is worth 8 marks (i.e. 1 bar of 4/4 time contains 4 crotchet beats worth 1 mark each, thus 8 beats in total for 2 bars), whereas a 2 bar 2/4 pattern is only worth 4 marks (2 crotchet beats worth 2 marks each).

This is not what was intended, and so scores were standardised to take this into account. The marking scheme was arrived at in consultation with the Head of Musicianship and was based on that used in RIAM examinations. Standardisation was done as follows: each score was expressed as a percentage and the percentage was then used rather than the raw score. This should become clear with an example. Suppose a subject scored the following for the first 4 patterns (i.e. those in 2/4 time): 3, 4, 3, 2.5. The maximum possible score for a 2/4 pattern is 4 marks. Therefore, the subject's scores are converted to percentages by dividing each by 4 and multiplying by 100, i.e. (3/4 * 100), (4/4 * 100), (3/4 * 100), (2.5/4 * 100). For the patterns in 3/4 time, the subject's score is divided by 6. For those in 4/4 time, a subject's scores is divided by 8, while for those in 6/8 time, the score is divided by 12.

Once all scores were converted to percentages, the maximum possible score in music became 1600 (16 patterns, 100% each). The minimum possible score is zero. In order to correct the rhythm patterns generated by the subjects, it was necessary to put them on an audio CD for the rater. Each subject had produced the patterns consecutively. Thus, patterns were stored in the form subject 1 pattern 1; subject 1 pattern 2; subject 1 pattern 3; … subject 41 pattern 1; subject 41 pattern 2; subject 41 pattern 3; … . It was necessary to present all examples of pattern 1, followed by all examples of pattern 2, etc. on the audio CD for the rater. In addition, as mentioned above, Anvil Studio creates files in MIDI sequence format, which is not suitable for audio CD. As above, all files had to be converted into .wav files in order to be transferred onto CD.

4.3.3.2 Language

Native Speaker Scoring

Native speakers scored the productions using a 5 point Likert scale. The Korean speaker judged pronunciation only for the Korean utterances. The Spanish speaker gave a mark for pronunciation and a mark for rhythm. Raters were given the scale along side a suggested standard for each score; e.g. the Korean rater was given the following instructions:

Pronunciation Score

5. *Perfect repetition of word as native Korean speaker would have pronounced it.*
4. *Near-perfect repetition of the word.*
3. *A reasonable repetition of the word but with a foreign accent.*
2. *Less than half the word correct but sounds are in the correct order.*
1. *The repetition bore little resemblance to the original (vowels and consonants are in the wrong order).*

The full instruction sheet given to the Korean rater can be found in Appendix 6. The maximum possible score in Korean is 65 (13 words, 5 marks each), and the minimum is 13 (13 words, 1 mark each).

The Spanish rater was given the following instructions for rhythm in addition to similar pronunciation instructions as were given to the Korean rater.

Rhythm Score

5. *Perfect native-speaker-like rhythm.*
4. *Near-perfect rhythm.*
3. *Reasonable attempt to imitate rhythm.*
2. *Some attempt to imitate rhythm.*
1. *Little resemblance to Spanish native-speaker rhythm.*

Again, the full instructions can be found in Appendix 6. Thus, the Spanish rater awarded a score of between 1 and 5 for pronunciation, in addition to a score between 1 and 5 for rhythm. In total, therefore a Spanish production received a score of between 2 and 10 - 2 being the minimum possible if a score of 1 was

awarded to both pronunciation and to rhythm. The maximum possible Spanish score is 150 (15 sentences, 10 marks each) and the minimum is 30 (15 sentences, 2 marks each).

After the two raters had completed their tasks, discussion revealed that the Spanish rater believed it was necessary to use half scores as she felt that she was not able to make adequately fine distinctions using whole numbers. Two additional raters were then recruited for each language. They were advised that they could use half scores if they felt that this was necessary. Thus, instructions for Spanish pronunciation for the second rater were as follows:

Pronunciation Score

Please give a score for pronunciation. You may give a mark between 1 and 5. You may use half scores if necessary.

5. Perfect repetition of the sentence as native Spanish speaker would have pronounced it.

4. Near-perfect repetition of the sentence.

3. A reasonable repetition of the sentence – more than half correct but part incorrect, with a foreign accent.

2. Less than half the sentence correct. Many of the vowels and consonants poorly pronounced.

1. The repetition bore little resemblance to the original.

You can also use 1.5, 2.5, 3.5, 4.5 if necessary.

Similar instructions were given to the second Korean rater. The consequences of this change are discussed below in terms of intra-rater reliability (see *Section 4.5.2).*

Pairwise Variability Index

In addition to the native speaker scores, all the utterances were segmented and a PVI value (Pairwise Variability Index) calculated for each subject. As noted above, PVI is a measure of rhythm. It is calculated using durational variability in successive acoustic-phonetic intervals (Grabe & Low, 2002). As durational variability is greater in stress-timed than syllable-timed languages, a PVI measure shows whether a speaker or a language tends towards stress-timing or syllable-

timing. PVI values for English speech are greater than those for Spanish speech as English is more stress-timed than Spanish.

The equation for calculating normalised PVI is as follows (Grabe & Low, 2002):

$$nPVI = 100 * \left[\sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1}) / 2} \right| / (m - 1) \right]$$

Equation 1. Calculation of normalised Pairwise Variability Index

m is number of items in an utterance;
d is the duration of the kth item.

The equation is used to calculate the difference in duration between each of pair of successive measurements. The absolute value of the differences may vary while the overall ratio remains the same. Thus, the absolute value is divided by the mean duration of the pair. The differences in duration between each pair of successive measurements are summed and divided by the number of differences.

In order to calculate the duration of each of the intervals, wideband spectograms were generated by Praat (Boersma & Weenink, 1992 - 2003) and intervals were measured left-to-right. Vowels were identified according to standard criteria (e.g. Peterson & Lehiste, 1960). In accordance with Grabe and Low (2002, p.15),

> "vocalic intervals were defined as the stretch of signal between vowel onset and vowel offset, characterised by vowel formants, regardless of the number of vowels included in the section. ... Intervocalic intervals were defined as the stretch of signal between vowel offset and vowel onset, regardless of the number of consonants included."

Duration measurements were made on each of the native speaker sentences, as well as on each imitation attempt by each subject. Hence, there were 630 sentences to be segmented (15 native speaker sentences + 41 subjects, 15 sentences each). Figure 5 shows a screenshot of one of the speaker's utterances segmented. The sentence illustrated is 'Aquí Radio Enlace'.

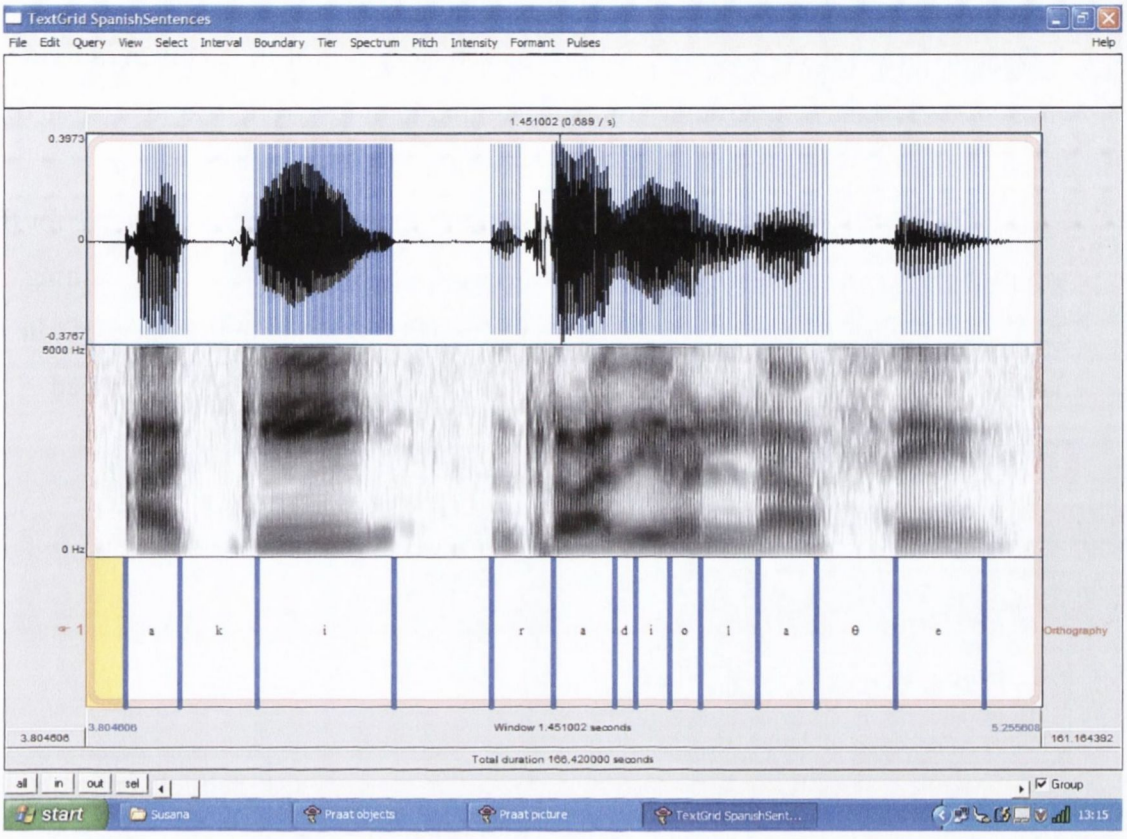


Figure 6. Utterance segmented in Praat

4.3.4 Subjects

As Phase II involved productive skills, tests had to be administered on an individual basis. This is clearly much more time consuming than the group administration of Phase I. As a result, it was impossible to ask all of the original subjects to participate in Phase II. In addition, some schools were unwilling/unavailable for the follow-up, e.g. due to mock examinations, parent-teacher meetings, staff training days, etc. Figure 7 shows the breakdown of subjects participating in Phase II. In *Section 5.3.1* below, it is shown the 41 subjects involved in Phase II are indeed representative of the full sample in Phase I.

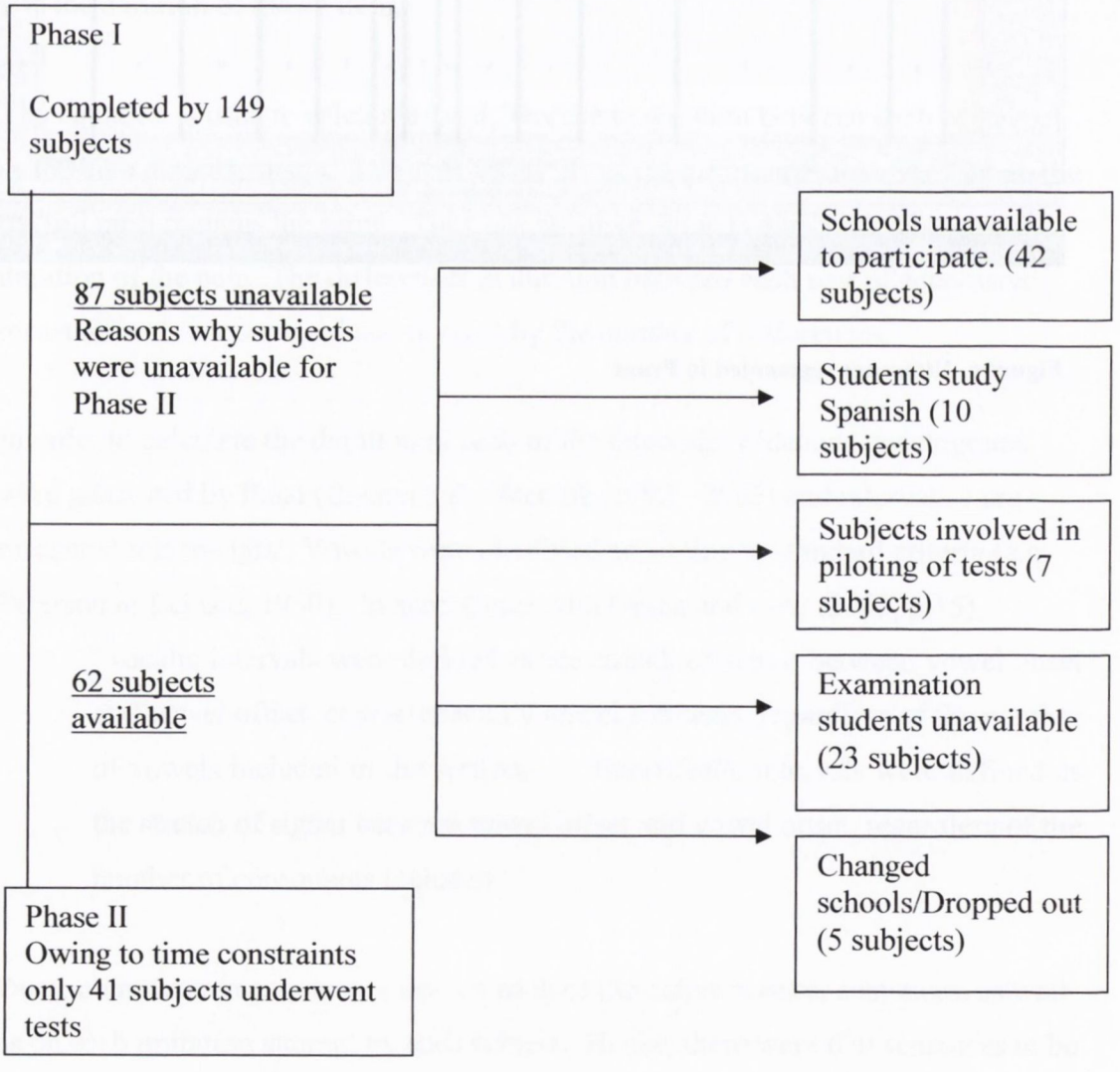


Figure 7. Subjects participating in Phase II

4.3.5 Administration of Tests

Forty-one subjects from three of the original five schools took part in Phase II. All subjects took both the language and the music test on the same day with a short break between the two. The order of the tests was random, i.e. some students took the music test first while others took language first. The order of items within each test could not be randomised for reasons associated with the construction of the tests. While randomisation would have been preferable, there is no reason to believe that order effects should unduly influence results.

In both music and language subjects listened to each item twice and then repeated it. Their attempt to imitate what they had heard was recorded. In the music test, subjects copied the pattern by pressing a key on a computer keyboard. Their imitation attempt appeared in the Anvil Studio program and was saved on the computer by the experimenter. The instructions on the music CD explained that each pattern was preceded by a number of beats to determine the tempo, and that subjects were not required to copy this. Subjects were given the chance to practice. In cases where subjects imitated the lead-in beats, the process was re-explained by the experimenter so subjects were told again that they did not have to copy the introductory beats which were played on the triangle. They were then given another practice opportunity. This continued until subjects got the practice pattern correct. In cases where it was clear that subjects were imitating the introduction yet were still capable of correctly imitating the pattern, they were allowed to begin the test. A transcript of the instructions given to subjects, along with the rhythm patterns used in the test, is given in Appendix 6.

4.3.6 Analysis

Variables of primary interest in Phase II are as follows: Language (Productive), Music (Productive) and PVI scores. Language (Productive) is the sum of the scores on the Korean, Spanish Pronunciation and Spanish Rhythm tests. Raven scores for these subjects are also available from Phase I. The analysis stage examines the relationship between Spanish Rhythm and PVI scores, the relationship between Language (Productive) and Music (Productive) and also how each of these relate to

Raven scores. The contribution of Spanish and Korean to overall language is also investigated. Results are presented in Chapter 5.

4.4 Phase III –Receptive & Productive

Taking variables from Phases I and II, Phase III aims to consider how receptive skills are related to productive skills. Thus, the main considerations are how Language (Productive) relates to Language (Receptive) and how Music (Productive) relates to Music (Receptive). For completeness, Language (Receptive) is also considered in relation to Music (Productive), and Music (Receptive) is considered in relation to Language (Productive).

In considering these correlations, data from 41 subjects are available as the productive scores come from Phase II which had 41 subjects. It is necessary to point out that there may be slight variations in the correlations between the receptive skills found in Phase III from those found in Phase I. This is because 149 subjects were investigated in Phase I and included in calculating the correlation coefficients. Only 41 productive scores are available from Phase II so only these 41 individuals may be considered in Phase III where productive scores are required. However, the correlations between music and language within these 41 individuals may vary somewhat from those in the larger group of 149. In fact this is not a very serious issue as it is shown in Chapter 5 that the group of 41 is indeed representative of the larger group of 149. Phase III can be said to involve 41 subjects.

4.5 Reliability

As tests in Phase II were developed for the purposes of this study they were not previously standardised. When administering tests it is important to know that they measure the construct they are supposed to measure and also that if they were administered on more than one occasion that they would reveal very similar results. The fact that a test measures what it is supposed to measure is known as validity. Validity is addressed in *Sections 5.4.1* and *5.6* below, where correlations between receptive and productive tests are considered. Reliability addresses the consistency of measurements.

When using human judges to score tests it is important to know that different individuals will award very similar scores. This is known as inter-rater reliability. It is also useful to know to what degree the different items within a test are correlated, i.e. do all items in the test measure the same underlying construct. This is known as internal consistency reliability. These issues are investigated in this section. Two other forms of reliability are test-retest reliability and parallel forms. Test-retest reliability is used to assess the consistency of a measure from one time to another. Parallel-forms reliability is used to assess the consistency of the results of two tests constructed in the same way from the same content domain. Time constraints meant that these could not be investigated in the current study.

4.5.1 Inter Rater Reliability

The aim of this section is to investigate whether the two raters for each language were in agreement on the scores which should be awarded to each subject. It was not possible to get two different raters to score the music tests. However, it was believed to be more important to get a number of raters for the language tests as judging pronunciation and rhythm is not something which a native speaker is really trained to do. The rater of the music tests on the other hand, is trained to judge musical productions as she regularly scores music exams.

Probably the greatest concern was that the raters should at least rank subjects in the same order, i.e. that both raters should agree which subjects were best and which were worst, even if they did not give the same absolute scores. Whether or not absolute scores are the same or different is less important, as the overall aim is to determine if those subjects judged good at the language tests were also deemed good in music. Hence, it is rank rather than absolute score which is important.

4.5.1.1 Korean

Figure 8 shows that Rater 1 generally awarded higher scores than Rater 2. It shows the median scores awarded for each student. Boxplots are described in detail in the next chapter in *Section 5.2.3*. It is seen here that in the case of Rater 1, the median score for each student is clearly higher than the median score awarded by Rater 2. However, as mentioned above, this is not a major problem so long as subjects were ranked in the same order.

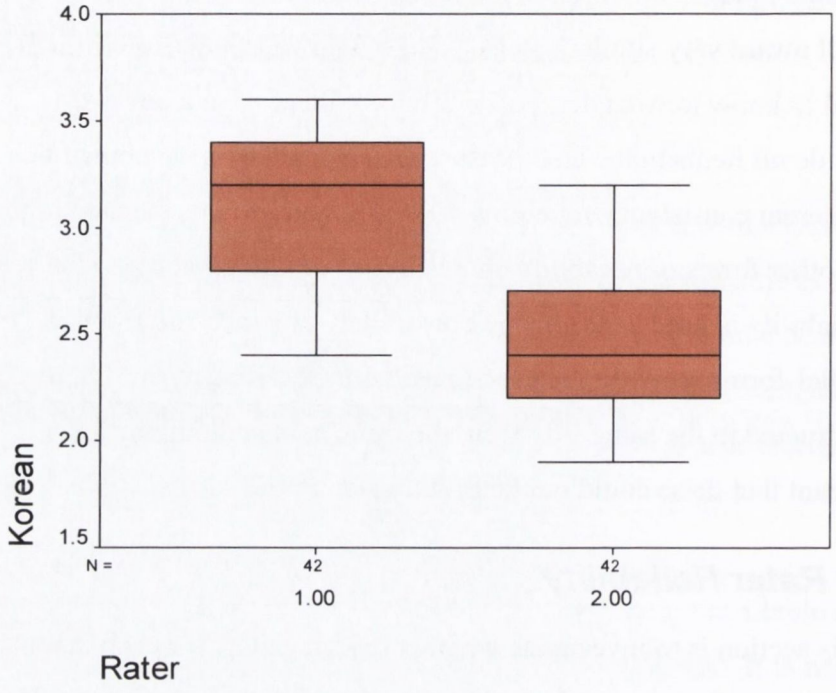


Figure 8. Boxplot of Korean scores

Figure 9 again shows that Rater 1 gave higher scores than Rater 2 but it also shows that the raters were generally in agreement as to who should receive high and low scores. This is seen from the fact that the lines are relatively parallel. In Figure 9 each data point represents the mean score for one particular student.

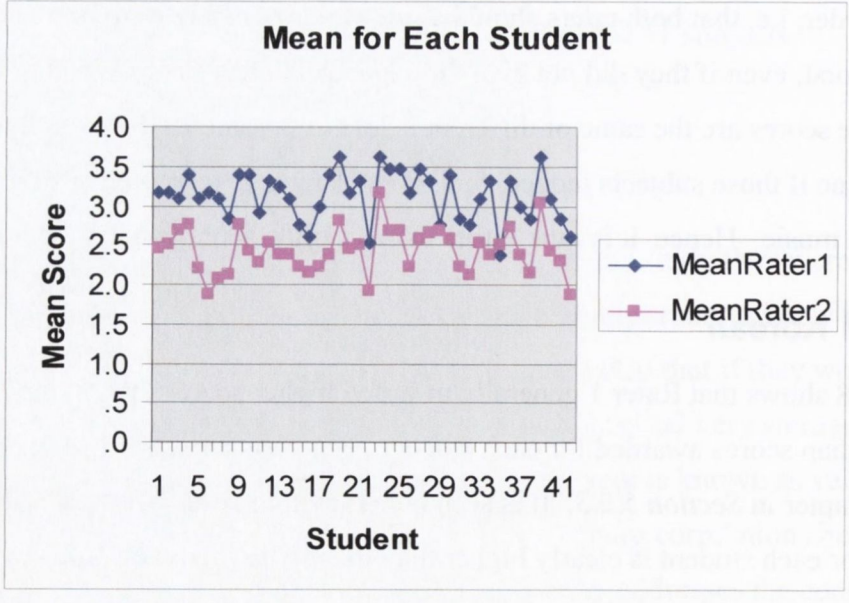


Figure 9. Korean raters' mean scores for each subject

As mentioned above, the most important question is whether or not the two raters agree on which subjects deserved high scores and which deserved low scores. Figure 9 shows that the raters indeed followed the same trend, awarding high scores to some subjects and low scores to others. They generally agreed on who these subjects were.

A more formal measure of this association is the Pearson Correlation which measures the linear association between two variables. If the two raters agreed perfectly this correlation would be one. In the absence of any agreement, this correlation would be zero. Correlation is discussed in more detail in *Section 5.2.2* below. In the case of the Korean raters, the Pearson correlation is 0.6 ($p<0.01$). This is a large correlation (see Chapter 5, *Section 5.2.2* for discussion of correlation magnitudes), which means there is very good agreement between them. In this case, it is also useful to look at the rank correlation, i.e. to consider whether the two raters rank the subjects in the same order. This is Spearman's rho and here is equal to 0.62 ($p<0.01$).

4.5.1.2 Spanish

Figure 10 shows that Rater 2 awarded slightly higher median scores to each student in the Spanish test. The difference however is very small.

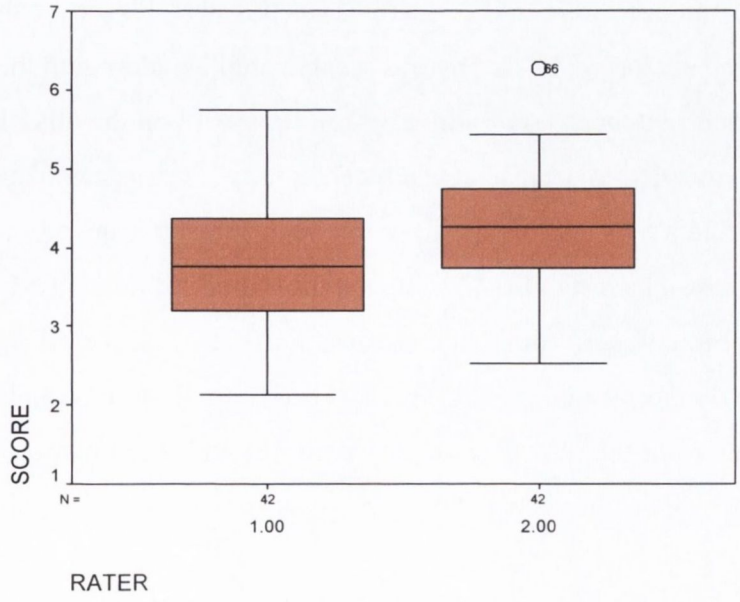


Figure 10. Boxplot of Spanish scores

This is also highlighted in Figure 11 and Figure 12, where it can be seen that the two raters awarded very similar scores indeed. Figure 11 shows the mean score for each student given by each rater. It can be seen that the means for both raters are indeed very similar.

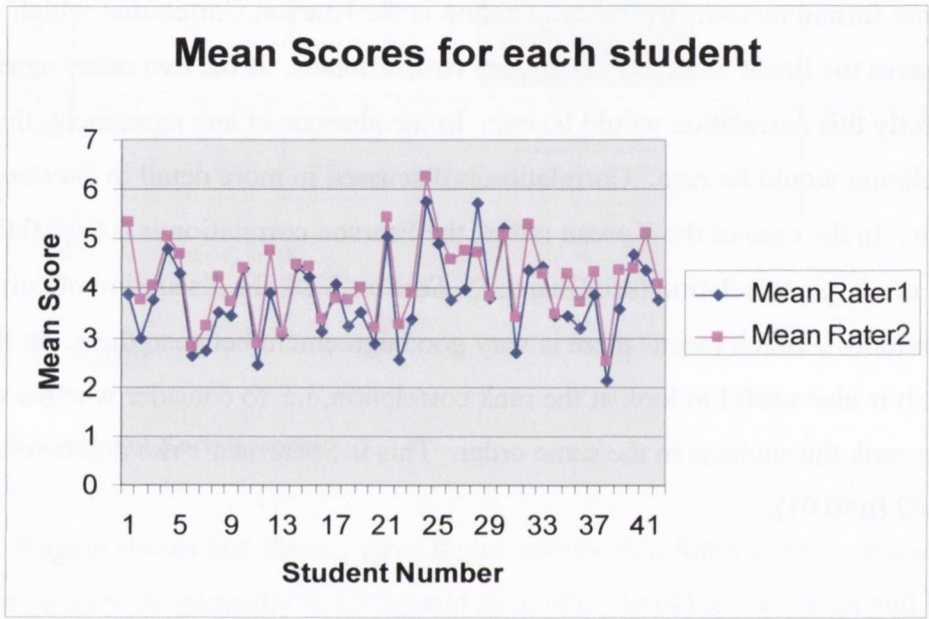


Figure 11. Spanish raters' mean scores for each subject

Figure 12 shows the mean score given by each of the two raters for each sentence. Thus, the difference in Figures 11 and 12 is as follows: Figure 11 shows the mean score for each subject. This is obviously calculated by adding up the subject's scores on each sentence and dividing by 15. Figure 12 on the other hand, shows the mean score awarded to each sentence by each rater. This is calculated for a particular sentence by adding up the scores each subject achieved on that sentence and dividing by 41. In Figure 12 it is seen that while for Sentence 1 and Sentence 2, Rater 2 awarded higher scores, by Sentence 3, the two raters converged. In Sentence 6, there was again a discrepancy with Rater 2 awarding a higher score than Rater 1. However, the overall trend is for similar scores to be awarded by both raters.

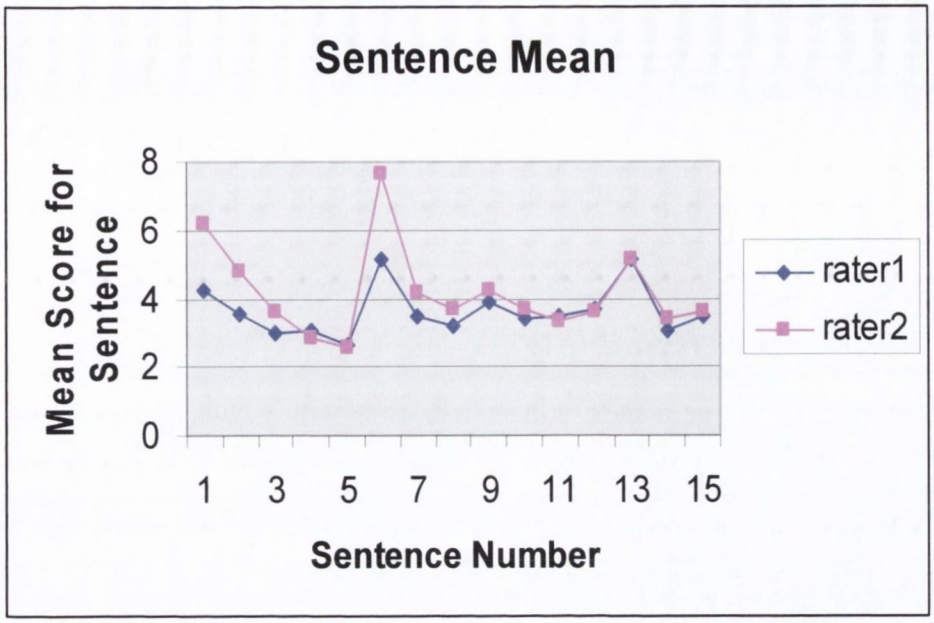


Figure 12. Spanish raters' mean scores for each sentence

The Pearson correlation between the two raters is 0.72 ($p<0.01$), while Spearman's rho is 0.71 ($p<0.01$). Thus, there is a very strong linear relationship between the two.

4.5.1.3 Kappa Statistic

While the Kappa statistic is commonly used for measuring rater reliability, it was decided in this case to use correlation instead. This decision is based on the fact that Kappa examines exact agreement between raters; e.g. how many times both raters awarded a score of 2 versus a score of 3. In this case, as mentioned above, the absolute scores are not as important as the rankings. Thus, if Rater 1 consistently gave a score two points above the scores of Rater 2, Kappa would find poor agreement whereas for the purposes of this study such a result would be better considered excellent agreement.

In addition, the second rater for both Korean and Spanish used 0.5 divisions of the rating scale whereas Rater 1 used whole numbers. This reduces the number of times where both raters will have awarded exactly the same score. While the scores of the second rater could be rounded to the nearest whole number, this would result in a loss of information.

4.5.1.4 Conclusion

Given the fact that there is good agreement between raters, the author feels justified in using the scores of Rater 1 for both Spanish and Korean in any further data analysis.

4.5.2 Intra Rater Reliability

Another important issue relates to the consistency with which raters awarded their scores. In order to investigate this issue, each rater rated a number of items twice although they did not know that items were repeated within the test.

4.5.2.1 Korean

Rater 1

Rater 1 rated 13 items twice. Figure 13 is a bubble chart where the size of the bubble represents the number of data points appearing at this point; e.g. there are 4 cases where Rater 1 awarded a score of 3 the first time she corrected the sentence and also the second time. Thus, there is a large bubble on the chart at [3, 3]. There were three occasions when she awarded a score of 2 on both occasions, 2 occasions when she awarded 3 the first time and 2 the second, 2 occurrences of [4, 3] and one each of [4, 4] and [5, 2]. The number in the centre of the bubble represents how many times this particular combination occurred.

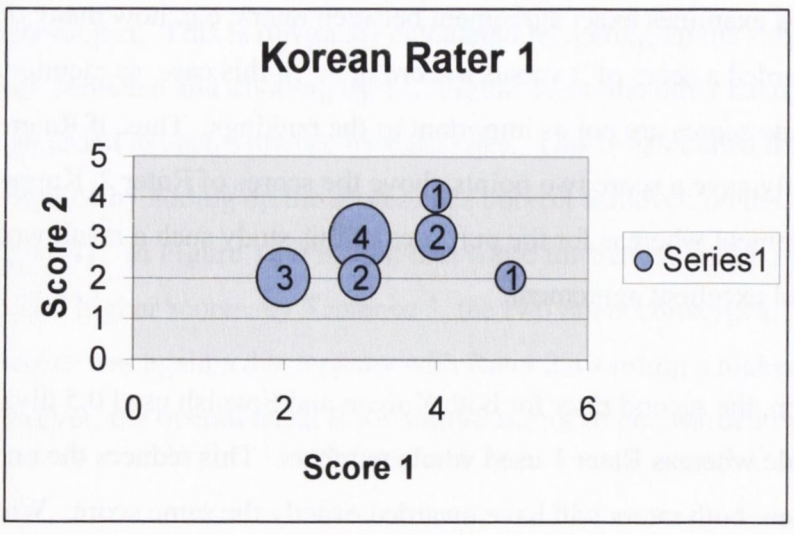


Figure 13. Bubbleplot for scores of Korean Rater 1

From this chart, it appears that the point [5, 2] is an outlier as in general this rater was quite consistent and normally awarded either the same score the second time around or a score which only differed by one from the score awarded the first time. Therefore, for the purposes of doing a Pearson correlation, this outlier is omitted, resulting in a highly significant and very strong correlation between scores awarded the first time and those awarded the second time around ($r=0.76$, $p<0.01$).

Rater 2

Rater 2 rated 21 items twice. Figure 14 shows that Rater 2 made much greater use of 0.5 values, e.g. 2.5, 3.5, etc. Given that this graph shows that she was not as consistent as Rater 1 in her use of scores, it seems that allowing half values introduced unnecessary complexity as the rater was not able to use them consistently. The Pearson correlation between the scores awarded at time 1 and at time 2 is 0.55 ($p<0.01$) and Spearman's rho is 0.59 ($p<0.01$). As will be seen in Chapter 5, this is considered a large correlation by Cohen (1988). However, it is clearly not as large as that found in the case of Rater 1.

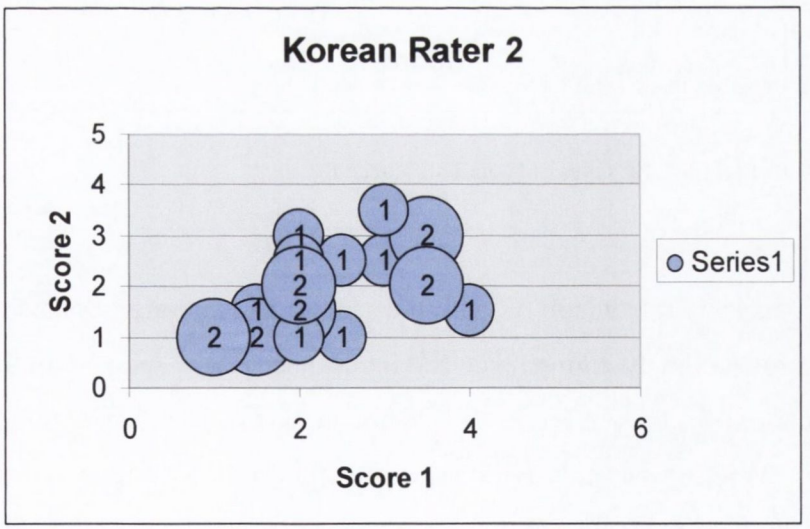


Figure 14. Bubbleplot for scores of Korean Rater 2

4.5.2.2 Spanish

Rater 1

Rater 1 rated 15 items twice. Figure 15 shows excellent consistency in the use of scores by Rater 1, with only one instance of the second score not being the same as the first score. In this case the Pearson correlation is 0.97 ($p<0.01$). Similarly, the Spearman's rho is equally high, as the score on the second occasion almost always ranks subjects in the same order as the score on the first occasion (rho=0.97, $p<0.01$).

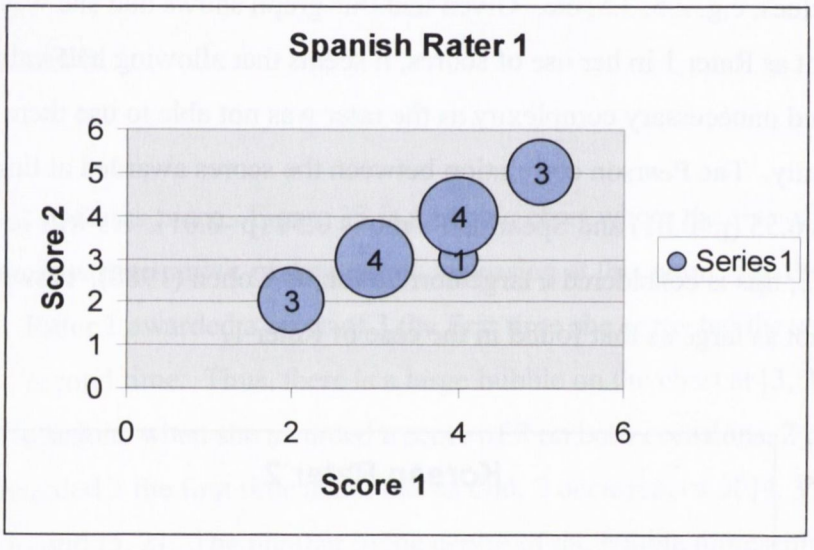


Figure 15. Bubbleplot for scores of Spanish Rater 1

Rater 2

Rater 2 rated 15 items twice. As with the second Korean rater, Spanish Rater 2 was instructed to use 0.5 divisions if she felt it was necessary. As a result there is a greater scatter of scores in Figure 16. While her scoring system is not as consistent as Rater 1, there are certainly no instances of high scores being awarded first time round and low scores the second. The Pearson Correlation is high ($r=0.76$, $p<0.01$) although Spearman's rho is somewhat lower at 0.57 ($p<0.05$). This means that although the scores are associated, subjects were not always ranked in the same order by Rater 2 the second time round.

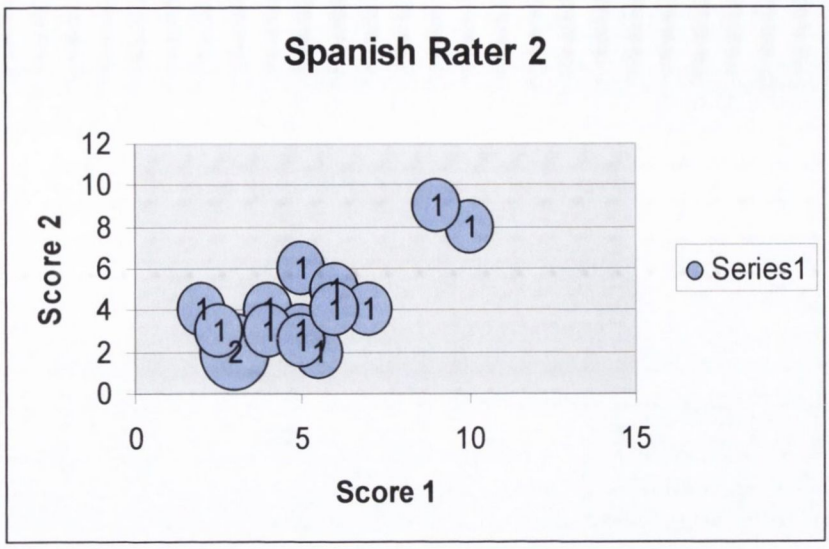


Figure 16. Bubbleplot for scores of Spanish Rater 2

4.5.2.3 Kappa Statistic

With respect to intra-rater reliability, it may be useful to examine the Kappa statistic. Krippendorff's alpha (α) is one particular calculation of the Kappa statistic. It is a reliability coefficient developed to measure the agreement between observers, coders, judges, raters or measuring instruments. The advantage of the Kappa statistic is that it corrects for chance, i.e. by chance alone it is likely that raters will agree some of the time and it is useful to factor this into the equation. The value of kappa (or alpha if using Kripendorff's calculation) can range between -1 and 1, where 1 indicates perfect agreement, -1 indicates perfect disagreement and 0 indicates that the raters agreed no more than was predictable by chance.

The general form of Krippendorff's alpha is as follows:

$$\alpha = 1 - \frac{D_0}{D_e}$$

Equation 2. Calculation of Krippendorff's alpha

D_0 is the observed disagreement;

D_e is the disagreement one would expect when the coding of units is attributable to chance.

Krippendorff's alpha (α) is proposed to be indeed appropriate in this situation although it was not above in the case of inter-rater reliability. This is because, when comparing a rater's scores against her scores the previous time, it is hoped that a rater will award identical scores. Therefore, it is hoped that high values of alpha are found.

Grove *et al.* (1981, p.412) point out that it is difficult to decide what counts as a high value of alpha, saying that:

> "A further complication is the fact that no useful statistical test is available to determine when reliability is acceptable. It can be determined that it is significantly better than chance, but this information has very little meaning in clinical or research settings. While investigators often make the simple assumption that a Kappa value of more than .5 or .6 is acceptable, in fact the meaning of all estimates of reliability will depend on all the factors that have been discussed in this report, such as stringency of design, base rate, sensitivity and specificity, and the coefficient chosen."

As mentioned above, it was only possible for each rater to rate a small number of items twice, i.e. 13 in the case of the Korean Rater 1 and 15 in the case of the Spanish raters. Ideally more items would have received two ratings. Values of Krippendorff's alpha for each rater are given below. These were calculated using a Perl script which is provided in Appendix 7 along with all output.

Korean Rater 1

Korean Rater 1 rated 13 items twice. Kappa is found to be 0.42, or 0.69 when the outlier above is omitted. Leaving out the outlier shows good agreement from one time to the next.

Korean Rater 2

Kappa is found to be 0.52. This value is clearly lower than that of Rater 1. This is likely due to the fact that Rater 2 had the option of using half scores. As a result of this option, she may have given identical scores fewer times than Rater 1.

Spanish Rater 1

Kappa is 0.96. This is an almost perfect score and shows excellent agreement from time point 1 to time point 2.

Spanish Rater 2

Kappa is 0.48. As in Korean, the value of Kappa is lower for Rater 2 than for Rater 1. Again it is possible that this is due to the option of using half scores.

4.5.2.4 Conclusion

This section has investigated the consistency with which raters awarded their scores. It is shown that the values of alpha for Korean Rater 1 and Spanish Rater 1 are high and certainly acceptable according to the standards used in the literature of accepting alpha greater than 0.6 (e.g. Grove *et al.*, 1981).

However, the values of alpha are lower in the case of Korean Rater 2 and Spanish Rater 2. It is likely that this is due to the fact that the second raters were instructed to use 0.5 divisions of scale if necessary. This instruction may have resulted in too many potential scores and the rater could not differentiate between, for example a score of 4 and a score of 4.5. The outcome of this is that the rater did not consistently use the scoring system so in many cases where the first time she awarded a score of 5, the second time she awarded a score of 5.5. As the first rater was more consistent both in the case of the Spanish and Korean tests, the scores of rater 1 are used in calculating results of the study. In addition it was shown that the raters generally show a high level of agreement with each other which further supports the use of Rater 1's scores for both Spanish and Korean.

4.5.3 Reliability of Scale Analysis

It is important to consider the reliability of the measurement scales used, i.e. it is necessary to ensure that all the items forming the scale measure the same thing. This property of a scale is known as internal consistency. If all the items in a scale measure the same underlying construct, the items will be correlated with each other. Cronbach's alpha is a useful coefficient for assessing internal consistency, i.e. it is used to measure how well a set of items (or variables) measures a single

unidimensional latent construct. Cronbach's alpha takes values between 0 and 1. The nearer the value is to 1, the higher the internal consistency of the scale. According to DeVellis (1991) an alpha below 0.60 is unacceptable; 0.60 - 0.65 undesirable; 0.65 - 0.70 minimally acceptable; 0.70 - 0.80 respectable; 0.80 - 0.90 very good; 0.90 excellent but may mean that the scale is too long.

4.5.3.1 Korean

Cronbach's alpha is calculated both for Rater 1 and for Rater 2. In the case of Rater 1, α=0.42 while in the case of Rater 2, α=0.06. These values are extremely small and imply very low internal consistency. This means that a subject could receive a very high score on a number of items but very low scores on other items. These values show that internal consistency is not acceptable for Korean. It seems that not all items on the Korean test are measuring the same underlying ability.

4.5.3.2 Spanish

Cronbach's alpha is calculated for both Spanish raters. For Rater 1, α=0.84 and for Rater 2, α=0.80. These values are much higher than in the Korean scale above and show very good internal consistency of the scale.

4.5.3.3 Music (Prod)

Cronbach's alpha is 0.86 when calculated for the items in the Music (Prod) test. This shows very good internal consistency for the music scale.

4.5.3.4 Summary

Internal consistency is very high and reaches an acceptable level in the case of the Spanish and Music tests. Internal consistency in Korean is very low and unacceptable. This must be kept in mind in Chapter 5 in discussion of the results when trying to establish the validity of the language tests.

4.6 Conclusion

This chapter has described the methodology employed in the current study. Potential limitations of the study have been acknowledged such as the method of

choosing the sample, sample size in Phase II and the fact that tests were not standardised on a large sample prior to use. In addition, test-retest reliability could not be considered. However, despite these limitations this chapter aims to show that care was taken at all times to minimise potential sources of bias. In addition, issues such as inter-rater and intra-rater reliability were addressed. Therefore, it is hoped that the reader may have confidence in the results of the study. These results are presented in the following chapter.

Chapter 5 Results and Discussion

5.1 Introduction

This chapter presents the results of the tests carried out as described in Chapter 4. The chapter consists of three main sections which detail the results of Phases I, II and III.

Phase I of the study examines receptive skills in language and music. In addition, gender and the non-verbal intelligence of subjects are also considered. Phase II investigates productive skills in language and music. Again, language and music skills are considered in relation to non-verbal intelligence and gender. Phase III examines how productive skills relate to receptive skills in the two domains.

Having presented the findings of the study, there are then two short sections discussing the problems of multiple testing and the validity of the language aptitude test used in Phase II. This is followed by the conclusion and suggestions for further study.

5.2 Phase I – Examination of Receptive Skills in Music and Language

Variables examined are Music (Recep), Language (Recep) and Raven. Music consists of four subtests - Pitch, Tunes, Chords and Rhythm. Language also consists of four subtests - Sounds, Chinese, Numerical and Grammar. I will firstly examine the data set as a whole (i.e. examine data from all 149 subjects together) and later consider how scores in music and language vary as a function of intelligence.

5.2.1 Distribution of scores

It is necessary to consider the distribution of scores for each variable as it becomes important when carrying out statistical functions such as correlation. Distribution

of scores for the three main variables (Language (Recep), Music (Recep) and Raven) are presented and commented upon in this section.

In examining the distribution of scores, it is useful to consider whether or not they conform to the normal distribution. The normal distribution is a symmetrical bell-shaped curve around the sample mean (or average). The extent to which the curve is compressed or flattened out depends on the standard deviation of the sample. The standard deviation is a summary measure of the differences of each observation from the mean. In other words, the standard deviation gives information about the spread of scores around the mean. If scores are widely spread the curve will be flatter than if scores are all very close to the mean, in which case the curve will be peaked.

Consider for example the following two data sets: {80, 20, 7, 3, 6, 4} and {21, 19, 18, 22, 20}. Both data sets have a mean value of 20. However, inspection reveals that the range of values is much greater in the first data set, i.e. the values are much more spread out from the mean. This is formally indicated by the standard deviation which is 30.03 for the first data set but only 1.58 for the second.

Let us now examine the distribution of Music (Recep), Language (Recep) and Raven. Music (Recep) is approximately normally distributed, with mean 33.8 and standard deviation 7.76, see Figure 17.

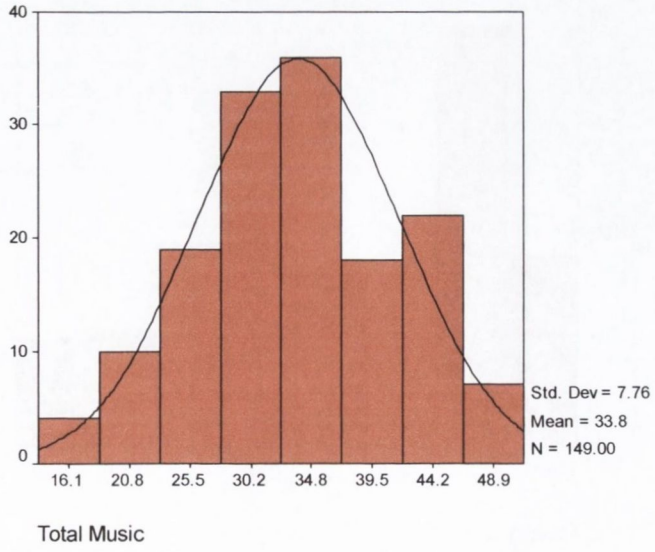


Figure 17. Distribution of Music (Recep)

Distribution of Language (Recep) scores is approximately normal, with mean 29.7 and standard deviation 6.74, see Figure 18.

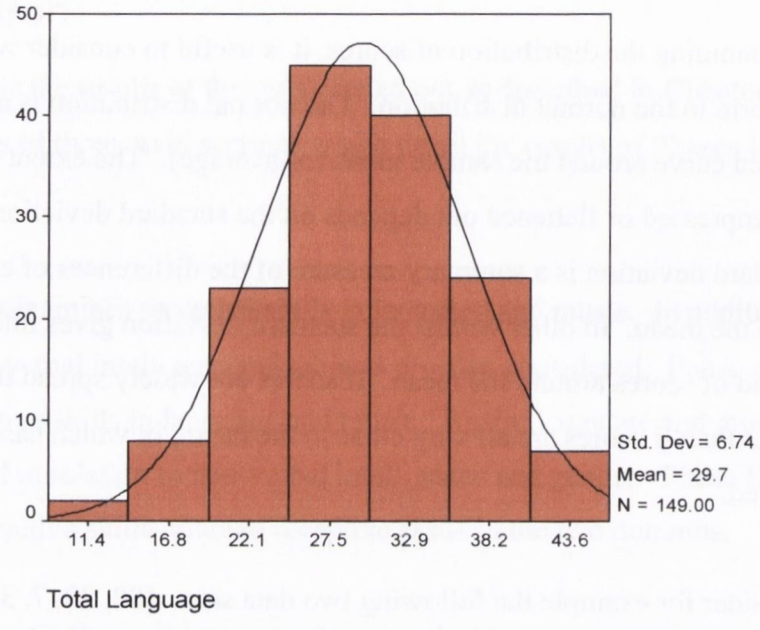


Figure 18. Distribution of Language (Recep)

The distribution of Raven scores is not normal, see Figure 19. Many more subjects received low scores than would be expected by the normal distribution. This means that overall non-verbal intelligence in this sample is below that of the general population, as distribution in the general population would be normal.

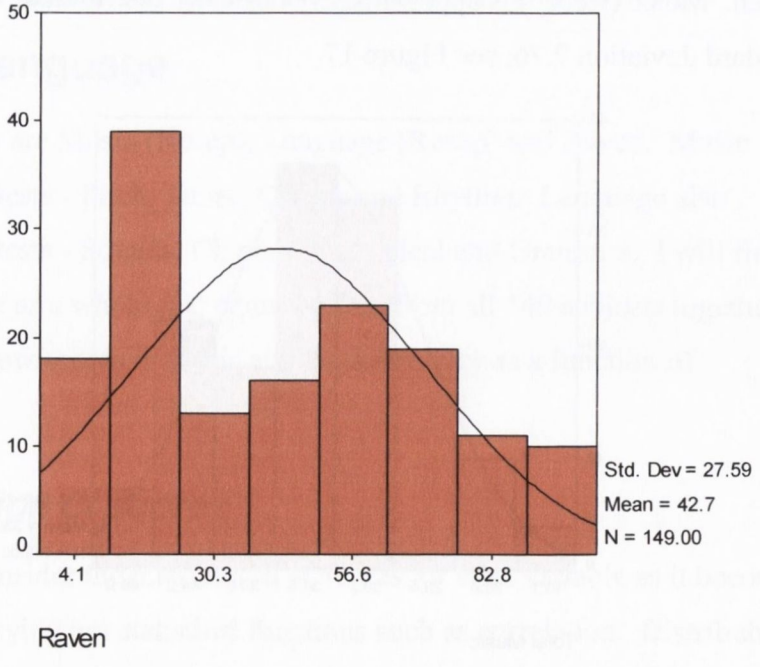


Figure 19. Distribution of Raven scores

Raven *et al.* (1998, p.61) provide a method of classifying scores on the Raven test. A score which lies at or above the 95th percentile is deemed intellectually superior (Grade 1). A score at or above the 75th percentile is "definitely above the average in intellectual capacity" (Grade 2). A score between the 25th and 75th percentiles is "intellectually average" (Grade 3). A score below the 25th percentile is "definitely below average in intellectual capacity" (Grade 4) and a score below the 5th percentile is "intellectually impaired" (Grade 5). The bar chart in Figure 20 confirms that the subjects in this sample indeed received low scores on the Raven test, i.e. it shows that the majority of subjects in this sample scores average or below average in Raven.

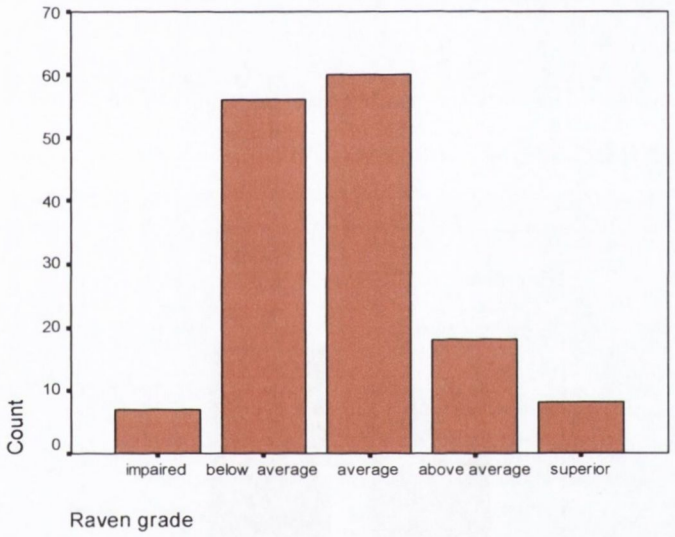


Figure 20. Distribution of Raven grades

This appears to be particularly influenced by one school (n=15) where no subjects scored Grade 1 (Superior) or Grade 2 (Above average), see Figure 21.

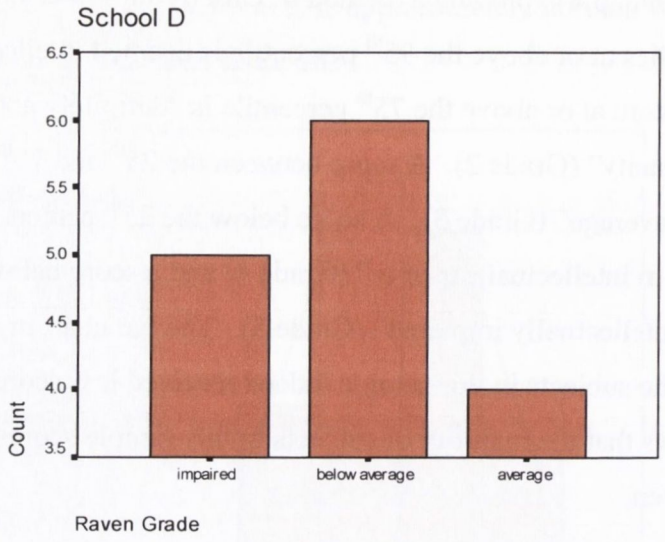


Figure 21. Distribution of Raven Grades in School D

In another school (n=18), few subjects scored Grade 2 (Above average) and no subjects scored Grade 1 (Superior), see Figure 22.

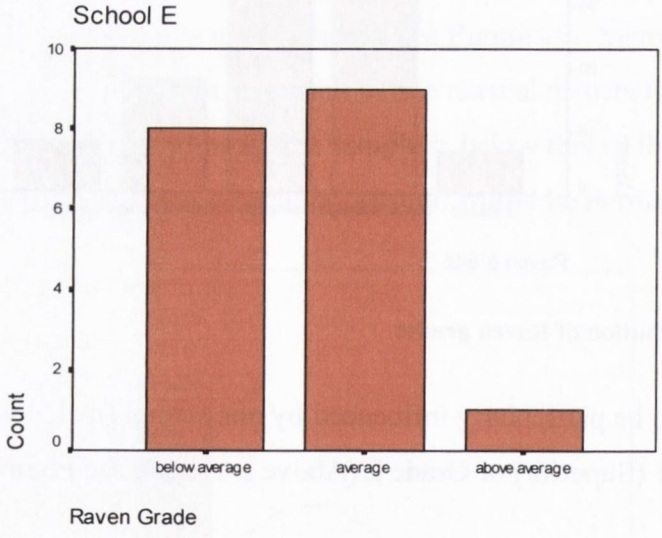


Figure 22. Distribution of Raven Grades in School E

The low Raven scores will be kept in mind throughout discussion of the results. Having noted the distribution of the variables, it is now useful to examine to correlations which exist between them. This is done in the following section.

5.2.2 Correlations

Pearson's correlation (r) is a measure of the linear association between two variables. A perfect positive correlation (r=+1) between two variables, e.g. height and weight, means that as one variable increases the other also increases, i.e. the taller an individual is, the greater his/her weight is. A perfect negative correlation (r=-1) means that as one variable increases, the other decreases; e.g. as length of day increases, hours of darkness decrease.

In general it is rare to see perfect correlations of +1 or -1. Rather, positive correlations vary in strength between 0 and +1 and negative correlations between 0 and -1. When a correlation is 0, no linear relationship exists between the variables. The stronger the correlation is, the nearer it is to +1 or -1. Simply because no linear relationship holds between two variables does not mean that they are unrelated; e.g. a brief consideration of the relationship between age and frailty reveals that very young people and very old people are quite frail, whereas middle aged people are not frail. There is a relationship between these variables, simply not a linear association. Rather the data follow a U-shaped curve.

In cases where a correlation is not -1, 0, or +1, it can be difficult to know whether or not it is large enough to be considered important. Cohen (1988) addresses this problem and suggests that a correlation of 0.5 is large, 0.3 is moderate, and 0.1 is small. Hopkins (2000) further develops Cohen's scale by adding further distinctions within the large category. He proposes that a correlation between 0 and 0.1 is trivial, 0.1 is small, 0.3 is moderate, 0.5 is large, 0.7 is very large, 0.9 is nearly perfect and 1 is perfect. Hopkins suggests that suitable synonyms for small are low or minor, while moderate is equivalent to medium. A large correlation may also be described as high or major. These terms will be used throughout this chapter to describe the magnitude of the correlations found.

Pearson's correlation assumes the normality of distribution of each of the variables used in the correlation and also bivariate normality. Bivariate normality means that both the x and y value are normal together; e.g. x could be normal and y not, or vice-versa. It was noted above that the Raven scores are not normally distributed here,

which suggests that Spearman's rho might be a more suitable statistical procedure to use when Raven is involved.

Spearman's rho is a nonparametric measure of correlation which requires the values of each of the variables to be ranked from smallest to largest. The Pearson correlation coefficient is then computed on the ranks. As the actual data are not used in the correlation, information is lost. Thus, as in this case, results are not substantially different for the Pearson correlation and for Spearman's rho, and as the Pearson correlation includes more information, the Pearson correlation is presented here. Values of Spearman's rho are presented following the values of the Pearson correlation in the appendices so that the interested reader may confirm that differences between the two are small.

A correlation is usually presented with a p-value. Before defining p-value it is necessary to understand the logic used in hypothesis testing. In the hypothesis testing framework, the null hypothesis (H_0) is assumed to be true. The null hypothesis is a statement of 'no effect'; e.g. that there is no association between the two variables entered into the correlation. In research, the aim is generally to demonstrate that the null hypothesis does not seem to hold. The alternative hypothesis is what is accepted if H_0 is rejected; e.g. that a correlation exists between two variables. The alternative hypothesis is generally the research question.

Having assumed H_0 to be true, it is necessary to evaluate the amount of support which the data lend to it. If there is not a lot of support, the null hypothesis can be logically rejected. If the null hypothesis is rejected, the alternative hypothesis may be accepted. The conclusion is that on the basis of the data, H_0 is rejected and that there is a statistically significant effect.

A p-value is a formal way of measuring the amount of support given by the observed data to the null hypothesis. When the p-value is very small, very little support is given to the null hypothesis. If there is very little support for the null hypothesis it can be rejected and the alternative hypothesis may be accepted. Thus, in research, a very small p-value allows the null hypothesis to be rejected and the research hypothesis to be accepted.

Consider the following example to illustrate the process: A researcher wishes to investigate the association between music and language aptitude. The null hypothesis states that there is no association between the two. Having collected a large amount of data, he/she investigates the amount of support for the null hypothesis. If there is little support, the p-value is small (e.g. less than 0.05), and H_0 can be rejected. On rejection of H_0 the alternative can be accepted to be the case, i.e. that there is an association between the two.

P-values of less than 0.05 are generally considered small enough to allow rejection of the null hypothesis. Another cut-off point that is commonly used is $p<0.01$. This means that alpha is 5% in the case of $p<0.05$ or 1% where $p<0.01$. The value of alpha determines the probability, which a researcher is prepared to accept, of making an error; e.g. in setting alpha at 5%, it means that in 5% of cases where the null hypothesis is rejected, it is rejected erroneously. It is clear that alpha must be set to a small value as if alpha is large there is a high probability that the null hypothesis will be incorrectly rejected. When alpha is 5%, we can say with confidence that there is 95% probability of correctly rejecting the null hypothesis. When alpha is 1%, there is 99% probability of correctly rejecting the null hypothesis.

In the following discussion, p-values are rounded off to either 0.01 or 0.05 but exact p-values are given where possible in the appendices. Where p-values are less than 0.0001 many statistical software packages including SPSS 11.0 simply give .000 rather than the exact value. In these cases the exact value is not given in the appendices.

5.2.2.1 Correlations between components of tests

Music (Recep)

Each of the components (pitch, tunes, chords and rhythm) makes a substantial and statistically significant contribution to Music (Recep). The correlations below are those which hold between total music and each of the individual music subtests. Each of these correlations is considered to be large/very large using Hopkins' (2000) classification described above.

Music (Recep) and Pitch, $r=0.72$ $p<0.01$.

Music (Recep) and Tunes, $r=0.71$, $p<0.01$.

Music (Recep) and Chords, $r=0.68$, $p<0.01$.

Music (Recep) and Rhythm, $r=0.60$, $p<0.01$.

The individual subtests also show correlations with each other. This is to be expected as they are all supposed to measure some aspect of musical aptitude and it is likely that there is some degree of overlap between the different components of musical aptitude. However, insofar as the subtests measure somewhat different aspects of musical aptitude, the correlations between them should be relatively far from perfect. Here it is seen that small to moderate correlations hold between the different subtests.

Pitch and Tunes, $r=0.41$, $p<0.01$.

Pitch and Chords, $r=0.20$, $p<0.05$.

Pitch and Rhythm, $r=0.23$, $p<0.01$.

Tunes and Chords, $r=0.29$, $p<0.01$.

Tunes and Rhythm, $r=0.35$, $p<0.01$.

Chords and Rhythm, $r=0.25$, $p<0.01$.

It is noteworthy however that Bentley's examination of the test battery did not reveal overlap between all the subcomponents. He found that pitch showed a significant ($p<0.01$) correlation with tonal memory and chord analysis but not with rhythmic memory (Bentley, 1966b, p.90). Similarly tonal memory correlated significantly ($p<0.05$) with chord analysis but not with rhythmic memory. Finally chord analysis, which correlated significantly with pitch and tonal memory, did not show a significant correlation with rhythmic memory. Thus, in Bentley's examination of the subcomponents, rhythmic ability seemed relatively separate from the other musical skills whereas here rhythm correlates positively with each of the other musical variables.

Language (Recep)

As with the music test above, all components in the language test make a significant contribution to the Language (Recep) score.

Language (Recep) and Sound, $r=0.65$, $p<0.01$.
Language (Recep) and Chinese, $r=0.32$, $p<0.01$.
Language (Recep) and Numerical, $r=0.70$, $p<0.01$.
Language (Recep) and Grammar, $r=0.82$, $p<0.01$.

While Sound, Numerical and Grammar make large to very large contributions to Language (Recep), the contribution of Chinese is much smaller (moderate according to Hopkins) albeit significant. It is perhaps not surprising that the ability to discriminate lexical tones is rather different from the abilities required in the other subtests, e.g. memory for phonemes and numbers and language analytic abilities.

Other researchers have shown English speakers to be poor at discriminating lexical tones, particularly when they are perceived as speech. According to Best's Perceptual Assimilation model, two syllables differing only on lexical tone would fall into the same category for English speakers. However, if the sounds were perceived as non-speech, English speakers could be quite good at discriminating them (Best, 1995). This hypothesis was investigated and English speakers were shown to discriminate tonal contrasts significantly better in a musical context than in a speech context (Burnham *et al.*, 1996). In a speech context, they performed better in filtered speech than in full speech. Thus, it is seen that English speakers can discriminate tones, but not very successfully in a speech context. This may explain the small correlation between the Chinese subtest and the overall language test here.

Whereas in the music test, all components correlated significantly with each other, this is not the case in the language test. With respect to correlations between language test components, Sound correlates significantly with Numerical and Grammar. Chinese correlates significantly with only Numerical. Numerical correlates significantly with Grammar.

Sound and Numerical, $r=0.31$, $p<0.01$.
Sound and Grammar, $r=0.33$, $p<0.01$.
Numerical and Grammar, $r=0.37$, $p<0.01$.

Numerical and Chinese, r= 0.22, p<0.01.
Chinese and Grammar, ns.
Chinese and Sound, ns.

The explanation above of Best's Perceptual Assimilation model may also account for these findings, i.e. while Sound, Numerical and Grammar draw on similar language abilities, the ability to discriminate lexical tones is quite distinct. It is an ability which is not often used by native speakers of English. The small correlation between Numerical and Chinese may be accounted for through recourse to memory – the numbers needed to be recalled over a period of time. Similarly the Chinese tones required memorization over a short period.

While the test components do not display large correlations with each other, the correlations between Sound, Numerical and Grammar are of approximately the same magnitude as those found between components of the music test. Thus, it can be concluded that it is likely that these three components of the language test measure somewhat different aspects of language aptitude but that they measure language aptitude nonetheless. Chinese, on the other hand, may be measuring a somewhat different ability as it examines subjects' ability to discriminate lexical tones. This might be important if the subject was going to learn a tonal language but is likely not to be of great significance if the subject is to learn an atonal language.

5.2.2.2 Correlations between tests

Music and Language Components

An analysis of all the subcomponents of the music and language tests (Pitch, Tunes, Chords, Rhythm, Sound, Chinese, Numerical and Grammar), reveals small to moderate significant correlations between many of these components. Correlations range from small, e.g. Tunes and Chinese (r=0.17, p<0.05), to moderate, e.g. Rhythm and Sounds (r=0.31, p<0.01) and Rhythm and Grammar (r=0.44, p<0.01). These correlations are of secondary importance to this study as the main aim is to consider language and music abilities overall. It might be of interest in the future to

consider in more detail the relationship between rhythm and grammar in order to cast further light on the association between the two.

Test components and Main Variables

It is not only the components of the language test which correlate with total language scores. Components of the music test also correlate with Language (Recep). Similarly components of the language test also correlate with Music (Recep). Of these correlations, one of the largest is between Language (Recep) and Rhythm, $r=0.47$, $p<0.01$. Indeed this correlation between Rhythm and overall language emerges as one of the largest in the test battery. A 95% Confidence Interval for the correlation between Rhythm and Language (Recep) is [0.34, 0.59].

A 95% Confidence Interval (CI) for a correlation gives the range within which the true population correlation should fall. The correlations calculated here are based on the sample of 149 subjects. However, it is impossible to be sure that identical correlations will hold in the population at large. Thus, confidence intervals are employed which give the range within which the correlation should fall for the population from which the sample is drawn. 95% of (95%) CIs contain the true value of the correlation for the population. Hence, it is the case that there is a 95% chance that the true value is included in the interval (see Swinscow, 1997).

Main Variables

The main aim of this phase of the study is to investigate the correlations between Music (Recep), Language (Recep) and Raven.

Music (Recep) and Language (Recep) show a moderate but significant correlation ($r=0.37$, $p<0.01$). A 95% Confidence Interval for this correlation is [0.22, 0.50]. Music and Language also correlate moderately with Raven - Music (Recep) and Raven ($r=0.34$, $p<0.01$), Language (Recep) and Raven ($r=0.40$, $p<0.01$). Ninety-five percent Confidence Intervals for these correlations are as follows:

Raven and Music (Recep): [0.19, 0.47].

Raven and Language (Recep): [0.25, 0.52].

As the correlation between music and language (r=0.37) is approximately the same as that between language and Raven (r=0.40) and that between music and Raven (r=0.34), it could be suggested that it is in fact non-verbal intelligence (as measured by Raven) which determines aptitude in music and language. Thus, schematically the idea is as shown in Figure 23 below.

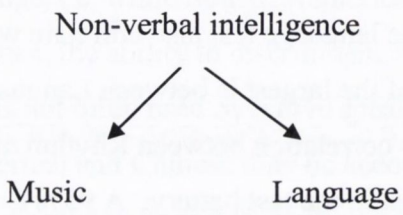


Figure 23. Association between music, language and intelligence

As noted in *Section 2.7.3*, non-verbal intelligence has been implicated in Working Memory ability. While findings in this area are not yet conclusive, it is likely that gF (general fluid or non-verbal intelligence) is related to WM capacity.

Working memory is known to be an important component of language aptitude. This was discussed in *Section 2.7.2.1* where it was mentioned that the ability to repeat non-words is a good predictor of later language learning success.

Given the connection between gF and Working Memory, and that between Working Memory and language aptitude, it is not surprising that there should be a link between gF and language aptitude. As WM is also likely to be an important component in musical aptitude, it is again unsurprising that a relationship should hold between gF and musical aptitude. The question which remains is whether or not a special relationship exists between musical and linguistic aptitude, over and above that accounted for by gF intelligence or Working Memory ability. Figure 24 attempts to examine the links between Working Memory, musical aptitude, musical training, language aptitude and language learning.

a ⟶ b Relates to links which have been extensively studied in the literature and are now widely accepted to exist. Where the lines contain arrowheads, this refers to direction of causality; e.g. 'a' causes 'b'.

- - - - - - - - - → Relates to links which have not been studied so extensively but have been alluded to. No clear consensus about their existence.

a —— b Where there is no arrowhead, the direction of causality is not known; e.g. in the case of evidence coming from a correlational study, direction of causation cannot be given.

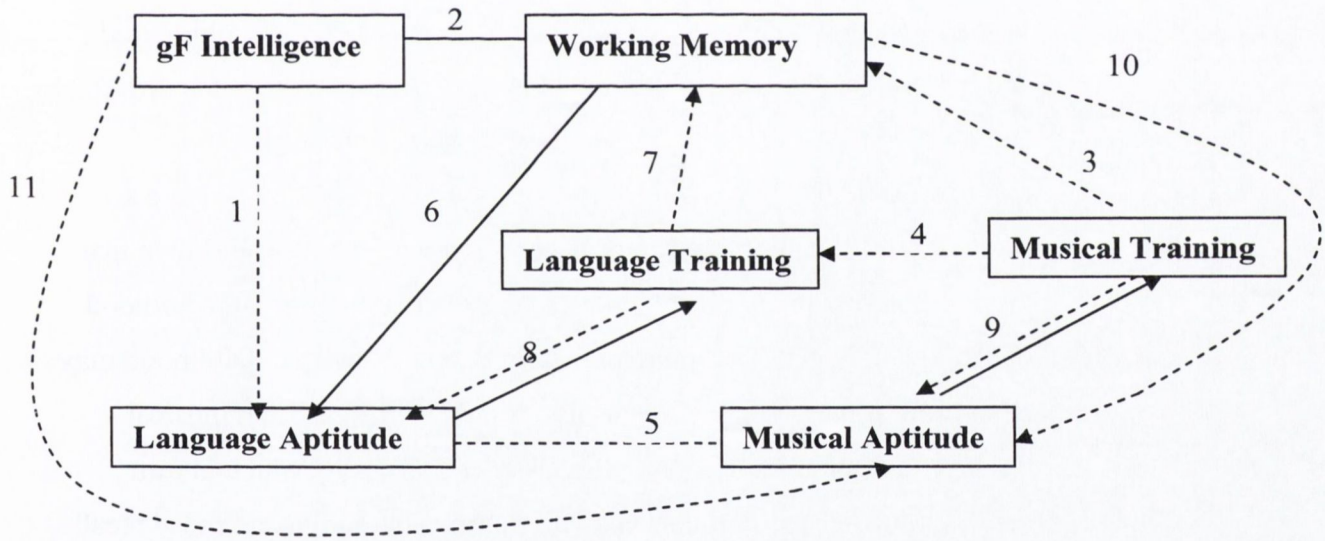


Figure 24. Connections between music, language and memory

The numbered links in Figure 24 are expanded upon below. For each link, an attempt is made to refer to literature which deals with this particular topic.

1. gF Intelligence --> Language Aptitude: Intelligence was originally considered to be a component of aptitude by Pimsleur (1962) (see *Section 2.5.2*). If this were the case, it is likely that verbal intelligence would be the primary correlate of language aptitude. However, it was suggested in *Section 2.6.2* above that verbal intelligence may be derived from fluid intelligence, i.e. it may depend on fluid intelligence. Therefore, a link could also be posited between gF intelligence and language aptitude. Recent research suggests that while fluid intelligence and language aptitude may be related, they are not completely synonymous (see *Section 2.5.6*). Fluid intelligence was found to be correlated with language aptitude in Phase I of this study ($r=0.40$, $p<0.01$).
2. gF Intelligence __Working Memory: Fluid intelligence has been linked to Working Memory ability by numerous researchers (see e.g. Colom *et al.*, 2004; R.W. Engle *et al.*, 1999b; Kyllonen & Christal, 1990). It is difficult to know if good WM ability causes high gF scores or vice-versa.
3. Musical Training --> Working Memory: Musical training may have a positive effect on Working Memory capacity; e.g. it has been shown that musical training in childhood leads to better verbal memory in adulthood (Chan *et al.*, 1998). The improvement in verbal memory in adulthood may be a result of increased WM capacity. It may be the case that musical training strengthens auditory temporal-order processing which in turn mediates the relationship between years of music training and prose recall (Jakobson *et al.*, 2003).
4. Musical Training --> Language Training: Musical training has been shown to have positive effects on reading ability, with children who undergo music lessons making faster progress than those who do not (Douglas & Willatts, 1994). For this reason, a connecting arrow is placed between musical training and language training. An alternative possibility is that this direct link should not exist but rather that musical training enhances Working Memory ability which in turn is exploited in language learning. This alternative can be seen in the graph by following links 3, 6 and 8.
5. Language Aptitude --- Musical Aptitude: Numerous studies have posited a link between musical and linguistic aptitude (e.g. Koren, 1993; Tucker,

2000). This study shows moderate to large correlations between the two in both Phase I ($r=0.37$, $p<0.01$) and Phase II ($r=0.55$, $p<0.01$), see *Section 5.3* below for details of Phase II.

6. Working memory → Language Aptitude: Working memory is now believed to be an important component of language aptitude (e.g. N. C. Ellis, 2001; Service, 1992).
7. Language Training --> Working Memory: Ellen Bialystok presented evidence from numerous studies at Eurosla 2004 showing that executive processes are boosted by bilingualism (Bialystok, 2004). Executive processes include Working Memory abilities, inhibitory controls and planning functions. Her work shows that bilingual children are more advanced in their use of executive processes than their monolingual counterparts (e.g. Bialystok & Martin, 2004). She also notes that adult bilinguals show slower decline in cognitive functioning than monolingual adults (Bialystok *et al.*, 2004). This suggests that language experience is helpful for efficient Working Memory functioning.
8. Language Aptitude → Language Training: Language aptitude is likely to predict the rate of progress which a learner makes in a language learning situation (Carroll, 1962). There is also an arrow in the other direction, i.e. from training to aptitude, as some researchers suggest that aptitude is modifiable by language learning experience (e.g. Grigorenko *et al.*, 2000; McLaughlin, 1990) (see *Section 2.5.1* above).
9. Musical Aptitude → Musical Training: Musical aptitude is likely to affect the outcome of musical training, i.e. it is expected that those with higher aptitude will derive greater benefit from training in music (e.g. Bentley, 1966b; Lehman, 1968; Wing, 1968). As long as the possibility exists that language aptitude may be modifiable by training, this possibility must also be included for the musical domain.
10. Working Memory --> Musical Aptitude: It seems likely that WM might also be a component of musical aptitude as music seems to require memory for sounds and rhythm in the same way as language processing.
11. gF Intelligence --> Musical Aptitude: gF may be related to musical aptitude. In Phase I, a moderate correlation of $r=0.34$, $p<0.01$ is found between the two.

From this graph then, it can be seen that one way to account for link 5, i.e. the association between musical aptitude and language aptitude is through the influence of WM on both (i.e. through links 6 and 10). If WM is adequately domain-general to influence both music and language aptitude, then music and language aptitude have in common at least their reliance on the same WM ability. This explanation is dependent however on the domain-generality of at least some Working Memory functioning, i.e. that it is general enough to influence both language and music and perhaps even more general. That some WM functioning is domain-general seems plausible given the discussion in Chapter 2 which points to a domain-general central executive and domain-specific slave systems.

However, if WM is so general that every cognitive task draws on it, and music and language share nothing more than a reliance on WM ability, they would show no greater relationship than that held between any other two cognitive abilities. Given the many studies which point to a particular relationship between music and language, it appears that they may have other features in common, over and above the common reliance on WM ability.

Perhaps in addition to a reliance on the domain-general central executive, music and language both draw on some of the same domain-specific skills, e.g. processing of sounds and sequencing ability. If this should be the case, it would account for a greater relationship between music and language than could be accounted for by the domain-general central executive. It would also mean that there should be a greater relationship between music and language than between other cognitive abilities which share no domain-specific abilities.

That music and language have a special relationship over and above that held between other cognitive abilities is difficult to show; e.g. music has also been shown to have an effect on spatial reasoning (see *Section 3.2* above). Clearly music and spatial reasoning may both require domain-general central executive functioning. However, it is more difficult to pinpoint the domain-specific skills which could be shared, but perhaps both involve sequencing abilities. Given this difficulty, this thesis must simply focus on investigating the relationship between music and language aptitude and quantifying the magnitude of that relationship in

the current sample. It is impossible to say definitively that this relationship has any special status until it is contrasted with the relationship between other cognitive abilities.

Another way to account for the connection between music and language or for the connection between any cognitive abilities is though recourse to general intelligence. It is noted in *Section 2.6.2* above that the 'g' factor of general intelligence now receives wide-spread support. One element of 'g' is gF, general fluid intelligence. gF can be seen to influence both language and music aptitude in Figure 24 through links 11 and 1. If music and language are even in part determined by gF intelligence, then gF intelligence accounts for some of their common variance. Again, as in the case of Working Memory, it is likely that music and language may share some processes over and above the influence of gF on both.

Yet another possibility is to look at the influence of musical aptitude on musical training (link 9) which in turn enhances Working Memory ability. Enhanced WM ability improves language aptitude. In other words, starting at musical aptitude and passing through links 9, 3 and 6 leads to language aptitude. However, this assumes that musical aptitude in itself has no direct link with linguistic aptitude and that it is the musical training which causes the effect on WM, basically arriving back at links 6 and 10. While training in music has been shown to be beneficial (see link 4), the subjects in the current study had certainly not all received musical training. Yet a significant link exists between music and language aptitude. Is this simply due to the importance of WM or gF to both?

In order to try to tease apart the contribution of WM to language and music, let us divide the subjects into those with above and below average Raven scores. As already stated, Working Memory is related to gF intelligence although it is not identical to it. Raven is a measure of gF intelligence and so may also be an indirect indicator of WM ability. This division is done in the following section.

5.2.3 Subjects divided by Raven scores

It is worth considering how higher and lower non-verbal intelligence affects musical and linguistic aptitude. If indeed music and language aptitude are primarily dependent on WM ability, it would be expected that high WM leads to high scores on the music and language aptitude tests. It would be unlikely to find low WM ability and high scores in one of music or language. The aim of this section is to consider this issue.

The mean score on the Raven test was 42.73 (standard deviation 27.59). While subjects could be divided into groups based on this score, i.e. one group where subjects scored equal to or above the mean and another where subjects scored below the mean, it was noted earlier that Raven scores are not normally distributed in this sample with more low scores than expected in the general population. Therefore, the mean Raven score in this sample is quite low.

It might therefore be preferable to use the categories set out above in *Section 5.2.1*, i.e. Superior, Above Average, Average, Below Average and Impaired, as these are categories which have been previously used in the literature. If the sample is divided into these five categories there are unequal numbers in the different groups, with only 7 subjects classified as impaired and 8 classified as superior. There are 56 in the below average group, 60 average and 18 above average subjects. Rather than have large differences between the groups, it seems better to use two groups with approximately equal numbers, i.e. average and above (AAS) versus below average and impaired (BAI). This results in 86 subjects classified as average or above and 63 as below average or impaired. Thus, grouping is done as discussed in the first paragraph of this section but using categories from the literature rather than simply the mean of this sample.

Let us now consider whether or not these two groups differ in their music and language scores. Boxplots are a useful way to do this. A boxplot is a summary plot based on the median, quartiles, and extreme values found in the data. A number of boxplots were found in Chapter 4 when looking at scores awarded by native speaker raters to subjects' productions.

The median is the value in the centre of the data, i.e. the value above and below which half the cases fall. If there is an even number of cases, the median is the average of the two middle cases when they are sorted in ascending or descending order. The median is a measure of central tendency not sensitive to outlying values - unlike the mean, which can be affected by a few extremely high or low values. Quartiles divide the observations into four groups of equal size. 25% of values fall below the first quartile (Q1) and 75% above. 75% of values fall below the 3rd quartile (Q3) and 25% above. The median is the 2nd quartile as 50% of values lie above the median and 50% below. The interquartile range is the difference between Q3 and Q1, i.e. Q3-Q1.

In a boxplot, the shaded box represents the interquartile range which contains 50% of values. Given the fact that 25% of values fall below the first quartile and 25% of values lie above the 3rd quartile, the region between the two, i.e. the interquartile range, must contain the remaining 50% of the values. The whiskers are lines that extend from the box to the highest and lowest values, excluding outliers. The line across the box indicates the median.

If there is an outlier in the data this is illustrated by a dot above or below the maximum or minimum value. There are no outliers in Figures 25 or 26 but one is shown in Figure 28. A point is considered an outlier, rather than the minimum or maximum value, if it is more than 1.5 times the interquartile range away from Q1 (minimum) or Q3 (maximum). This should become clear with an example.

Consider the following data set in order to illustrate median, interquartile range, minimum, maximum and outliers: {60, 73, 67, 62, 62, 70, 65}. The median is calculated by ordering the data, i.e. {60, 62, 62, 65, 67, 70, 73}, and selecting the middle value, i.e. 65.

The first quartile (Q1) is the median of all values below the median, in this case the median of {60, 62, 62}, i.e. 62. The third quartile (Q3) is the median of all values above the median, i.e. the median of {67, 70, 73}, which is 70. The interquartile range (Q3-Q1) is 70 – 62, i.e. 8.

The minimum value is 60 and the maximum value is 73. Neither of these values is an outlier because neither is more than 1.5 times the interquartile range away from the nearest quartile, i.e. 1.5 times the interquartile range here is 1.5 * 8 = 12. The maximum value possible here which is not considered an outlier is as follows:

Q3	+	(1.5 * interquartile range)	=	
70	+	(1.5 * 8)	=	
70	+	12	=	82

As 73 is the maximum value in this data set and 73 is less than 82, it is therefore taken as the maximum and not an outlier.

Suppose 86 is then added to the data set so that it becomes {60, 62, 62, 65, 67, 70, 73, 86}. The median is now 66, Q1 is 62, Q3 is 71.5 and the interquartile range is 9.5. Multiplying 1.5 by the interquartile range gives 14.25, as follows: 1.5 * 9.5 = 14.25.

Adding this to Q3 gives the maximum possible value that will not be considered an outlier, i.e. Q3 + (1.5*interquartile range) = 71.5 + 14.25 = 85.75. Therefore, the maximum possible value which could be present in this dataset and not be considered an outlier is 85.75.

However, the maximum value in the data set is 86, but 86 > 85.75. 86 is therefore classified as an outlier and the maximum is deemed to be 73, i.e. the next largest value in the dataset.

A boxplot is useful for showing the distribution of the data as it shows the box in which 50% of the values lie, in addition to how far away from the centre the most extreme values occur. If the whiskers are far away from the box, this means that the most extreme values are far away from the median so data is quite spread. If whiskers are short, this means that the most extreme values are not very far away from the centre of the data, so values are less spread. A narrow box shows that the 50% of values contained within it are close to the median. A wide box shows that even within the 50% of values contained in it, there is a large spread of values.

Figure 25 shows that the median score for Language (Recep) is slightly higher in the Above Average/Superior (AAS) group. Both boxes are quite narrow which shows that the 50% of values contained within the boxes are not widely spread. The whiskers are quite long which shows that the minimum and maximum values are quite far away from the median. This demonstrates a good spread of scores, with some subjects receiving low scores and others receiving high scores.

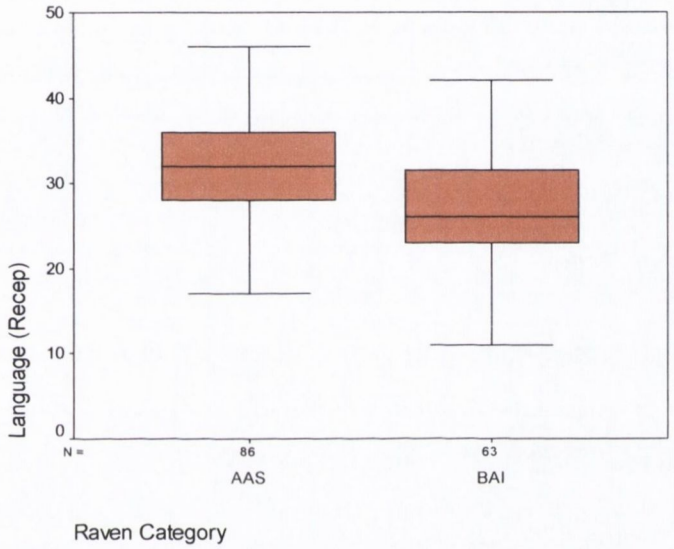


Figure 25. Language (Recep) scores for different Raven categories

A t-test is a formal way to test if there is a significant difference between the mean scores of two groups. Looking at Language (Recep), a significant difference is found between the two groups (equal variance not assumed, $t=4.67$, $df=120.45$, $p<0.01$). Details of these t-tests are presented in Appendix 1. A 95% Confidence Interval for the difference is [2.87, 7.11]. If zero is contained in the confidence interval, it means that the difference between the two groups may be equal to zero. Here, zero is not contained in the interval, so the difference between the groups is not equal to zero.

Figure 26 shows that the median for Music (Recep) is also higher in the AAS group.

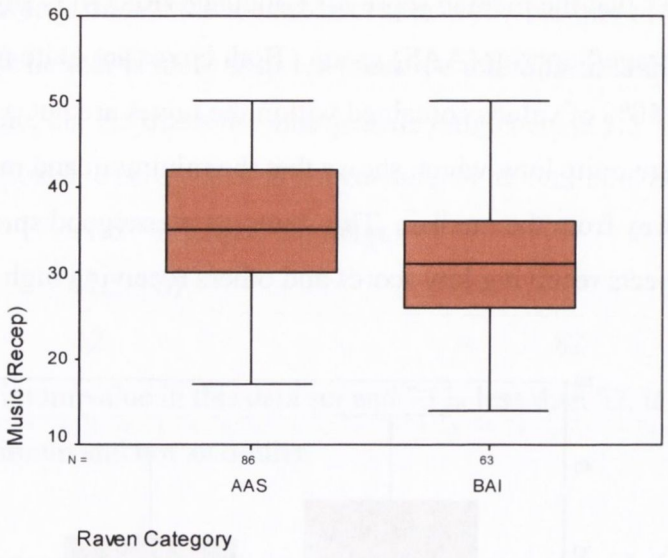


Figure 26. Music (Recep) scores for different Raven categories

Looking at Music (Recep), a significant difference is also found between the two groups (equal variance not assumed, $t=3.50$, $df=132.42$, $p<0.01$). A 95% Confidence Interval for the difference is [1.89, 6.81].

95% Confidence Intervals for the differences in mean scores between high Raven and Low Raven Groups are as follows:

Music (Recep): [1.89, 6.81].

Language (Recep): [2.87, 7.11].

Thus, whether or not the difference is 'caused' by Raven, slightly higher scores in music and language go hand in hand with higher Raven scores. Perhaps greater Raven scores reflect better Working Memory skills, as it has been shown that Raven is related to WM (see *Section 2.7.3*). Higher WM skills, in turn, may lead to higher scores in language, music and other cognitive skills. Thus, suppose that WM and Raven are placed in the dark shaded area in Figure 27 on the grounds that they influence most cognitive skills and some level of WM ability is essential in order to do any task. Placing WM and gF intelligence in the centre allows us to account for the link between music and language (link 5 in Figure 24) through the common influence of WM and gF on both, i.e. links 6, 10, 1, and 11 in Figure 24.

The question of whether or not language and music share some commonalities over and above the influence of WM is addressed by the positioning of auditory skills in the grey shaded area of Figure 27. Auditory abilities which are shared between music and language could be placed in the grey area on the grounds that they are common to music and language but are not required by other cognitive processes such as mathematical reasoning. This suggests that music and language may have some common processes over and above WM and gF intelligence.

Of course music and language also involve abilities which are specific to each domain; e.g. natural language semantics is much more highly developed than semantics in music.

Figure 27 allows for the existence of correlations between the more general skills required in Raven and WM tests and the more specific skills required in language and music tests as both the grey area and the darker area constitute the intersection of the music and language sets. It also allows for the possibility that even with a high level of WM ability, musical or linguistic aptitude could be low due to insufficient domain-specific skills. This would account for the fact that it is not always the case that high Raven scorers have high language aptitude.

This organisation also allows for the possibility of having different types of language learner; e.g. a learner may have high musical ability. This means that he/she is strong in A and B on Figure 27. He/she may therefore use the abilities in B in language learning. On the other hand a learner may be high in C, e.g. language analytic abilities, so may take this path into language learning.

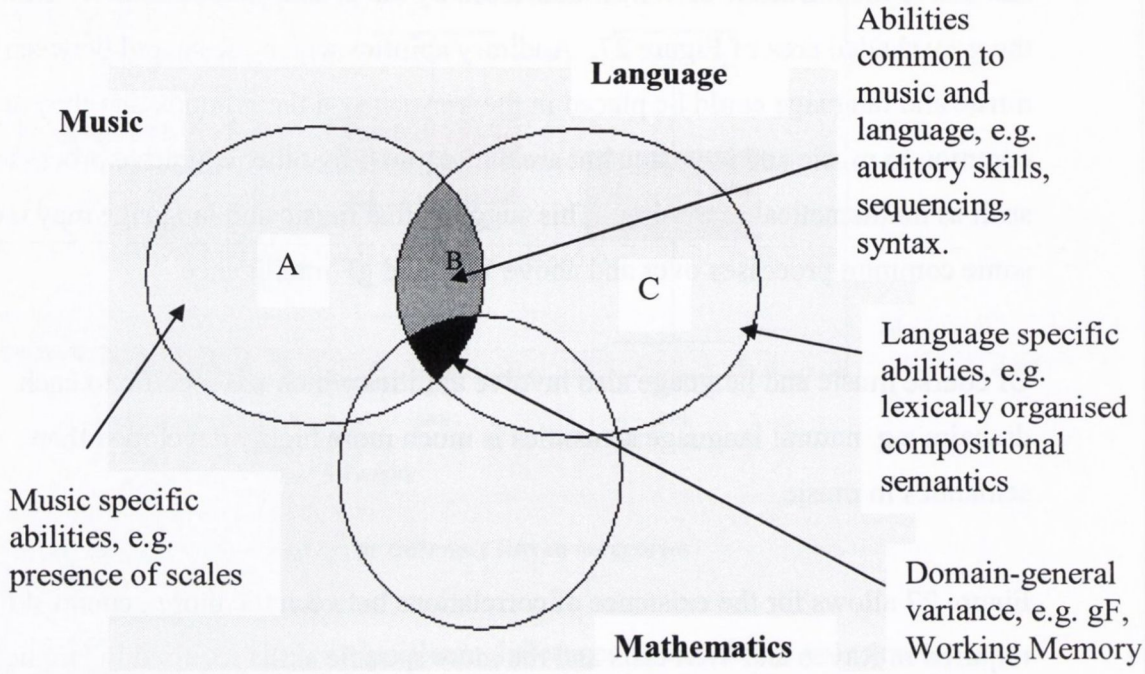


Figure 27. Cognitive processes in music and language

5.2.3.1 Correlations within high and low Raven groups

It was established in the previous section that those subjects with higher Raven scores generally receive higher music and language scores. It is also interesting to consider whether music and language abilities are more closely or less closely related in the different groups.

In fact, no significant differences emerge between the pattern of correlations in the High and Low Raven groups. It was shown above that those with above average Raven have significantly higher scores in music and language than those with below average Raven scores. Here it is shown that the overall pattern of correlations in the two groups is the same.

Music Receptive and Language Receptive

Low Raven: $r=0.44$, $p<0.01$

High Raven: $r=0.18$, ns.

The difference between the groups is not significant because a 95% CI for the difference contains 0, ([-0.55, 0.04]).

Music Receptive and Raven

Low Raven: $r=0.23$, $p<0.05$.

High Raven: $r=0.2$, ns.

Again the difference between the two is not significant. A 95% CI for the difference containing zero is as follows: [-0.37, 0.27].

Language Receptive and Raven

Low Raven: $r=0.40$, $p<0.01$.

High Raven: $r=0.13$, ns.

The difference is not significant. A 95% CI for the difference is as follows: [-0.56, 0.03]

Although none of the differences emerge as significant, it is noteworthy that correlations seem greater in the low Raven group. It is possible that if larger numbers of subjects were present in each group, these differences would in fact

emerge as significant. It is to be expected that a certain baseline level of intelligence is essential in order to complete the language and music tests. Perhaps once this minimum level is present, fewer correlations are seen between the different abilities.

This is in line with the proposal that in more intelligent individuals, abilities are more differentiated whereas in individuals of lower intelligence, a general factor is more pervasive. This is known as the Ability Differentiation Hypothesis (or Divergence Hypothesis) and was originally proposed by Spearman (1927, p.219) who notes that at high levels of 'g', abilities are not as closely associated as they are at lower levels of 'g'. Similar findings come from recent research (e.g. Deary *et al.*, 1996).

5.2.3.2 Summary and conclusion of Raven groups

- Those with average or above Raven scores have significantly higher scores in music and language. This is perhaps somewhat surprising as many of the tests involved discriminating between different sounds. It was not predicted that more intelligent subjects should necessarily be better at discriminating sounds. It had been expected that the grammar component in the language test might be influenced by intelligence but these effects were not predicted for the language and music tests in their entirety. As mentioned above, the Raven scores in this sample are quite low and perhaps a minimum level of intelligence is required to successfully complete the tests. While the above average group here receives higher scores, it would be interesting to consider a group with exceptional Raven scores. It would be very surprising indeed if exceptional Raven scores automatically led to better sound discrimination abilities.
- There are no cases of those in the low Raven group achieving language scores as high as the maximum scores achieved in the high Raven group. (This is seen in Figure 25). This points to an important role of gF intelligence in determining language aptitude scores.
- There are a number of cases of those in the low Raven group achieving music scores as high as the maximum scores achieved in the high Raven group. This suggests that high music scores may be achieved even in the

absence of high Raven scores. Above it was stated that if WM or gF intelligence was solely responsible for the link between music and language, it would be unlikely to find cases of low WM ability and high scores in either music or language. Here are examples of low WM ability and high scores in music. Of course this is based on the assumption that Raven scores are related to WM ability. Therefore, musical aptitude depends on more than simply WM ability.

- The pattern of correlations does not differ significantly between the two groups although there is a slight trend for the low Raven groups to show greater correlations.

To conclude, it seems here that music and language aptitude follow predictions made by the Ability Differentiation Hypothesis with a tendency for greater correlations between the two to exist in those individuals with lower intelligence. This section therefore does not provide substantial support for the existence of a special relationship between these two abilities. Rather they seem to act as other abilities and show greater correlations in less intelligent individuals and lower correlations in more intelligent individuals. However, it was seen that an individual may have a high score in music even in the absence of a high score in Raven. Therefore, it is seen that Raven does not completely determine musical aptitude. Hence, it is argued that as Raven may be related to Working Memory ability, Working Memory does not entirely determine musical aptitude.

In any case, the analysis in this section is an informal way of looking at the relationship between two variables. The following sections describe partial correlation and regression, which are formal statistical methods of investigating the relationship between more than two variables. Before that an analysis of the results broken down by gender is presented.

5.2.4 Subjects divided by gender

The aim of this section is to investigate whether males (n=82) and females (n=67) differ in their Music (Recep), Language (Recep) and Raven ability. Figures 28, 29 and 30 present this information graphically.

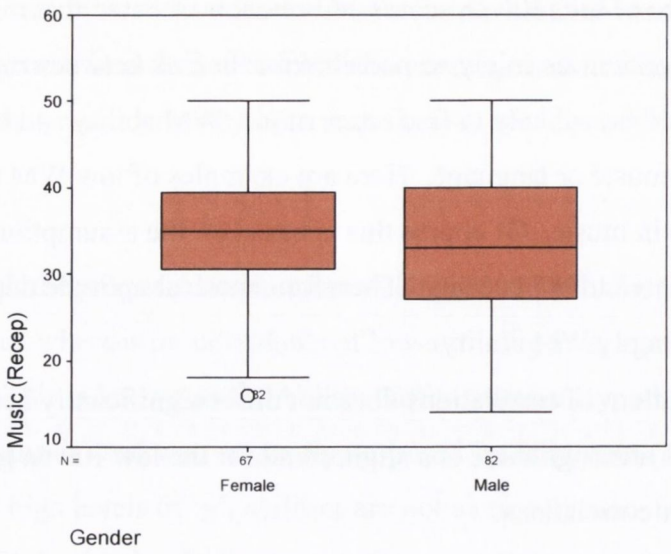


Figure 28. Music (Recep) in Males and Females

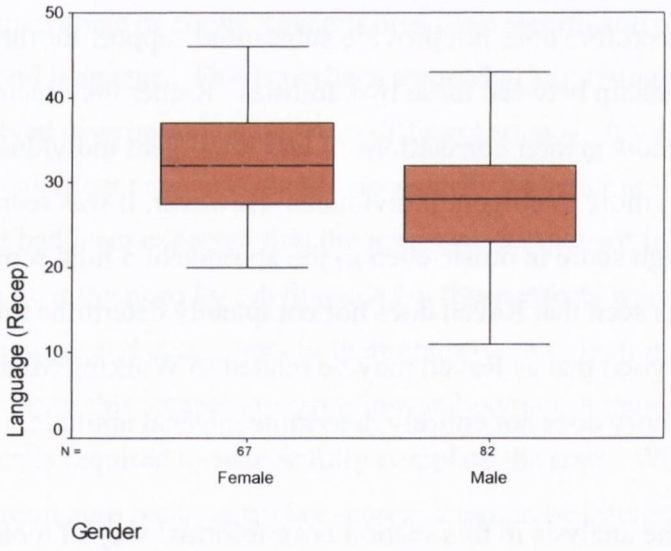


Figure 29. Language (Recep) in Males and Females

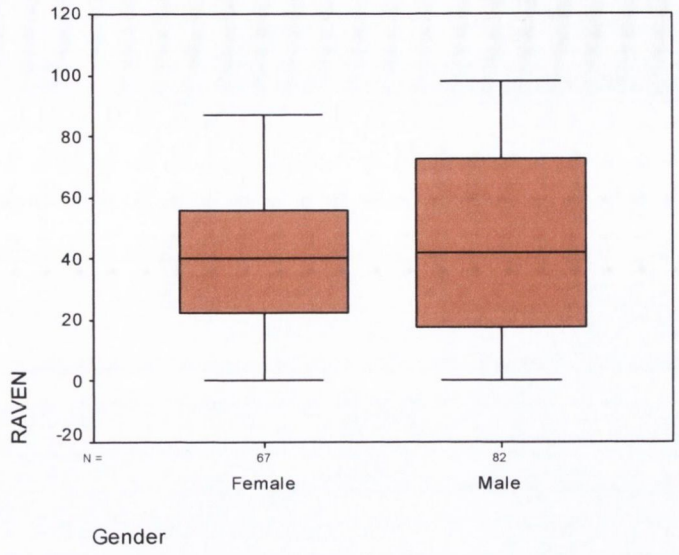


Figure 30. Raven in Males and Females

The mean value for Music (Recep) in females is 34.61 (std. dev 7.534) and in males 33.16 (std. dev 7.923). The mean value for Language (Recep) in females is 32.27 (std. dev 6.059) and in males 27.61 (std. dev 6.565). The mean value for Raven in females is 39.67 (std. dev 22.306), while in males the mean value is 45.23 (std. dev 31.171).

T-tests were carried out to investigate the significance of the differences between the groups (see *Appendix 1* for details). Results show that the only significant difference between the groups is in the case of Language (Recep), ($t=4.496$, $df=144.793$, significant $p<0.01$).

95% Confidence Intervals of the differences are as follows (equal variances not assumed):

Language (Recep): [2.61, 6.71].

Music (Recep): [-1.06, 3.96].

Raven: [-14.23, 3.12].

The only interval which does not contain zero is that for Language (Recep). Thus, the difference between the groups is significant only in the case of Language (Recep), with females scoring significantly higher than males.

5.2.5 Partial Correlation

Figure 27 above showed that the relationship between music and language may be accounted for in a number of ways. One way to account for the relationship is to say that it is all due to the darkest part of the diagram, i.e. all due to Working Memory or non-verbal intelligence. Another way is to propose that the relationship is due to the shaded part only and a third explanation is a combination of the first two. Correlational studies cannot take into account confounding variables and so cannot tell us which of these possibilities is most likely to be the case. It is therefore necessary to introduce the idea of partial correlation. A partial correlation is used to investigate whether a significant correlation holds between two variables when the effects of a third variable are held constant; e.g. in this case, is the correlation between music and language significant when the effects of Raven are held constant.

A partial correlation between two variables while accounting for a third, is calculated based on the correlations between each of the original variables, say *A*, *B* and *C*. It is necessary to know the correlation between *A* and *B*, *B* and *C*, and between *A* and *C*.

Altman (1999, p.296) describes how to calculate partial correlations as follows:

> "We can calculate the correlation between two variables after adjusting for a third if we have the correlation coefficients between each pair of variables, say r(*AB*), r(*AC*) and r(*BC*). To adjust the correlation between variables *A* and *B* for the possible effect of variable *C*, we calculate the partial correlation of *A* and *B* adjusted for *C* as:"

$$r(AB|C) = \frac{r(AB) - r(AC)r(BC)}{\sqrt{[1 - r(AC)^2][1 - r(BC)^2]}}$$

Equation 3. Calculation of partial correlation

Thus, in order to calculate the numerator, it is necessary to multiply the correlation between *A* and *C* by the correlation between *B* and *C*. This is then subtracted from the correlation between *A* and *B*.

The denominator again uses the correlation between *A* and *C* and that between *B* and *C*. These values are squared and subtracted from 1 to give two new values. Then the denominator is calculated by taking the square root of the product of these two new values.

It can therefore be investigated if the correlation between music and language is still significant even with the effects of non-verbal intelligence partialled out. Partial correlation gives answers to the following questions:

- Is the correlation between music and language still significant if adjusted for intelligence?
- Keeping Raven (i.e. gF intelligence) constant, what is the correlation between music and language?

Let *A* = Language (Recep), *B* = Music (Recep) and *C* = Raven. Therefore, the necessary values to substitute in the formula are as follows:
r(*AB*) = 0.371, r(*AC*) = 0.395, r(*BC*) = 0.337.

Substituting these values into the formula above gives[25]:
r(Language Recep, Music Recep|Raven) = 0.275, p=0.0007
95% Confidence Interval = [0.120, 0.417] .

Therefore, the correlation between Music (Recep) and Language (Recep) is still considered moderate and is highly significant when Raven is partialled out. Another way to investigate this issue is by using Regression to consider the contributions made by music and Raven individually in predicting language scores. This is done in the following section.

5.2.6 Regression

5.2.6.1 Introduction

Regression is used to estimate the linear relationship between a dependent variable and one or more independent variables or covariates (SPSS Inc., 2001). In other words, regression is used to explain the extent to which one variable increases or

[25] Calculations carried out using http://home.clara.net/sisa/correl.htm, last verified 26.04.2005

decreases with respect to another; e.g. as musical aptitude scores increase, is a corresponding increase (or decrease) found in language aptitude scores?

Regression is a very useful tool as it can tell us the amount of variation in language aptitude accounted for by musical aptitude and Raven scores. This information is given by R-squared (or R-squared adjusted). R-squared is the proportion of variation in the dependent variable (e.g. language aptitude) explained by the regression model. It ranges in value from 0 to 1, or 0% to 100%. Small values indicate that the model does not fit the data well (SPSS Inc., 2001).

Regression involves the calculation of residuals. Residuals are the difference between the observed values of the dependent variable (e.g. Language (Recep) in this study) and the values expected by the model; e.g. suppose that observed values are {20, 25, 28, 30} and values expected by the regression equation are {20, 23, 26, 29}, then residuals are {0, 2, 2, 1}. Large residuals indicate that the observed values are very different from the predicted values. In other words, large residuals mean that the model is not a good fit. Residuals are often divided by their standard deviation in order to make them more directly comparable with each other. When residuals are standardised in this way, 95% of them are expected to fall between –2 and +2. Figure 33 below shows that the majority of residuals here fall between -2 and +2.

Before conducting a regression analysis, it is useful to examine scatterplots of the variables in order to see if a linear relationship holds between them. Figure 31 presents graphs of Language (Recep) scores plotted against Music (Recep) scores and Rhythm scores.

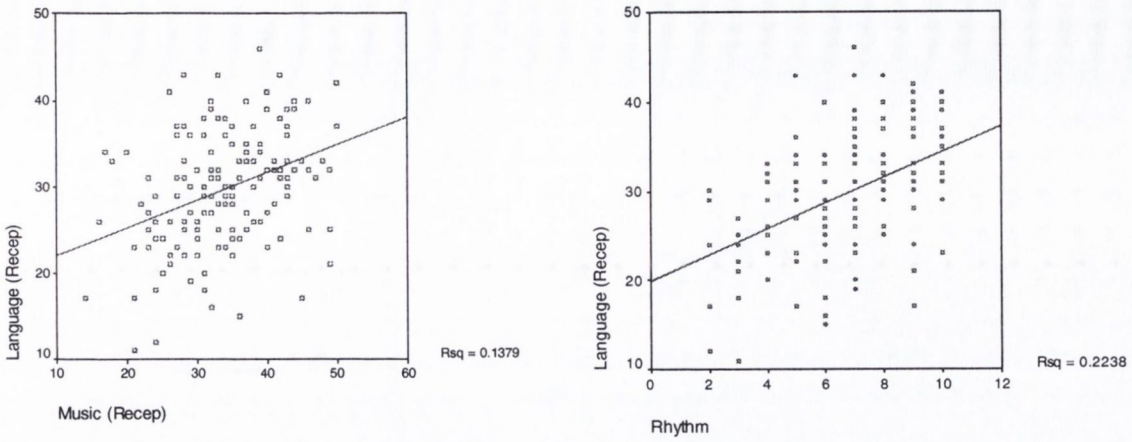


Figure 31. Music (Recep) and Language (Recep) scores

Looking at a scatterplot of Music (Recep) and Language (Recep) scores suggests that there is a trend for language scores to increase as music scores increase. R-squared is 13.79%. Thus, approximately 14% of the variance in Language (Recep) scores can be accounted for by Music (Recep) scores. The linear relationship between music and language is slightly greater when the rhythm component of music is considered separately. Here R-squared value is 22.38%. Because Rhythm is a better predictor of Language (Recep) than Total Music, the remainder of this section uses rhythm as a predictor variable.

Clearly there is a certain amount of noise in these scatterplots. It is evident that some subjects received high scores in music, yet still received low scores in language. Others received low scores in music but high scores in language. This is to be expected when dealing with human subjects and two distinct abilities. Of course music and language are not perfectly related and in the course of doing this research I was told numerous anecdotes about individuals who could speak many languages yet didn't have a musical note in their head.

Language scores also increase in response to an increase in Raven scores. Figure 32 shows this information. R-squared is 15.59%. Hence, approximately 16% of the variance in Language (Recep) scores can be accounted for by Raven scores. Rhythm scores do not particularly increase in response to Raven scores. R-squared is only 7.23%.

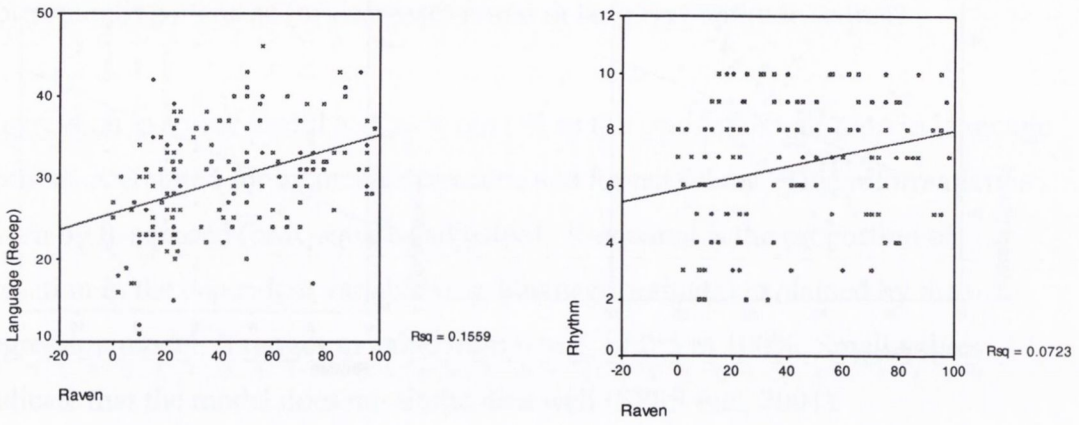


Figure 32. Music, Language and Raven scores

5.2.6.2 Multiple Regression

Multiple Regression allows more than one predictor variable to predict or explain a single response variable. In this case, Rhythm and Raven are used to predict Language (Recep). Therefore, Language (Recep) is the dependent variable and Rhythm and Raven are predictor variables. Multiple regression shows whether or not the inclusion of both variables is more useful in predicting language scores than simply either variable alone. As the two variables are only weakly correlated, it is valid to include both in the regressions analysis. The outcome of a regression analysis is a regression equation. Here the regression equation is as follows:

$$\text{Language (Recep)} = 18.48 + 1.22\ \text{Rhythm} + 0.07\ \text{Raven}.$$

The coefficients are all significant $p<0.01$. Raw data is provided in Appendix 1.

This gives a means of predicting Language (Recep) scores from Rhythm and Raven scores; e.g. suppose a student scores 5 in Rhythm and 40 in Raven, his/her Language (Recep) score will be as follows:

$$\begin{aligned}\text{Language (Recep)} &= 18.48 + 1.22(5) + 0.07(40)\\ &= 18.48 + 6.1 + 2.8\\ &= 27.38\end{aligned}$$

Suppose a student has an excellent Rhythm score (maximum possible score in Rhythm is 10) and an average Raven score (50th percentile), his/her language score is predicted to be as follows:

$$\begin{aligned} \text{Language (Recep)} &= 18.48 + 1.22(10) + 0.07(50) \\ &= 18.48 + 12.2 + 3.5 \\ &= 34.18 \end{aligned}$$

On the other hand, a student could have a good Rhythm score (e.g. 6) and an excellent Raven score (90th percentile). In this case, his/her Language (Recep) score is predicted to be the following:

$$\begin{aligned} \text{Language (Recep)} &= 18.48 + 1.22(6) + 0.07(90) \\ &= 18.48 + 7.32 + 6.3 \\ &= 32.1 \end{aligned}$$

Standardised coefficients allow more direct comparison of the contribution of Raven and Rhythm. Standardised coefficients are shown to be as follows: 0.396 Rhythm, 0.288 Raven which suggests that Rhythm makes a greater contribution than Raven.

R-squared (adj) for this model is 0.291. Thus, 29.1% of the variance in Language (Recep) scores can be accounted for by considering Raven and Rhythm together. R-squared (adj) is used as it takes into account the fact that extra variables were included in the model and penalises them if they do not add any further predictive power to the model.

In order to check that Regression is an appropriate analysis in this case, I present the following plots:

- Residuals versus the Order of the Data
- Residuals versus the Fitted Values
- Normal Probability Plot of the Residuals

Figure 33 shows that there is no particular ordering of the data. This is what is expected, i.e. there should be no relationship between the order in which the data was gathered and the size of the residuals.

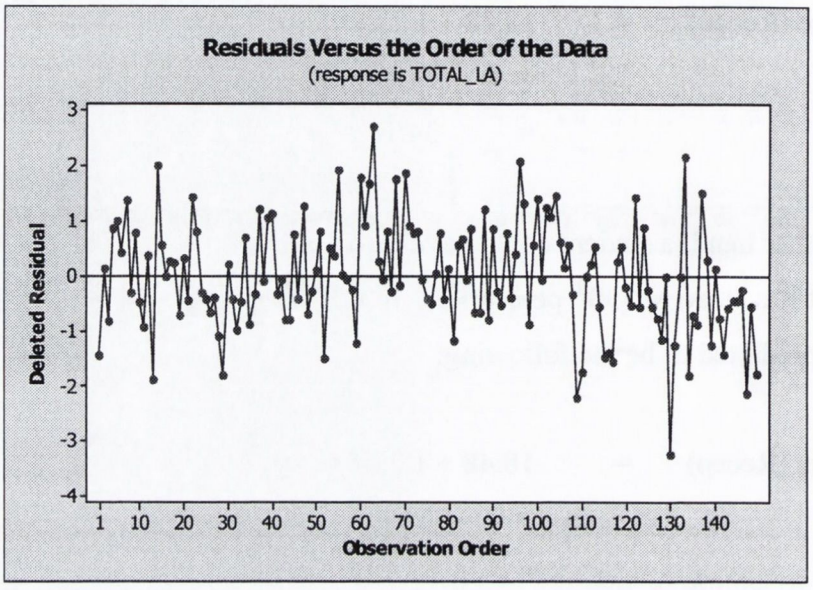


Figure 33. Residuals vs. order of data

In Figure 34, random scatter of points is expected. This is indeed what is seen.

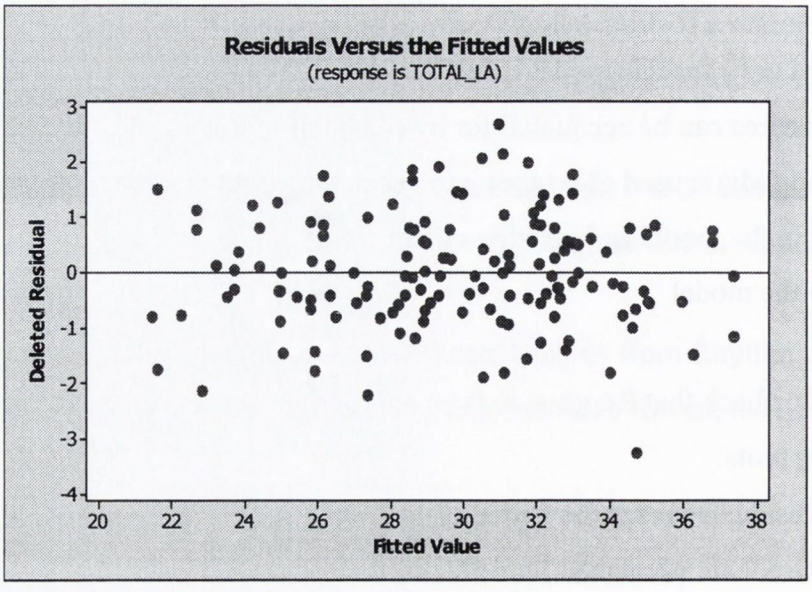


Figure 34. Residuals vs. fitted values

In the Normal Probability Plot of the Residuals, points are expected to sit on a straight line. This is basically the case in Figure 35 except for one outlier.

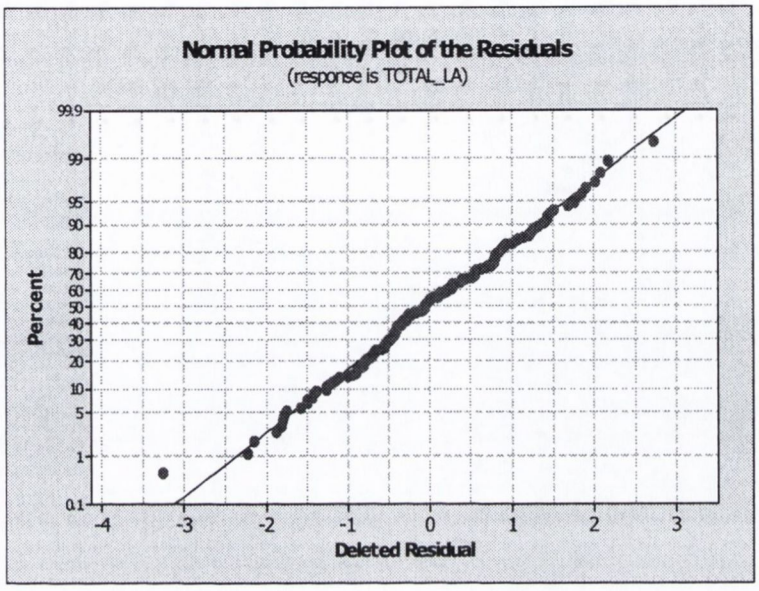


Figure 35. Normal probability plot of residuals

Figure 35 above shows that one point appeared to have a much smaller residual than expected. Hence, Cook's Distance is plotted (see Figure 36). Cook's Distance is a measure of how much the residuals of all cases would change if a particular case were excluded from the calculation of the regression coefficients. A large Cook's Distance indicates that excluding a case from computation of the regression statistics, changes the coefficients substantially (SPSS Inc., 2001).

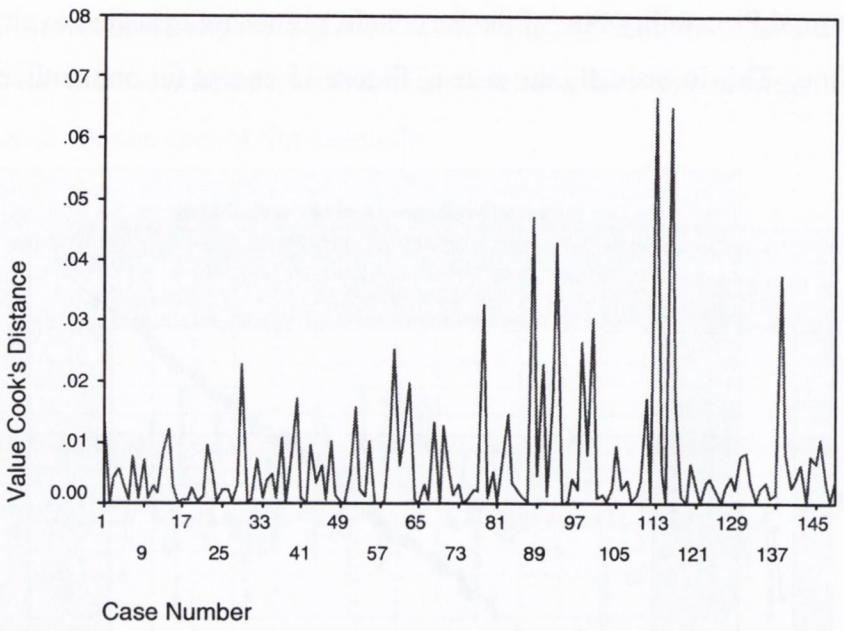


Figure 36. Cook's Distances, Phase I

Points 113 and 115 in the database are seen to have large Cook's Distances and so are taken to be outliers. These two points with the largest Cook's Distances are omitted and the analysis re-run (see *Appendix 1* for raw data).
The Regression equation is now:
Language (Recep) = 17.978 + 1.31 Rhythm + 0.07 Raven.

Standardised coefficients are 0.44 Rhythm, 0.28 Raven, both significant $p<0.01$.

R-squared (adj) is now 32.7% for this model. Thus, 32.7% of the variation in Language (Recep) scores can be accounted for by Rhythm and Raven.

As it was noted in *Section 5.2.4* above that females scored significantly higher than males in Language (Recep), gender should be added as a predictor variable into the regression equation. Gender was coded into 0 for female and 1 for male. This results in the following regression equation:
Language (Recep) = 21.27 + 1.09 Rhythm + 0.08 Raven – 4.15 gender.
All predictor variables significant, $p<0.01$. R-sq(adj) is 41.7%.

This equation shows that when gender is zero (i.e. female), -4.15 is multiplied by zero so nothing is subtracted from the value for Language (Recep) in the case of

female subjects. However, in the case of male subjects, -4.15 is multiplied by 1 with the result that -4.15 is subtracted from Language (Recep) for males.

In order to say definitively that using Raven, Rhythm and Gender better predicts Language (Recep) than using either variable alone or in combination with just one other, stepwise regression is used. In stepwise regression, each variable is entered one at a time into the model. At each stage it can be seen if the addition of another variable improves or reduces the predictive power of the model. Table 4 presents the first model which includes only Rhythm as a predictor of Language (Recep) scores. R-sq (adj) is 25.8%. The second model adds Raven as an additional predictor. R-squared (adj) then increases to 32.7%. Addition of gender causes R-squared (adj) to increase to 41.7% so it is concluded that the addition of the extra variables is useful.

Model Summary[d]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.513[a]	.263	.258	5.693
2	.580[b]	.336	.327	5.421
3	.655[c]	.429	.417	5.046

a. Predictors: (Constant), Rhythm

b. Predictors: (Constant), Rhythm, Raven

c. Predictors: (Constant), Rhythm, Raven, Gender

d. Dependent Variable: Language (Recep)

Table 4. Summary of Regression Models, Phase I

From the variables measured here, it can be concluded that the best way to predict Language (Recep) scores is by using Raven, Rhythm and Gender. All three variables make a significant contribution in predicting Language (Recep) scores.

5.3 Phase II – Examination of Productive Skills

Variables measured in Phase II are Music (Prod) and Language (Prod). Language consists of Spanish pronunciation, Spanish rhythm and Korean. Raven scores are available from Phase I. A Pairwise Variability Index was also calculated for each subject. These are compared with the scores awarded by the native speakers for rhythm.

5.3.1 Subjects

The subjects in Phase II (Group 2) are a subset of the subjects in Phase I (Group 1). In order to examine whether or not the participants in Phase II (n=41) are representative of the complete group (Group 1), it is necessary to consider the scores of the two groups on the Music (Recep) test, Language (Recep) and Raven tests, i.e. the tests that all subjects completed in Phase I. If no significant differences emerge between the two groups, it is justifiable to say that Group 2 is representative of the whole sample in Group 1. In all cases a two sample t-test shows that differences between the groups are insignificant.

It was noted above (*Section 5.2.2.2*) that a 95% Confidence Interval gives the range within which the true population statistic should fall. If there is no true difference between the groups, the difference in the mean scores of the two groups should be zero or as near to zero as makes no difference statistically; i.e. zero should be included in a 95% Confidence Interval. Ninety-five percent confidence intervals for the differences in mean scores of Group 1 and Group 2 always include zero in this case. Thus, the difference between the two groups may be zero. Even if the difference is not zero, it is insignificant.

95% Confidence Intervals of the difference in mean scores between Group 1 and Group 2:

Music (Recep): [-4.07, 1.55].

Language (Recep): [-3.96, 0.78].

Raven: [-9.01, 9.51].

5.3.2 Distribution of scores

As explained in *Section 5.2.1* above, it is worthwhile examining the distribution of scores for each variable. By looking at the distribution of Total Music (Prod) it can be seen that subjects scored higher than predicted by the normal distribution, see Figure 37. The reader should not be unduly concerned about the fact that this plot may look as if it does not fit the normal distribution. In fact a Normal Q-Q plot shows that indeed these data do follow the normal distribution. The interested reader should refer to Appendix 2.

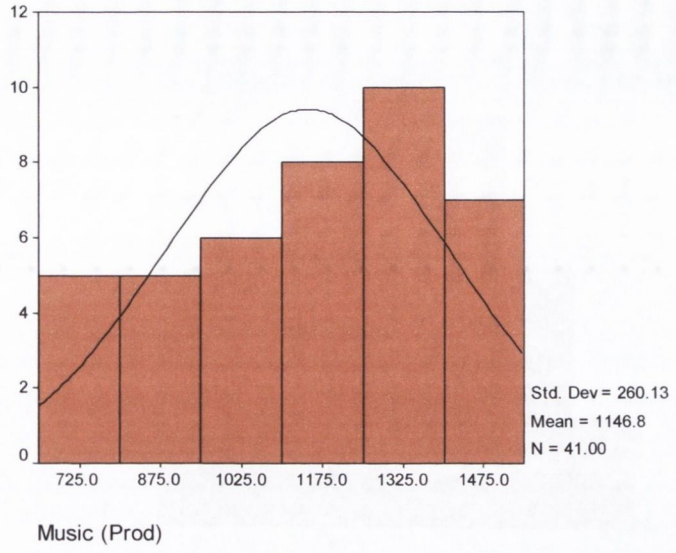


Figure 37. Distribution of Music (Prod) scores

Figure 38 tells us that in Language (Prod) subjects scored slightly lower than predicted by the Normal distribution but the distribution is approximately normal (see Appendix 2 for Normal Q-Q plot). The fact that subjects scored somewhat lower than predicted by the normal distribution in language means that the language test proved relatively difficult. As subjects scored higher than predicted by the normal distribution in music, this means that the music test was somewhat easier.

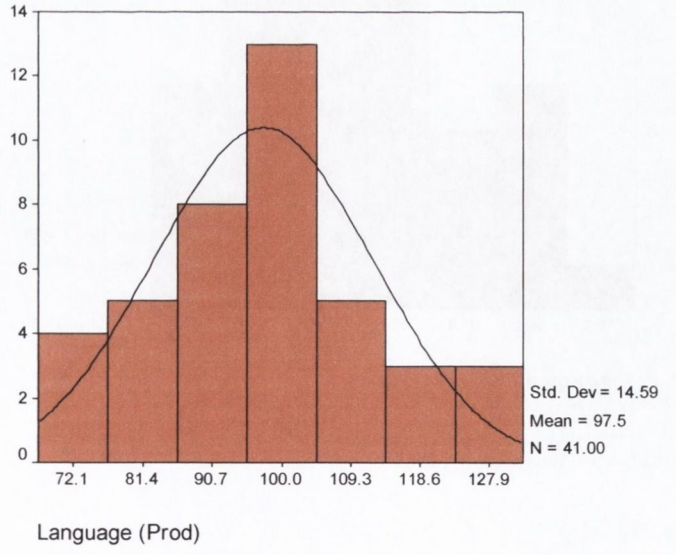


Figure 38. Distribution of Language (Prod) scores

In Spanish scores were lower than predicted by the normal distribution.

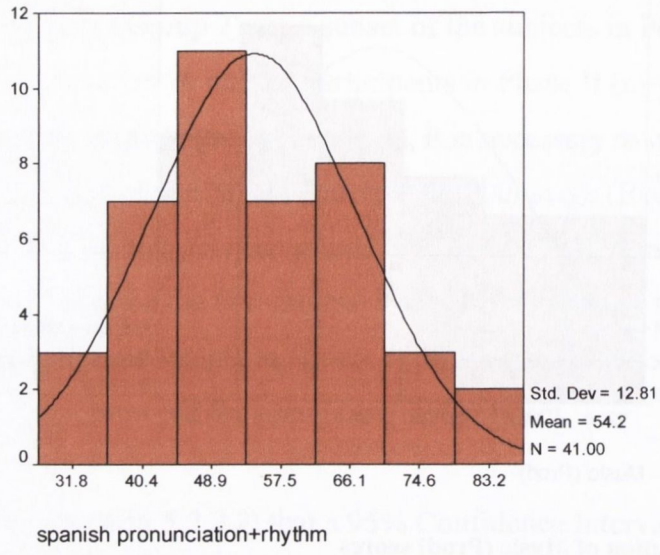


Figure 39. Distribution of Spanish scores

In Korean distribution is approximately normal but there are more high scores than low scores.

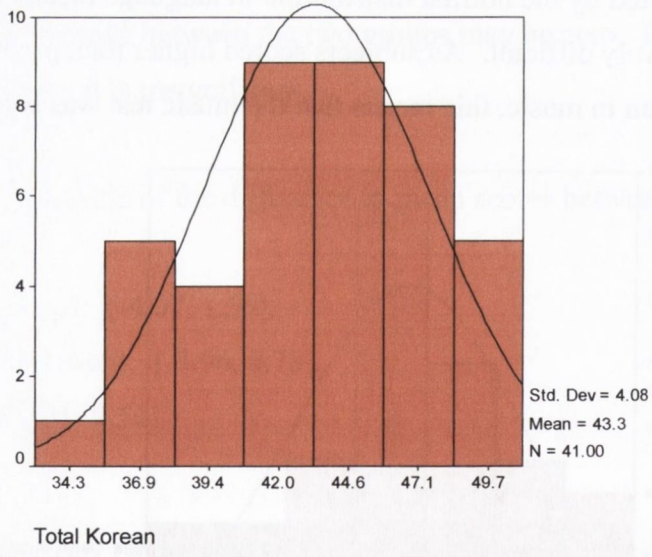


Figure 40. Distribution of Korean scores

5.3.3 Correlations

As discussed in above in *Section 5.2.2*, the Pearson correlation provides more information than Spearman's rho and differences in results between the two are negligible here. Therefore, the Pearson correlation is presented here. Spearman's rho is available in Appendix 2.

5.3.3.1 Correlations between components of Language test

Spanish Pronunciation and Spanish Rhythm

Spanish pronunciation and rhythm show a highly significant correlation ($r=0.61$, $p<0.01$). This is perhaps not unexpected, particularly as both scores were given by the same native speaker. However,it can be seen that pronunciation and rhythm are not perfectly correlated, thus some subjects are deemed better in one area than in the other.

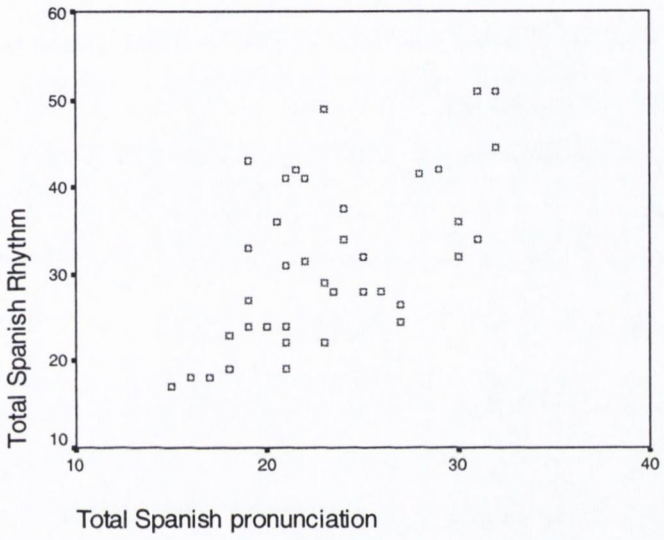


Figure 41. Spanish Pronunciation and Rhythm scores

Total Spanish and Korean

Total Spanish and Korean show only a very low correlation which just fails to reach significance ($r=0.31$, $p=0.051$). This is surprising if both are indeed tests of language aptitude and the consequences of this lack of correlation will be discussed in *Section 5.6* below.

5.3.3.2 Correlations between tests

As in Phase I, it is necessary to examine correlations between the main variables of interest. Correlations are given with exact p-values where possible in Appendix 2.

Language (Prod) and Music (Prod)

Language (Prod) and Music (Prod) show a correlation of 0.55 ($p<0.01$). This is a large correlation according to Cohen (1988). A 95% Confidence Interval for this

correlation is [0.29, 0.73]. The magnitude of this correlation is certainly noteworthy. It provides strong evidence in favour of the research hypothesis of this study that there is a link between skills in the musical and linguistic domains. Of course this correlation cannot give information on confounding variables such as intelligence, but this is considered below (see *Section 5.3.4*).

Spanish and Music (Prod)

There is a correlation of r=0.55 ($p<0.01$) between Spanish and Music (Prod). A 95% CI for this correlation is [0.29, 0.73].
Given that this is the same as the correlation between Music (Prod) and total Language (Productive), i.e. Korean and Spanish together, it can be seen that it is actually the influence of Spanish which causes the positive correlation between music and language. Korean does not add anything to the relationship. This is discussed further in the following sections.

Korean and Music (Prod)

There is not a significant correlation between Korean and Music (Prod), r=0.24, (p=0.13). A 95% CI for this correlation is [-0.07, 0.51].

When zero is included in a 95% confidence interval for a correlation, it means that a significant effect is not found. This is because a correlation of zero implies that no linear relationship holds between the two variables. The confidence interval presented above shows that the correlation here just fails to be significant because zero is at the very edge of the interval. If the sample size was slightly larger, it might result in a slightly narrower confidence interval. This is because an increase in sample size means that it is possible to be more confident about the true value of the statistic, which in turn decreases the width of the confidence interval. Decreasing the width of the interval might remove zero, in which case the correlation would become significant. However, in any case, a correlation of 0.24 is small using Cohen's (1988) classification system. Therefore, even if the correlation became statistically significant, in practice, there is still only a weak relationship between Korean and Music (Prod).

It is interesting that Spanish shows a large correlation with Music (Prod) whereas Korean reveals only a small, insignificant correlation. This suggests that there may be some fundamental difference between the two. However, considering both Korean and Spanish as tests of language aptitude and comparing the two confidence intervals produced for their correlations with music, it is found that the two are indeed compatible with an underlying significant correlation of between 0.29 and 0.51 between music and language aptitude, i.e. the CI for Korean and music [-0.07, 0.51] overlaps with the CI for Spanish and music [0.29, 0.73] between 0.29 and 0.51. Interestingly this interval of [0.29, 0.51] is also compatible with the correlation found between Music (Recep) and Language (Recep) in Phase I ($r=0.37$). Taken together, these findings provide strong evidence of a moderate to large correlation between music and language aptitude in both the receptive and productive domains.

Inspection of a 95% CI of the difference between the two correlations, i.e. Music (Prod) and Spanish compared to Music (Prod) and Korean, reveals that the CI includes zero, [-0.68, 0.07]. This shows that the difference between the two correlations is not statistically significant, i.e. there is not a significant difference between the correlation found between music and Spanish and that found between music and Korean.

Taking Korean and Spanish together as language aptitude tests, this study provides compelling evidence of a strong link between musical aptitude and language aptitude in the productive domain. If, on the other hand, we wish to claim that there is a true difference between the Korean and Spanish tests, one suggestion could relate to the length of the Korean words. The Korean words are short; the mean length is 0.65 seconds (std. dev. 0.26). This is likely to be well within the capacity of the phonological loop, given the suggestion in *Section 2.7.1* that its capacity is about two seconds.

On the other hand, the music patterns and Spanish sentences are somewhat longer. The mean length of the music patterns is 3.34 seconds (std. dev. 0.78), while the mean length of the Spanish sentences is 1.72 seconds (std. dev. 0.44). It is likely that the music patterns are beyond the capacity of the phonological loop. The

Spanish sentences may or may not be within this span as if indeed there is a limit of approximately two seconds, 1.72 seconds is quite near this limit.

Therefore, it could be the case that the Spanish sentences and musical patterns are processed somewhat differently to the Korean words as a result of the differences in length. The result of this could be that the correlation between Korean and Music (Prod) is insignificant while that between Spanish and Music (Prod) is significant. Clearly this would require much closer investigation. For now, it seems preferable to focus on the fact that the confidence intervals for the two correlations (i.e. between Korean and Music and between Spanish and Music) are actually compatible and that there is no significant difference between the two. Also, the point of primary interest is that a large and significant correlation is found between language aptitude and music aptitude in the productive domain.

Raven and Language (Prod)

Language (Prod) is the sum of the Spanish and Korean scores. Correlation between Raven and Language (Productive) is small ($r=0.23$) and not significant. A 95% Confidence Interval for this correlation is [-0.08, 0.50].

The question which arises is whether this lack of significant correlation is indeed due to the absence of a correlation in the underlying population, or simply due to a lack of power in this study. The power of a study is the chance of finding a genuine effect when there is one, i.e. correctly rejecting the null hypothesis.

It is proposed that this study is not adequately powerful to pick up such small correlations; e.g. in order to detect a correlation of 0.4 with 80% power, a sample size of 47 is needed. Here there were only 41 subjects. Hence, it cannot be concluded that no correlation exists between Raven and Language (Prod), merely that this study is not adequately powerful to detect it. Given a sample size of 41, the power of this study to pick up a correlation of 0.23 is only 31%, i.e. there is only 31% chance of finding a correlation of 0.23 even if one truly does exist. Unfortunately this power calculation was done after the study was conducted so it was not possible to increase the number of subjects.

In Phase I a correlation of r=0.40 was found between Language (Recep) and Raven. A 95% confidence interval for this correlation was given as [0.25, 0.52]. The CI found here [-0.08, 0.50] is actually compatible with that of Phase I, i.e. the two overlap. Therefore, it can be concluded that a similar correlation is likely to hold between Language (Prod) and Raven as between Language (Recep) and Raven. This correlation is likely to be in the region [0.25, 0.50] as this is the overlap between [-0.08, 0.50] and [0.25, 0.52].

Raven and Music (Prod)

The correlation between Raven and Music (Prod) is small (r=0.17) and is not significant. A 95% Confidence Interval for the correlation is [-0.15, 0.45]. Again this may be an issue of power rather than the absence of a significant correlation. Given a sample of 41, the power to detect a correlation of this size is only 19%.

As in the case of Raven and Language (Prod) discussed above, the confidence interval found here for the correlation between Raven and Music (Prod) is actually compatible with that found in Phase I for the correlation between Raven and Music (Recep). Phase I produced a 95% Confidence Interval of [0.19, 0.47] for the correlation between Raven and Music (Recep). Phase II produced a 95% CI of [-0.15, 0.45]. These confidence intervals overlap between 0.19 and 0.45 and therefore are compatible with an underlying correlation of between 0.19 and 0.45.

5.3.4 Subjects divided by Raven scores

As in Phase I, subjects are divided into two groups on the basis of Raven scores. The groups are as in Phase I, where subjects with average or above Raven scores form one group (AAS) while subjects with below average Raven scores form the other group (BAI). The problem with this division here is that it results in small groups, i.e. 18 subjects in BAI and 23 in AAS. Unless a correlation is very large indeed these small samples will not permit the study to be adequately powerful to pick it up.

The mean Language (Prod) score appears to be higher in the AAS group, see Figure 42, and scores also seem to be more widely dispersed within this group. There is much less variation of scores in the BAI group.

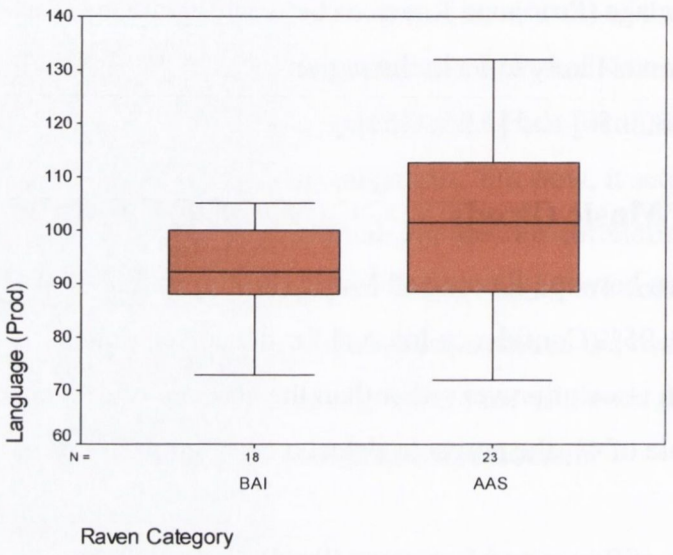


Figure 42. Language (Prod) scores according to Raven category

A formal examination of the differences using an independent samples t-test reveals that the difference in Language (Prod) scores between the two groups is significant (equal variances not assumed, $t=-2.21$, $df=35.69$, $p<0.05$, see Appendix 2 for details). A 95% Confidence Interval for the difference is as follows: [-17.53, -0.76].

The mean score for Music (Prod) also seems to be higher in the AAS group, see Figure 43. In this case, scores are less variable in the AAS group than the BAI group.

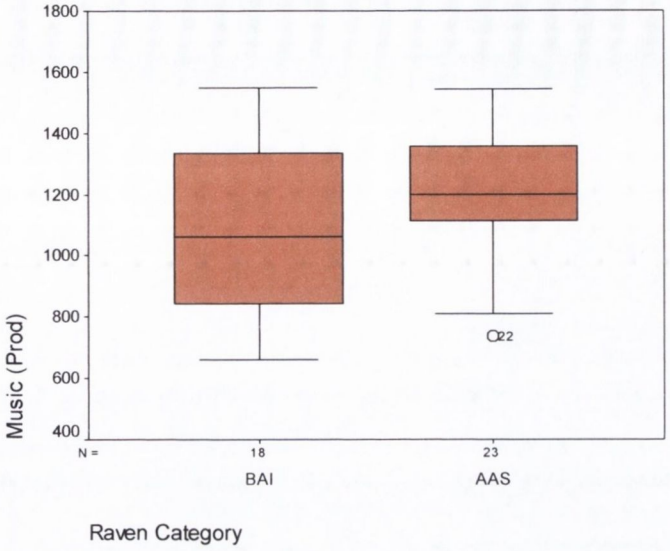


Figure 43. Music (Prod) scores according to Raven category

The difference in Music (Prod) scores between the groups is not significant, (equal variances not assumed, t=-1.36, df=29.46, p=0.19).

Looking more closely at the Spanish and Korean tests individually shows that the difference between the two groups in Spanish is significant while the difference in Korean fails to reach significance. A 95% CI for the difference in Spanish scores is: [-16.67, -2.2], (equal variances not assumed, t=-2.64, df=36.5, $p<0.05$). While a significant difference is shown, this is a very wide interval and the difference between the groups may be as little as 2 points. Given that the maximum possible score on the Spanish test is 150, a 2 point difference is really rather small. Even the possible 16 point difference is not huge, given the range of possible scores.

A 95% CI for the difference in Korean scores is [-1.35, 3.69], (equal variances not assumed, t=0.94, df=37.91, ns.).

Therefore, it can be concluded that there are no major differences between the high (AAS) and low (BAI) Raven groups in performance on the music and language tests in Phase II, although the high Raven group show slightly superior performance on the Spanish test.

To summarise, the 95% Confidence Intervals given above for the difference in mean scores between high Raven and Low Raven Groups are given again below:

Korean: [-1.35, 3.68]

Spanish: [-16.67, -2.20]

Music (Prod): [-286.64, 57.90].

Having established that there are no major differences in mean scores between the groups, let us now turn to the pattern of correlations within the two groups.

5.3.4.1 Correlations within high and low Raven groups

This section examines the correlations which hold between the different variables within the two groups. This allows comparison of the relationship between music and language in the two groups which should reveal if the two are more closely or less closely related in either group. In Phase I it was noted that there was a tendency for greater correlations in the low Raven group – a finding which is in keeping with the literature on the Ability Differentiation Hypothesis. This may or may not also arise here.

Music Productive and Language Productive

Correlations between Music (Prod) and Language (Prod) in each of the two groups are as follows:

Low Raven group: $r=0.44$, ns.

High Raven group: $r=0.70$, $p<0.01$.

While at first sight it seems interesting that the correlation should be significant in one group and not the other, closer consideration reveals that the difference in the correlations is not in fact significant. A 95% CI for the difference is as follows: [-0.76, 0.34]. Thus, the difference is not significant.

Music Productive and Raven

Correlations in the two groups are as follows:

Low Raven: $r=0.13$, ns.

High Raven: $r=-0.07$, ns.

As above, the difference in correlations between the two groups is not significant. A 95% CI for the difference is as follows: [-0.44, 0.70].

Language Productive and Raven

Correlations in the two groups are as follows:

Low Raven: -0.06, ns.

High Raven: -0.05, ns.

Again, a 95% CI for the difference reveals that it is not significant and is as follows: [-0.13, 0.84].

Within Language Test

Spanish and Korean

Low Raven: r=0.3, ns.

High Raven: r=0.52, $p<0.05$.

Again the difference between the groups is not significant. A 95% CI for the difference is as follows: [-0.75, 0.35].

Music Productive and Spanish

Correlations for the two groups are as follows:

Low Raven: r=0.58, $p<0.05$

High Raven: r=0.56, $p<0.01$.

A 95% CI for the difference is as follows: [-0.57, 0.60]. The difference is not significant.

Music Productive and Korean

Correlations for the two groups are as follows:

Low Raven: r=-0.10, ns.

High Raven: r=0.65, $p<0.01$.

Here a significant difference emerges between the two groups with the high Raven group showing a large significant correlation between music and Korean and the low Raven group showing almost no correlation. Results for the low Raven group are quite puzzling as they clearly go against the Ability Differentiation Hypothesis.

A 95% CI for the difference between the two correlations is as follows: [-0.91, -0.21]. This shows that there is indeed a significant difference between the correlations in the two groups. The difference in the correlations is anywhere

between 0.21 and 0.91. However, the confidence interval is very wide - almost as wide as possible without including zero. It is therefore difficult to be certain about the actual size of the difference between the two groups. Nonetheless a difference does seem to exist. This issue would clearly require a much more detailed study than is possible here as the aim of this study is to consider language aptitude overall rather than one component of the aptitude test.

5.3.4.2 Summary of Low/High Raven split for Phase II

- There is little difference between mean scores in the two groups although Total Spanish is slightly and significantly higher in the High Raven group. This in turn means that Language (Prod) is significantly higher in the High Raven group.
- There is much more variability in language scores in the High Raven group, (see Figure 42, where the AAS box is seen to be much wider than the BAI box). Whereas all those in the BAI group received low language scores, at least some of those in the AAS group received high scores. This suggests that while high Raven does not guarantee a higher score in Language, it is a possibility for those with high Raven scores to achieve high Language scores but much less likely for those with low Raven scores.
- There is no significant difference between the groups in the mean music score although there is much more variability in the Low Raven group than in the High Raven group, see Figure 43. This seems to suggest that given a certain level of non-verbal intelligence (average or above), a certain minimum level will be achieved in the music test, i.e. all those in the high Raven group achieved a certain minimum level. Even without this level of non-verbal intelligence, a good score may be achieved but it is also possible to get a much poorer score, i.e. some of those in the low Raven group also achieved this minimum level but some did not. It is interesting that this finding is in line with the findings from Phase I which also showed that high Raven was not a necessary requirement for achieving a high music score.
- When the above two points are taken together, it is seen that in the area of language, high Raven scores seem to confer an advantage insofar as those with higher Raven scores have the possibility of achieving much higher

language scores than those in the low Raven group. However,in music even those with low Raven scores may achieve the level of the high Raven scorers, although many do not. In summary it seems that high language scores will not be achieved without high Raven scores. High music scores may be achieved without high Raven scores.

- In general, there is little difference between the groups in the pattern of correlations uncovered. However, there is a significant difference between the groups in the correlation between Music (Prod) and Korean. The High Raven group shows a significant correlation between the two whereas the Low Raven group does not. It is also noted that the confidence interval for the difference is very wide and therefore much uncertainty surrounds the size of the difference. Therefore, it is important not to overemphasise the importance of this difference until further research is carried out.

5.3.5 Subjects divided by gender

The aim of this section is to investigate whether males and females differ in their productive musical and linguistic aptitude.

In Music (Prod), females (n=25) have a mean score of 1192.25, [std. dev. 265.01] and males (n=16) have a mean score of 1075.78, [std. dev 243.43]. In Language (Prod), females have a mean score of 96.24, [std. dev 15.69], while males have a mean score of 99.59 [std. dev 12.91]. Neither of these differences is significant.

95% CIs for the differences in both cases include zero and are as follows (equal variance not assumed):

Music (Prod): [-47.52, 280.46].

Language (Prod): [-12.48, 5.77].

It is important to note that the sample size here is relatively small (n=41). This may be a contributory factor in accounting for the lack of significant differences between males and females. Of course, the other possibility is that there is indeed no true difference between males and females in the Language (Prod) or Music (Prod) tests.

5.3.6 Partial Correlations

As in Phase I, it is useful to consider partial correlations between Music (Prod), Language (Prod) and Raven. Substituting the values in the formula given in *Section 5.2.5* results in the following:

r(Language Prod, Music Prod|Raven) = 0.535, p=0.0003
95% Confidence Interval = [0.273, 0.724].

This is highly significant and a large correlation (Cohen, 1988). Removing the contribution of Raven reduces the correlation between Music (Recep) and Language (Recep) only from 0.553 to 0.535. This is a small reduction and shows that Raven does not play a major part in accounting for the correlation between these two variables. Again regression is another useful way of considering the predictive power of music and Raven for language aptitude scores. This is done in the following section.

5.3.7 Regression

Figure 44, which shows a normal probability plot of the residuals, reveals a reasonably straight line. Regression is therefore an appropriate analysis.

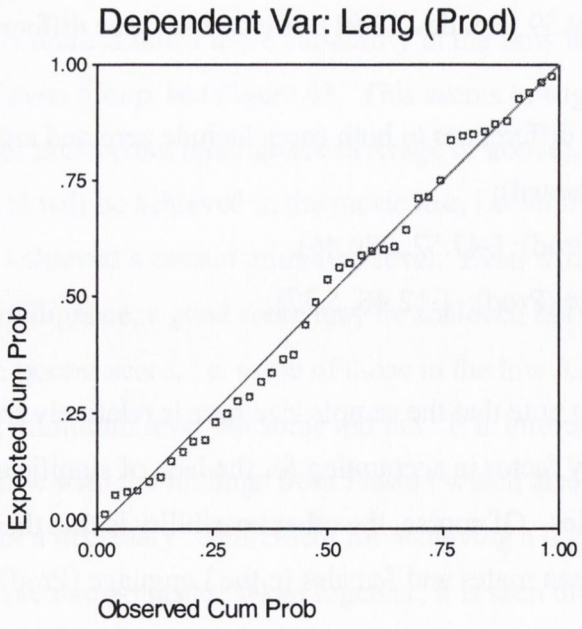


Figure 44. A normal probability plot of residuals

Using all subjects in Phase II (n=41), 28.9% of the variance in Language (Prod) is accounted for by Raven and Music (Prod) as R-squared (adj) = 28.9%. The regression equation is as follows (see Appendix 2 for details):

Language (Prod) = 60.204 + 0.08 Raven + 0.03 Music (Prod).

Raven does not make a significant contribution to the model, p=0.31. Music (Prod) is significant, p<0.01.
The fact that standardised coefficients are 0.14 Raven and 0.53 Music (Prod) underlines the greater role played by music than Raven in determining Language (Prod).

A plot of Cook's Distances shows those points which strongly influence the coefficients so that they can be excluded.

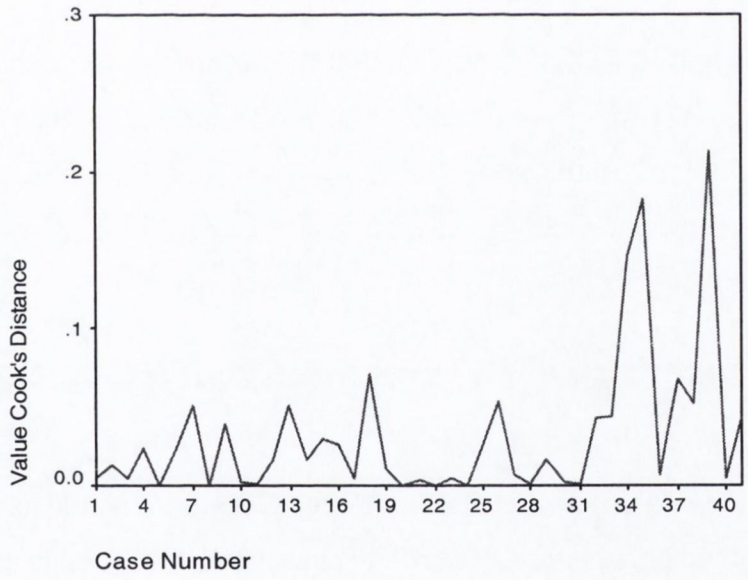


Figure 45. Cook's Distances, Phase II

With the three largest Cook's Distance points omitted, the regression equation is as follows:
Language (Recep) = 61.65 + 0.11 Raven + 0.03 Music (Prod).
Raven is not significant, p=0.16; Music (Prod) is significant p<0.01.
Standardised coefficients are 0.20 Raven, 0.51 Music (Prod).

R-sq(adj) is now 29.0%. Thus, 29% of the variance in Language (Prod) scores is accounted for by variation in Music (Prod) and Raven scores.

Stepwise Regression highlights the fact that only Music (Prod) is a significant predictor of Language (Prod). In choosing the best model it omits Raven. The model it produces has R-sq (adj)=27.0%. The Regression equation is:
Language (Prod) = 64.35 + 0.029 Music (Prod).
Music (Prod), $p<0.01$. Standardised coefficients: 0.538 Music (Prod).

Gender was also considered as a predictor variable but it is omitted by stepwise regression as an insignificant predictor. One possibility is that sample size may be a contributory factor in accounting for the failure of gender to emerge as a significant predictor in Phase II.

This section has shown that Music (Prod) is a more useful predictor than Raven of Language (Prod). This is significant in light of the question in *Section 5.2.2.2* which asks if the relationship between music and language aptitude is simply due to the influence of gF intelligence on both. Here it is shown that music makes a significant contribution to the prediction of language aptitude scores over and above gF intelligence.

5.3.8 Comparison of PVI measures with Native Speaker scores

Grabe and Low (2002) predict that stress-timed languages would exhibit high vocalic nPVI and high intervocalic rPVI values. Syllable-timed languages would have low vocalic nPVI and low intervocalic rPVI values, (see *Section 4.3.1*). Thus, they found that British English (stress-timed) has nPVI=57.2 and rPVI=64.1, whereas Spanish (syllable-timed) has nPVI=29.7 and rPVI=57.7. From this it can be seen that a stress-timed language has a much higher nPVI value than a syllable timed language. In addition a stress-timed language has a higher rPVI value but the difference is relatively small, see Figure 46. Henceforth the Grabe and Low (2002) study will be referred to by the initials GL.

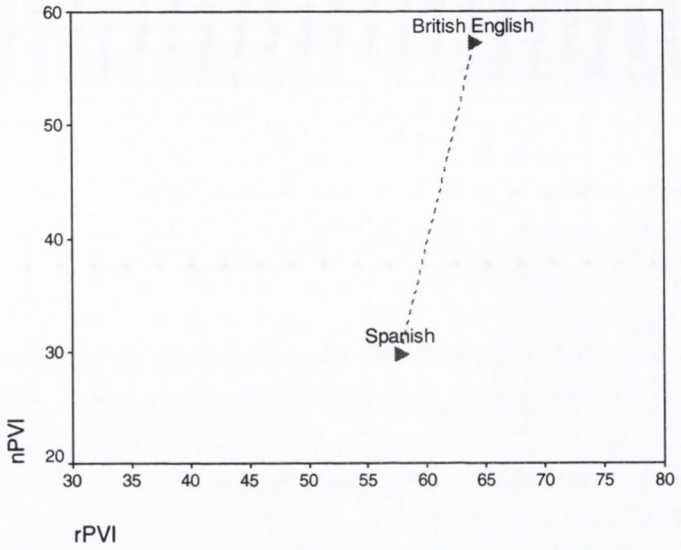


Figure 46. Findings of Grabe and Low (2002)

Measurements of the intervals produced by the native Spanish speaker in the current study reveal an nPVI value of 49.1 and rPVI of 52. English sentences produced by the native English speaker give values nPVI=51.4 and rPVI=72. While these values follow the trend shown in GL, i.e. English nPVI>Spanish nPVI and English rPVI >Spanish rPVI, the differences are not in the same proportion. Here English nPVI is only slightly larger than Spanish, whereas in the GL study the difference is much greater. Here English rPVI is much greater than Spanish, whereas GL found a small difference. Figure 47 presents this information in a graphical format.

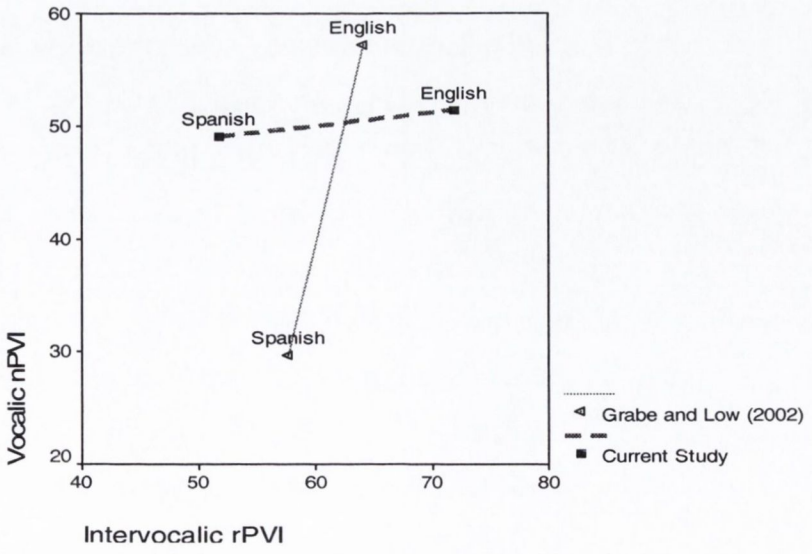


Figure 47. PVI values found in Phase II compared to literature

5.3.8.1 Why the differences?

It is important to note that in both the Grabe and Low (2002) study and in the present study for calculating reference values, measurements were taken from only one speaker per language. According to Miller (1984), it is likely that some rhythmic variation can be due to differences between speakers.

Secondly, the sentences in this study were short so as to be suitable for imitation and there were only 15 sentences in total. Thus, perhaps the length of the text was an issue. However, as the main focus of the current study is how non-native speakers compare to a native speaker, it is not the absolute values which are important but rather the ratios of PVI values of non-native speakers to native speakers.

With these results in mind, let us examine the PVI scores of the subjects who took the test. The productions of forty subjects were subjected to PVI analysis. While Phase II had forty-one subjects, one subject's recording was of poor quality and could not be segmented accurately. Figure 48 shows the values of the native speakers of Spanish and English in both this study and that of Grabe and Low (2002) (as in Figure 47). In addition the Vocalic nPVI and Intervocalic rPVI of each of the 40 subjects are given.

Clearly the GL value for Spanish (point D on Figure 47) is far below anything produced by the participants in this study. Even the native Spanish speaker (point A Figure 48) in this study failed to approximate the GL value. On the other hand, the GL British English value (point C) is reasonably near the native English speaker value found in this study (point B). Thus, if we consider a point halfway between point B and point C to represent a native speaker of English (say point X), we can see that the majority of utterances produced by subjects in this study lie to the left of this point. This means that most subjects produced utterances with smaller rPVI values than that produced by an English native speaker. In this way subjects incorporated at least one aspect of Spanish rhythm into their productions.

Let us now consider a point Y between points A and D which represents an average Spanish speaker. In general points representing subjects in this study lie to the right of this point. This means that subjects produced utterances with larger rPVI values than that produced by a Spanish native speaker. Many points lie between points X and Y. These subjects produced utterances with rPVI values somewhere between those of native English speakers and native Spanish speakers. This is what would be expected as the subjects are native speakers of English attempting to approximate Spanish rhythm.

This did not happen on the nPVI axis, i.e. it is not the case that most of the points lie between X and Y. All nPVI values are higher than the nPVI value of point Y. Many are higher than the nPVI value of point X. Thus, in the vocalic nPVI dimension, all subjects produced nPVI values higher than those of a native Spanish speaker. Many subjects even produced values higher than those found in the speech of an English native speaker. Some suggestions are presented below as to why this might be the case.

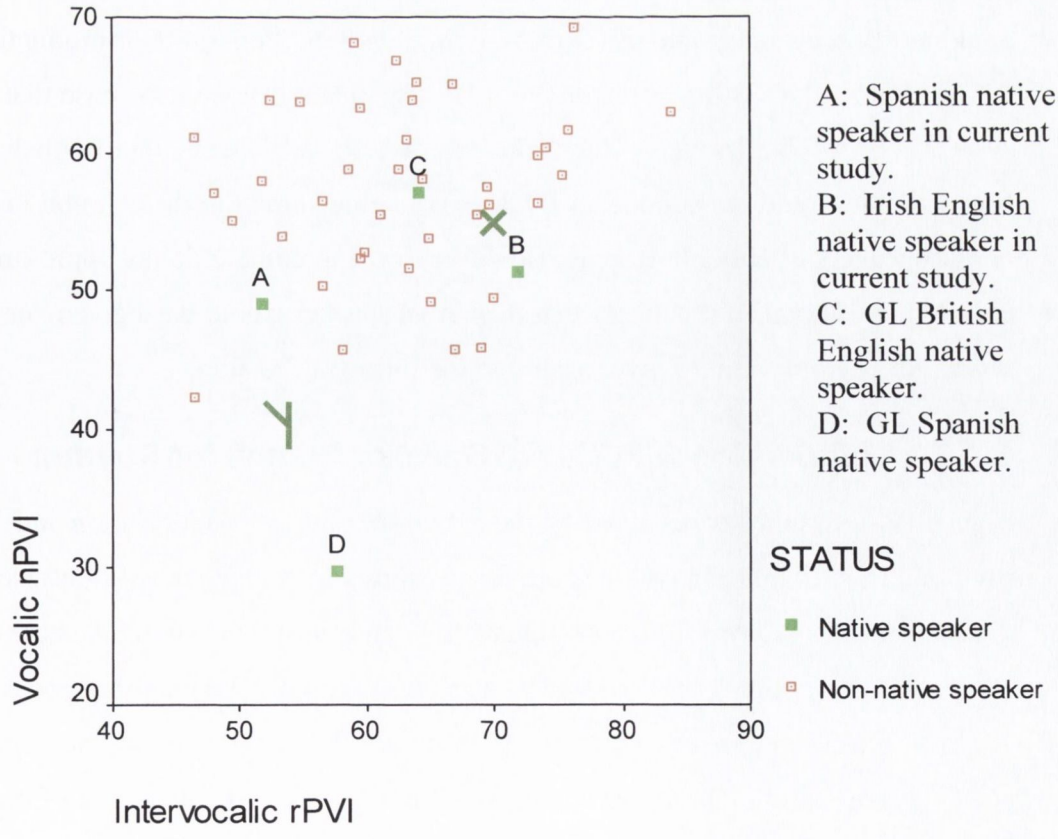


Figure 48. NS PVI values contrasted with NNS values

Figure 48 shows clearly that there is no particular pattern of clustering around either English or Spanish values.

5.3.8.2 Why values are not near English or Spanish values?

The aim of this section is to consider why there is no apparent clustering around English or Spanish values. It was evident during the process of segmenting the utterances that in some cases where subjects did not remember the utterance, they produced a relatively long vowel sequence as a filler, e.g. /aioria/. A few exceptionally long vocalic intervals would inflate the vocalic nPVI score for an individual. It is easy to produce long vowels. Some consonants on the other hand are difficult or impossible to stretch; e.g. plosives such as /b/ or /p/ are almost impossible to elongate. This may have contributed to the fact that subjects could not stretch consonants and therefore could not get very high intervocalic scores.

More generally, in order to account for the poor rhythm scores as measured by PVI, it could be proposed that subjects focussed on pronunciation to the detriment of rhythm. Hence, if they paid careful attention to individual sounds, maybe they could not fit the sounds into the correct rhythm pattern. Perhaps their productions did not follow the rhythm of either English or Spanish. If it were the case that subjects produced utterances which sounded rhythmically like neither English nor Spanish, this would explain the PVI values not being similar to those found in native speakers of either language. However, if the utterances do not approximate the rhythm of Spanish it is likely that the native speaker would have given very low scores for rhythm. This is investigated in the following section.

5.3.8.3 Comparison with Native Speaker Scores for Rhythm

Figure 49 shows the scores given by the native speaker for pronunciation and rhythm. The minimum possible score for rhythm is 15, as the rater was advised to give 1 mark if there was little resemblance to native-speaker rhythm. A score of zero is not an option. There are fifteen sentences in total. The maximum possible score is 75 (15 sentences, 5 marks each). Similarly for pronunciation, the minimum possible score is 15 and the maximum possible score is 75. It can be seen from Figure 49 that the rhythm scores are generally higher than pronunciation scores but are a lot more variable. Thus, in rhythm there is a very large range of scores (range

[17, 51], mean=31.12, standard deviation=9.51). In general it seems that the native speaker deemed subjects better at imitating rhythm than pronunciation.

However, even the best score was only 51 out of a possible 75. Clearly even the best subject was still quite far away from achieving native speaker rhythm. Perhaps this is in fact similar information to that presented above, i.e. according to PVI values, no subjects were very close to Spanish native speaker rhythm. Similarly, according to native speaker judgements, even the best attempts were still well below native speaker level.

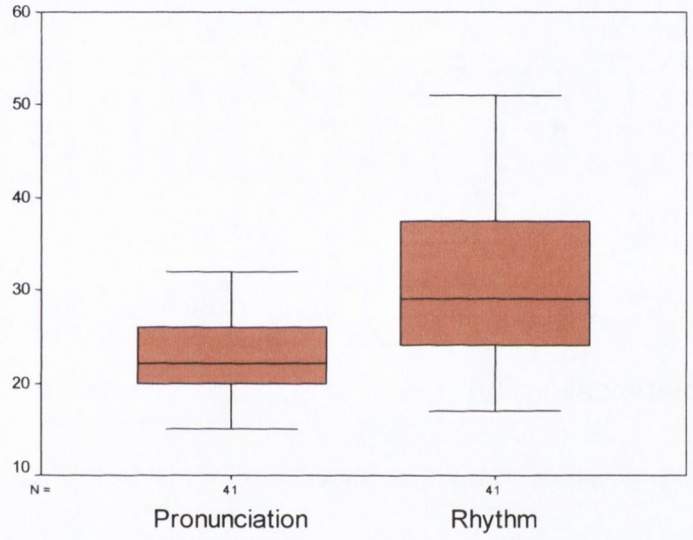


Figure 49. Boxplot of Spanish Pronunciation and Rhythm NS scores

Figure 50 shows almost the same information as Figure 48, i.e. the PVI measures of subjects in Phase II. However, in addition to the information found in Figure 48, native speaker results for rhythm are ranked and the top ten are marked in Figure 50. Interestingly, the person given the best score by the native speaker rater has an rPVI value equal to that of the native Spanish speaker. However, there are many other points with similar rPVI values to the native speaker value not given high scores by the native speaker corrector. Thus, there seems to be little correlation between the subjects judged to have good rhythm by the native speaker and those who appear to imitate the rhythm well as judged by their PVI measure. While it is not clear why this might be the case, it is clear that no subjects succeeded in achieving a very close approximation to Spanish native-speaker rhythm, as judged by both PVI and a native speaker.

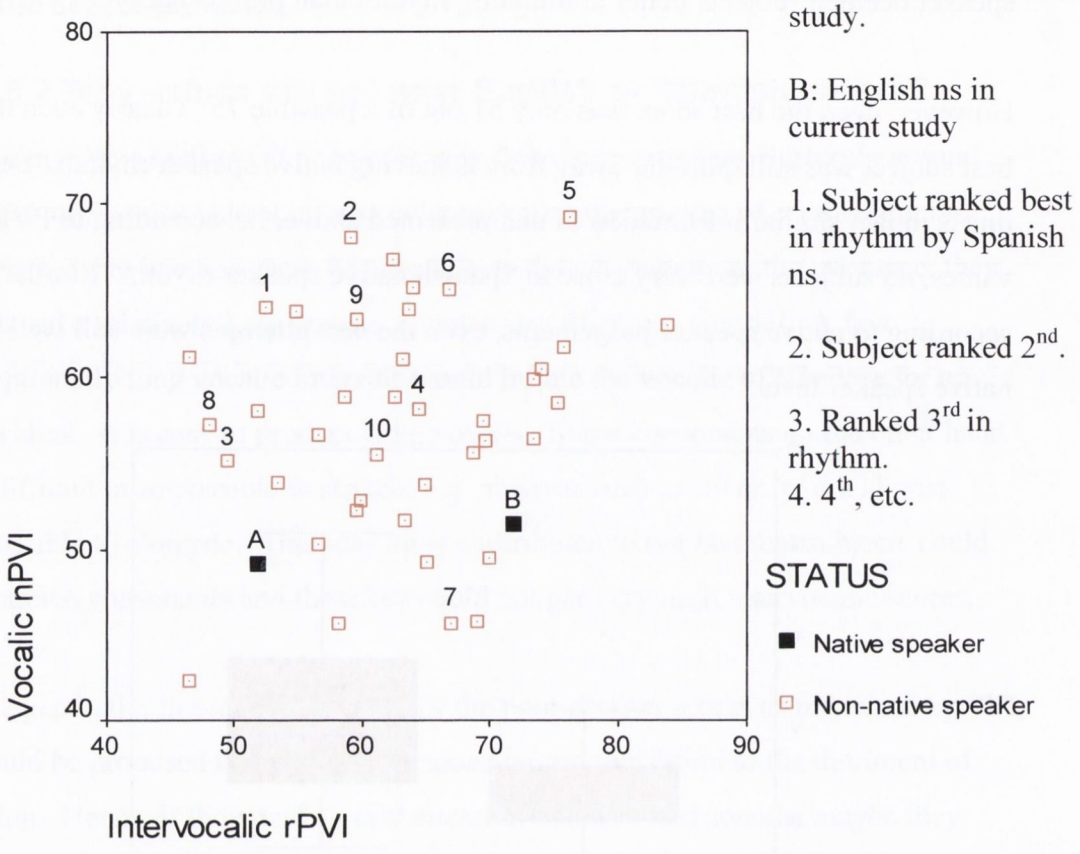


Figure 50. PVI scores ranked according to NS scores

5.3.8.4 PVI scores vs. ns scores

It is necessary to ask why 'good' PVI scores do not reflect native speaker scores. One possibility is that the native speaker is strongly influenced by pronunciation errors, i.e. if pronunciation is poor, rhythm is severely penalised. Indeed there is a high Pearson correlation between native-speaker scores for rhythm and pronunciation ($r=0.61$, $p<0.01$) so it seems that either subjects who were good at pronunciation were also good at rhythm or that in awarding rhythm scores the rater was heavily influenced by pronunciation. However, as the rater is a trained phonetician, it is likely that she was aware of this issue.

Otherwise, how do we account for the fact that the subjects who appear to most closely approximate syllable-timed rhythm according to their PVI score do not receive high scores from a native speaker rater? Perhaps it is necessary to consider what exactly a native speaker scores when asked to judge "rhythm". How is he/she influenced by intonation and speed of speech? It is also necessary to consider the

reliability of the PVI measure under non-optimal conditions, i.e. when speech contains filled pauses. As the PVI is a relatively new measure, it requires further examination of its usefulness under non-optimal conditions. Further research into the relationship between the PVI measure and native speaker judgements is beyond the scope of this thesis.

5.3.9 Conclusion

Phase II revealed a large significant correlation between music and language aptitude in the productive domain. This is a particularly interesting finding as productive tests are relatively new to the literature. Many issues emerged which may be worthy of further consideration; e.g. the difference between the Korean and Spanish tests and the difference between native speaker scores for rhythm and the Pairwise Variability Index. Additional work on these areas is beyond the scope of this thesis which set out to deal with music and language aptitude overall. However, brief consideration is given to the Korean and Spanish tests in *Section 5.6* below.

5.4 Phase III: Links between Receptive and Productive Skills

Phase III examines the association between the receptive and productive domains by considering the relationship between the results of Phases I and II. As Phase II involved 41 subjects, only the results of these 41 subjects may be considered for Phase III given that productive music and language results are only available for these 41 subjects. Therefore, for the sake of consistency, in Phase III when results of the receptive music and language tests are considered, it is the results of the 41 subjects who also participated in Phase II rather than the results of the whole 149 subjects who underwent tests in Phase I which are examined. It was noted above in *Section 5.3.1* that the subjects in Phase II are indeed representative of the whole group in Phase I. Hence, it is valid to look at this subset of results.

5.4.1 Correlations

5.4.1.1 Correlations Between Tests

Music Receptive Music Productive

Music (Recep) is measured in Phase I and Music (Prod) in Phase II. A correlation of $r=0.54$, $p<0.01$ is found between the two. This is the correlation which holds in the 41 subjects who underwent Phase II.

While this is a large correlation according to Cohen (1988), it is by no means perfect. It would seem plausible that the productive test is more susceptible to the influence of musical training, particularly the influence of percussion training, than the receptive test. It is likely that a subject who has received percussion training would have an advantage in this test more so than in the receptive test. This is surely one reason to account for the absence of a perfect correlation between the two.

Secondly, the productive test requires motor skills not involved in the receptive test. Subjects must control their tapping so as to sound out the pattern that is in their head. In the receptive test subjects simply have to notice differences but they are not required to reproduce them. The difference in production and perception is illustrated by a recent article which distinguished between inaccurate singers who were inaccurate because they failed to discriminate pitch accurately and those who, despite accurate discrimination abilities, failed to sing correctly for other reasons such as problems with the vocal mechanism (Bradshaw & McHenry, in press). A similar problem could occur with the Music (Prod) test in this study. A subject could fail to reproduce the pattern because he/she failed to hear and memorise the pattern correctly or because he/she failed to tap it correctly despite correct memorisation. While administering the test it was clear that some subjects seemed disappointed with their productions. They seemed to know that they were incorrect, suggesting that they may have memorised them correctly but failed at the reproduction stage.

It is also interesting to consider whether Music (Prod) is correlated with the rhythm scores from Phase I, as the Music (Prod) test seems to be primarily a test of rhythmic ability. The correlation between Music (Prod) and Rhythm proves to be $r=0.34$, $p<0.05$.

Given that this correlation between Rhythm and Music (Prod) is smaller than that between Music (Recep) and Music (Prod), it seems that Music (Prod) measures something more general than merely rhythmic ability. However, it is important to note that the difference between these two correlations, i.e. the difference between the correlation for Music (Prod) and Rhythm ($r=0.34$) and that for Music (Prod) and Music (Recep) ($r=0.54$), is not statistically significant, so it cannot be definitively said that there is a difference in the correlation between Rhythm and Music (Prod) and that between Music (Recep) and Music (Prod). A 95% CI of the difference between the two correlations is [-0.20, 0.60]. This study was not designed to determine if Music (Prod) is more closely related to Music (Recep) or rhythm specifically. Should this be a topic of interest, it would require more detailed study than is possible here.

Language Receptive Language Productive

There is an almost zero insignificant correlation between these two variables ($r=-0.062$, $p=0.70$). As in the case of the music tests, the productive language test requires motor skills which are not required in the receptive test. Not only do subjects need to hear phonemic contrasts, they must correctly position the articulators to produce them. It is likely that this is quite a demanding task particularly as subjects have had no previous experience of practising Spanish sounds.

This issue receives further consideration in *Section 5.6* below where the absence of a correlation between the receptive and productive language tests is considered.

Music Receptive Language Receptive (Phase I)

In Phase I, a correlation of $r=0.37$, $p<0.01$ was found between receptive skills in music and language. In the subset of subjects investigated in Phase II ($n=41$), a correlation of $r=0.59$, $p<0.01$ is found between Music (Recep) and Language

(Recep). However, the difference between these two correlations is not significant. Thus, it can be said that a similar correlation is found in both groups. A 95% CI of the difference between the two groups is as follows:
[-0.57, 0.07]. This is what is expected as the subjects in Phase II are representative of all the subjects in Phase I. Therefore, given that the correlation held in the whole group and that the subgroup is representative of the large group, the correlation should also hold in the subgroup.

Music Productive Language Productive (Phase II)

This relationship is discussed in *Section 5.3.3.2* above. A 95% CI for the correlation mentioned in this section is [0.30, 0.74].

Music Receptive Language Productive

Music (Recep) is measured in Phase I, Language (Prod) is measured in Phase II. There is no significant correlation between the two, ($r=0.181$, ns). As mentioned above, the Language (Prod) test demands quite specific skills in articulating phonemes which are unfamiliar to Irish subjects. It is perhaps for this reason that there is no significant correlation between Music (Recep) and Language (Prod). Hearing sound contrasts is not sufficient for the Language (Prod) test whereas it is sufficient for the Music (Recep) test.

Music Productive Language Receptive

Music (Prod) is measured in Phase II, Language (Recep) is measured in Phase I. The correlation between the two is moderate and significant, ($r=0.35$, $p<0.05$). One might expect that these two tests would not show a significant correlation on the grounds that the music test requires productive ability whereas the Language (Recep) test simply requires recognition skills. Above it was shown that Music (Recep) and Language (Prod) tests do not correlate significantly and it might be expected that a similar result would be found here.

However, the productive skills required by the music test are perhaps not as specialised as those required by the productive language test. In the productive music test, subjects simply have to tap a key on a keyboard whereas in the productive language test, they have to make sounds some of which do not exist in

English. In this way, perhaps a good deal of the Music (Prod) test depends on remembering sound patterns which is in fact what is also required for the Language (Recep) test. Therefore, a correlation emerges between Music (Prod) and Language (Recep).

5.4.2 Subjects divided by Raven scores

Group 2 subjects are again divided on the basis of Raven scores into an average or above group (AAS) and a below average group (BAI). As in Phase II, it is necessary to keep in mind the relatively small numbers in each group. There are 18 subjects in the below average group and 23 average or above. This may be a contributory factor in the failure to find significant differences between the groups. Of course it is also possible that no differences actually exist. The primary aim of this section is to investigate if additional issues emerge which may be worthy of consideration at a later stage. If so, work could be conducted using a larger sample size.

5.4.2.1 Correlations within high and low Raven groups

Music Receptive Music Productive

Correlations are as follows:

Low Raven: $r=0.76$, $p<0.01$.

High Raven: $r=0.19$, ns.

A 95% CI for the difference between the correlations in the two groups is as follows: [0.12, 0.90]. This interval does not contain zero and is therefore significant. As noted in *Section 5.2.2*, a positive correlation takes a value between 0 and 1. This CI is therefore very wide as it ranges from 0.12 to 0.90. A very wide CI means that the location of the value is not very certain – it can be located anywhere within this interval. Thus, the difference between the two may be as little as 0.12 or as large as 0.90. This issue would need much closer examination before any definitive conclusions can be drawn regarding differences in the two groups.

Language Receptive Language Productive

Correlations in the two groups are as follows:

Low Raven: $r=0.02$, ns.

High Raven: r=-0.20, ns.

The difference is not significant. A 95% CI for the difference is as follows: [-0.42, 0.71].

Music Receptive Language Receptive (Phase I)

Section 5.2.3.1 above shows that the difference between the groups is not significant. However, the correlation is slightly higher in the Low Raven group.

Music Productive Language Productive (Phase II)

Section 5.3.4.1 above shows that the difference between the groups is not significant. In this case the correlation between the two variables is slightly higher in the High Raven group.

Music Receptive Language Productive

The correlations in the two groups are as follows:

Low Raven: r=0.50, p<0.05

High Raven: r=-0.04, ns.

The difference is not significant. A 95% CI for the difference is as follows: [-0.08, 0.85].

Music Productive Language Receptive

Low Raven: r=0.59, p<0.01

High Raven: r=0.03, ns

A 95% CI for the difference is as follows: [-0.01, 0.87]. The difference is not significant.

Again it is found that while differences between the two groups generally do not reach significance, the trend is for the low Raven group to show slightly greater correlations than the high Raven group. This is in keeping with the Ability Differentiation Hypothesis discussed in *Section 5.2.3.1* above.

5.5 Multiple Testing

Clearly, very many tests are carried out in the course of this study. A danger of carrying out many tests is that it may give rise to Type I error. Type I error arises if

the null hypothesis is rejected erroneously; e.g. a study concludes that there is a correlation between two variables where in fact none exists. While formal correction methods (e.g. the Bonferroni method) are available to correct for Type I error, it was decided, as is common practice in the literature, not to employ them in this study. Rather, exact p-values are given where possible in the appendices in association with the correlation coefficients and the reader is invited to consider the strength and significance of the correlations.

In addition and more importantly for this study, the main emphasis is placed on tests which were chosen prior to data collection, i.e. testing of relationships between music and language (both receptive and productive) and non-verbal intelligence. Tests were therefore not carried out randomly but in accordance with the design of the study. While some additional tests were carried out; e.g. consideration of whether scores vary according to gender, secondary importance is placed on relationships which were discovered to be significant but which had not been explicitly set out as aims of the study.

5.6 Korean and Spanish as tests of Language Aptitude

As mentioned above, there is no (strong) linear relationship between Korean scores and Spanish scores. This is illustrated in Figure 51.

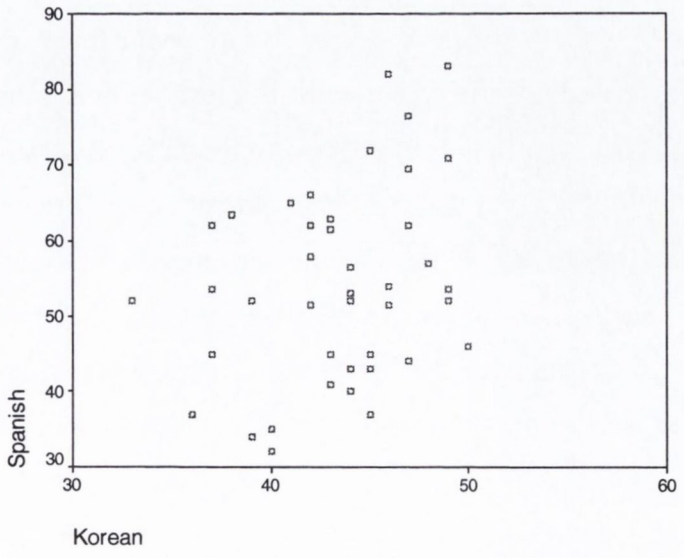


Figure 51. Scatterplot of Spanish and Korean scores

It is necessary to consider why this may be the case. The Korean words were either one-syllable, two-syllable or three-syllable in length. Thus, the stimuli were short (mean duration 0.65 seconds) and in all cases subjects were able to make an attempt at imitation.

It is important to note that Korean stops are distinguished on the basis of being aspirated, tense or plain[26]. The voiced/voiceless distinction used in English does not make phonemic contrasts in Korean. Voiced stops do arise but only in certain intervocalic environments; e.g. they never occur utterance initially. For Irish subjects, it is very difficult to hear the difference between aspirated, plain and tense stops but for a Korean native speaker, these distinctions are essential for distinguishing between minimal pairs. Three of the test words belonged to minimal pairs, e.g. item 2. A/PP/A with a tense stop means 'dad', while A/P/A (plain stop) means 'Abba' (the Swedish pop group) and A/PH/A (aspirated stop) means 'painful'. Item 6. /CH/AYK means 'book' while /CC/AYK means the sound of birds' singing. Item 12. /CH/ANGMWUN forms a minimal pair with /C/ANGMWUN. Item 13 /PP/ALLAY with a tense stop forms a minimal pair with /P/ALLY which has a plain stop. Thus, it seems likely that the Korean native speaker awarded scores on the basis of subjects being able to make very fine distinctions between sounds which are not used for phonemic contrasts in English.

The Spanish sentences were quite long - the mean number of syllables in each utterance was nine but the range was from five to twelve (mean duration of sentences 1.72 seconds). Hence, in the Spanish test, it seems more likely that very fine distinctions between sounds were not as important. Simply because there were more sounds in question, the Spanish rater could not have listened to each individual sound as closely. In this case perhaps misplaced aspiration in consonants would not be penalised as the sound was basically right. Perhaps the ability to actually produce foreign phonemes very, very accurately is different from the ability to remember a longer sequence of words and more broadly imitate it.

[26] Jeong-Young Kim (September 2004, personal communication)

As mentioned above, additional evidence for the difference between the Spanish and Korean tests comes from the area of memory research. As noted in *Section 2.7.1*, it has been hypothesised that the duration of short term memory is approximately two seconds. It would seem plausible then that the Korean words may be stored in STM, whereby the Spanish sentences require some other storage mechanism. Fraisse (1984, p.9*f.*), in his discussion of the perception and estimation of time, discusses estimation and perception of duration. He notes that perception of duration involves the psychological present which has no fixed duration. Rather it is based on what is perceived and refers to our capacity to apprehend sets of objects. He suggests that this psychological present is what is now commonly called short-term memory and proposes concrete examples of it as follows:

> "it applies to the perception of a telephone number, for example, or of a sentence simple enough to be repeated, or of a rhythmic pattern identified as a rhythm provided that its elements are perceived as being linked to each other in such a way as to form a unified group. The psychological present corresponds to the duration of an experiential process and not to a given period of duration. However, it has an upper limit which hardly exceeds 5 sec, and has an average value of 2 to 3 sec".

The suggestion of five seconds may be somewhat of an overestimation as other authors have proposed two seconds as a realistic upper limit, (see *Section 2.7* above). Therefore, as mentioned above, the Korean stimuli clearly fit within this duration whereas it is not clear in the case of the Spanish and musical stimuli that they are within this duration. This may lead to differences in the modes of processing of the different stimuli.

It is necessary to consider which of these, i.e. the ability to imitate sounds exactly or the ability to remember longer sequences of words, is a more realistic predictor of foreign language learning aptitude. It is possible that both abilities are useful although they may be somewhat separate. As shown in *Section 2.7.2.1*, previous researchers have shown that the ability to imitate short foreign words or non-words is a significant predictor of language learning success (Gathercole *et al.*, 1991; Papagno & Vallar, 1995; Service, 1992).

Others have used Working Memory tasks where subjects were required to remember information over a longer period of time while also carrying out a processing task (e.g. Daneman & Carpenter, 1980). These tests too are useful predictors of language learning success and seem to have a slightly different focus to the imitation of words tests. To the best of my knowledge, tests requiring simply the memorization of unfamiliar material longer than just a word, but without the processing component, have not previously been used in the language learning domain as tests of aptitude. Thus, as the Korean test follows others used previously in the literature, it is the more likely to be a valid language aptitude test. However, it was noted in Chapter 4 (*Section 4.5.3*) that the Korean test shows poor internal consistency. This means that not all items on the test are measuring the same underlying construct. The Korean test contrasts with the Spanish test which shows very good internal consistency. In this way, it is known that the Spanish test seems to measure a single underlying construct. This casts some doubt on the Korean test as a test of language aptitude.

The question which remains is whether or not the Spanish test is in any way a useful predictor of language learning success. If it is a valid test of language aptitude, the results of Phase II surely provide strong evidence for an association between musical and linguistic aptitude, given the correlation between Spanish and Music (Prod). As mentioned above, the fact that the Spanish test shows very good internal consistency is certainly one point in its favour. If the Spanish test is not a valid test of language aptitude, there is less evidence for this association. However, we are still left in a position where we must account for the correlation between Spanish and Music and the dissociation of the Spanish test from language aptitude.

Perhaps this could be accounted for through consideration of the work of Burnham *et al.* (1996), mentioned in *Section 5.2.2.1* above, who found that English speakers were better able to perceive differences in pairs of tones when they were presented as music rather than speech. While of course Spanish is not a tonal language, it was nonetheless unfamiliar to the subjects in this study. Perhaps here the subjects did not process the Spanish sentences in a language mode of processing given that the language was unfamiliar. Maybe they sought recourse to a musical mode of processing which might have allowed them to consider the sentences in their totality

rather than as individual words. Should this be the case it would allow for the correlation between Spanish and Music (Prod) and the absence of a correlation between Spanish and Language (Recep) or between Spanish and Korean.
Another way to consider the validity of the language tests in Phase II is to investigate why there is no association between the receptive language test in Phase I and the productive language test in Phase II. It was noted above that there is no apparent linear relationship between the two. This is shown in Figure 52.

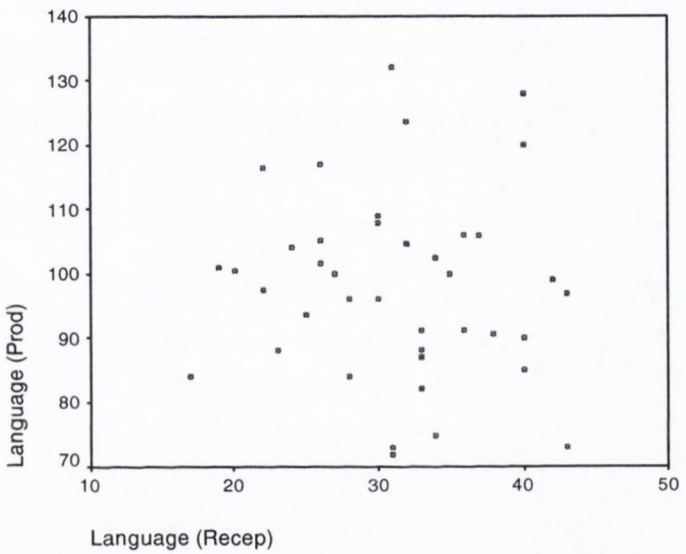


Figure 52. Language (Prod) and Language (Recep) scores

This is quite worrying as it was believed that the receptive test of Phase I was a valid test of language aptitude as it is based on the Modern Language Aptitude Test. Similarly the productive test was believed to test language aptitude as it tests phonological Working Memory which is a good indicator of language aptitude. One suggestion is that both tests do in fact test language aptitude but different aspects of it, i.e. the receptive test should test Carroll's factors of language aptitude (particularly Phonetic Coding Ability, Grammatical Sensitivity and Rote learning ability. Inductive Language Learning Ability is only reflected weakly in the MLAT). The productive test measures phonological Working Memory. Indeed Carroll himself (Carroll, 1990) suggests that it is necessary to update the MLAT with some type of memory test. If this were to be the case, it might be expected that while the tests do not correlate with each other, both may correlate with the eventual outcome of the language learning process.

It is perhaps surprising that both Language (Recep) and Language (Prod) should show similar correlations with Raven as they do not relate to each other. While significant correlations between Raven and Language (Prod) were not shown in this study (see *Section 5.3.3.2*), it was shown above that this may be related to a lack of power and that in fact the difference in correlation between Language (Recep) and Raven and that between Language (Prod) and Raven is not significant. The correlation between Language (Prod) and Raven has been accounted for on the grounds that gF intelligence has been shown to correlate with Working Memory ability. The correlation between Language (Recep) and Raven arises primarily because of the correlation between the grammar component of the test and Raven, $r=0.41$, $p<0.01$. Given that the definition of what Raven is supposed to measure given in Raven *et al.* (1998) is "the ability to forge new insights, ..., and the ability to identify relationships", it is not surprising that there should be an overlap between Raven and Grammar. Hence, the correlation between Raven and Language (Recep) can be accounted for. Thus, it can be seen how Raven correlates with both Language (Recep) and Language (Prod) even though they do not show a significant relationship to each other.

5.7 Conclusion

This chapter presented the results of the tests which were carried out as described in Chapter 4. Many interesting findings emerged including the presence of a significant relationship between music and language. This relationship holds even when non-verbal intelligence is taken into account. A more detailed summary and conclusion is presented in Chapter 6. Some suggestions for future work are also presented in this context.

Chapter 6 Conclusion

6.1 Introduction

The aim of this chapter is to draw together the findings of the previous chapter and to relate them to the earlier chapters which dealt with previous investigations of the relationship between music and language aptitude. An attempt is made to highlight new methodologies employed in this study and new findings which emerged. I also hope to demonstrate where these findings depart from existing findings in the literature and where they concur.

6.2 Summary of research

Phase I examined receptive skills in language and music aptitude and their relationship to general fluid intelligence. The test used in the music domain, i.e. the Bentley Measures of Musical Abilities, is a standardised test, relatively widely used in the field of music aptitude testing. The language aptitude test is based on the Modern Language Aptitude Test but is not a standardised test as a standardised test was not available. General fluid intelligence is tested using the Raven Progressive Matrices test – a very widely used and respected measure in educational settings. The gender of participants was also recorded.

Findings of Phase I point to a significant correlation between musical and linguistic aptitude, which exists even when the effects of non-verbal intelligence are taken into account. Musical aptitude, non-verbal intelligence and gender are all significant predictors of language aptitude scores, with females who are high in musical aptitude and non-verbal intelligence receiving the highest language aptitude scores. These results were found in the age group sampled, which was 12 years 1 month to 16 years 6 months.

Phase II investigated productive skills in language and music and considered these in relation to each other and in relation to gF intelligence. Productive skills in language were tested by asking subjects to imitate words in Korean and short sentences in Spanish. Productive skills in music were examined by asking subjects

to imitate rhythm patterns taken from Irish folksongs. Raven scores were available from Phase I. The language tests were scored by native speakers. This resulted in a subjective score. An objective score was also calculated in the form of the Pairwise Variability Index, which is a measure of language rhythm.

Results show a moderate and significant relationship between productive skills in language and music. Fluid intelligence is not as influential in this case as in Phase I, although the lower power of the second part of the study may be one contributory factor. It may of course also be related to the fact that the tests in Phase II were less typical of normal classroom activities which are likely to require gF intelligence. In addition, the tests in Phase II required motor skills which are unlikely to be closely related to fluid intelligence.

Phase III considered receptive skills in music and language and their relationship to productive skills in music and language. There is a moderate and significant correlation between receptive and productive skills in music. There is no correlation between receptive and productive skills in language. One suggestion to account for this is that the receptive language test involves primarily analytical skills and memory skills whereas the productive language test involves greater memory ability and the ability to actually produce foreign phonemes.

6.3 New approaches and findings

One of the strengths of this study is that intelligence and gender are taken into account. Previous studies (e.g. Koren, 1993; Tucker, 2000) have investigated the relationship between music and language aptitude but have generally not given serious consideration to the suggestion that intelligence may have a confounding effect on the relationship between the two. Here it is found that gF intelligence interacts with music and language but that the relationship between the two is still significant, albeit small to moderate, when the effects of intelligence are partialled out.

It is perhaps not surprising that gF intelligence is particularly important in Phase I of this study. General fluid intelligence is closely related to 'g', the general factor

of intelligence. If indeed 'g' is important for academic tasks, it is likely that it is important for Phase I, as tests were administered in a classroom setting and are similar to the type of schoolwork that subjects would often do, particularly in the case of the language aptitude test. Certainly the grammar component of the Language (Recep) test is closely related to common classroom activities.

A novel approach of measuring language aptitude was taken in Phase II. Not only were subjects asked to repeat foreign words, they were also asked to imitate short sentences. The idea of testing ability to imitate foreign language rhythm is new to the literature. Here it is shown that the ability to imitate Spanish rhythm as judged by a native speaker is quite closely related to the ability to imitate sounds of the language. Both pronunciation and rhythm are related to the ability to reproduce musical patterns.

It was interesting to note that raters generally showed quite high levels of agreement when awarding scores on the productive language test. This suggests that it is indeed possible to score such a test reliably in a subjective way. Therefore, there must be certain elements which native speakers consciously or subconsciously attend to when listening to speech in order to decide how native-like it is.

The idea of scoring language productions objectively for rhythm using PVI measures is also a new approach. While the PVI measures may not have correlated with native speaker judgements of rhythm, they present a new objective approach to assessing rhythm which may prove useful in the future. This approach might be more suitable with advanced learners rather than with complete beginners. While the full potential of PVI was not reached here, further studies may be more successful in exploiting the benefits of this objective measure.

A productive music aptitude test was developed in the course of this study. It is noteworthy that it showed a good relationship to the music receptive test. This demonstrates that both tests measure some similar underlying construct. The productive music test therefore appears to be a useful way to measure musical aptitude.

Another interesting finding to emerge is the suggestion of a difference between the Korean and Spanish tests of language aptitude. It is too early to say if this difference is due to properties of the tests themselves or if in fact there may actually be a difference in attempting to imitate one language over another. One possible indication that an actual difference between the two languages may exist comes from the work of Young-Won Kim (2005) who suggests that people cannot hear a foreign language properly without experience of articulating the foreign language using "good speaking posture of the language". Unfortunately, Kim's book is only available in Korean, with only a brief summary made available in English.[27] Speaking postures appear to be related to place of articulation. Kim proposes that there is a big difference in the speaking postures of Korean and English. From the brief summary available in English, it appears to be the case that the speaking postures of Korean and English show greater diversity than the speaking postures of English and other European languages. Kim notes that Americans can easily shift their speaking postures to imitate Korean even without special training, but that Koreans do not recognise the possibility of changing their speaking positions for English without special training. While of course the subjects in this study were native speakers of English attempting to imitate Korean, this issue may be relevant to the findings of this study and would seem worthy of further consideration.

It was found that the language aptitude test used in Phase I, i.e. the receptive test, shows no correlation with the productive test used in Phase II. One possible suggestion to account for this is the lack of emphasis placed on Working Memory ability in the receptive test.

A Chinese test was devised as part of Phase I in order to investigate if subjects could discriminate between lexical tones. This was a new addition to the receptive test but it proved very difficult for the subjects in this study. This is perhaps not surprising as others have shown that native speakers of English place two syllables differing in only lexical tone into the same category. Other points of convergence with the existing literature are presented in the following section.

---

[27] See http://www.academici.com/cgi-bin/forum.fpl?op=showforum&id=7681 (last verified 26.09.2005).

6.4 Relationship to existing literature

The findings of Phase I that receptive skills in language aptitude are related to receptive skills in musical aptitude have been previously noted in the literature. A number of these studies were described in Chapter 3. The emergence of rhythm as the component of musical aptitude to be particularly related to language aptitude was noted in this study and has also been previously documented. Again this issue was highlighted in Chapter 3. It was noted there that rhythm is a candidate for common processing in music and language given that both music and language involve left hemisphere processing. Empirical studies such as that by Douglas and Willatts (1994) have shown evidence to this effect.

It is interesting to note that two previous studies highlighted the link between Spanish and musical aptitude which is what also emerged in this study. Dexter and Omwake (1934) found that ability to discriminate pitch was more important in predicting Spanish scores than German scores. Eterno (1961) pointed to a link between musical aptitude and Spanish grades after a course of instruction. He did not investigate abilities in other languages. The current study found Spanish to be related to musical aptitude to a greater extent than Korean, although the difference between the two did not reach statistical significance. Further work is necessary to investigate why this difference may hold or if indeed a greater relationship holds between Spanish and music than between music and other languages.

Dividing the sample on the basis of Raven scores provided some evidence in favour of the Ability Differentiation Hypothesis. This hypothesis proposes that in individuals of lower intelligence, a general factor of intelligence is more pervasive, with the result that different abilities are more closely related than in individuals with higher levels of intelligence. While results here were not significant, there was a trend for results to be in the direction of this hypothesis.

One point of departure from the existing literature was the finding that all components of the music test in Phase I were correlated. The author of the Bentley Measures of Musical Abilities, Arnold Bentley, had found that rhythmic ability was somewhat distinct from the other musical skills.

6.5 Limitations of this study

This study was conducted by an individual researcher and as such was subject to certain constraints. A convenient sample was chosen rather than a purely random sample. Testing conditions in Phase I were not identical for all subjects as some subjects took all three tests on one day whereas others took the tests over a two day period. It was not possible to standardise tests on a large group prior to use. However, in other respects the study was conducted in the most rigorous manner possible. It is hoped therefore that these limitations do not impact too severely on the results.

6.6 Suggestions for future work

It would be very interesting to consider in more detail the Spanish test used in this study. For example, does the ability to repeat short sentences predict later language learning success? Or, is it possible that subjects are processing Spanish in some way more similar to music processing than language processing, given that it is an unfamiliar language? It is important to remember that in this study, there is no statistically significant difference between the correlations found between Spanish and music and between Korean and music, although Spanish appears more closely related to music than Korean, i.e. where Spanish shows a large significant correlation with music, the correlation between Korean and music is smaller and insignificant. However, a closer examination revealed that there is actually no statistically significant difference between the two correlations. Therefore, it is important not to attach too much importance to this difference at this stage. Further work is necessary to determine if this is a true effect or simply a statistical artefact.

The relationship between PVI measures and native speaker scores is worthy of further investigation. It would be useful to administer this test to intermediate/advanced language learners in order to investigate the rhythmic properties of their productions. Advanced language learners would not have the same difficulties in repeating the sounds and words as the subjects in this study. Therefore, it could be hoped that they would manage to repeat whole sentences or that they could even be asked to read a text. Their productions could then be segmented and PVI measures calculated and compared to native speaker ratings. If

there is still no relationship between the two, it must be asked what it is that native speakers judge when asked to consider rhythm.

The finding that music and language are consistently related over and above the influence of intelligence could be of interest to educators. It is possible that new teaching approaches involving music would be particularly suited to those with high musical aptitude. Perhaps the musical aptitude of students could be better exploited than it is today. Smith Salcedo (2002) presents an interesting review of language teaching approaches which exploit music and song. She mentions numerous books and courses in Spanish which teach through song. It would be interesting to investigate how the musical aptitude of students may influence their attitudes towards, and success in, such approaches to learning.

Given the suggestion in the literature that musical training may be useful for language learning and also the idea presented here that even musical aptitude in the absence of training may be beneficial, it would be interesting to consider if the benefits of musical training on language learning depend on the musical aptitude of the subjects, i.e. if the subject has very poor musical aptitude, will he/she derive any extra musical benefits from musical training?

Another possible avenue which could be explored in the future is the contrast between the music-language relationship and the relationship between music and other cognitive abilities. Much work has already been done in investigating the relationship between music and spatial reasoning. It would be interesting to consider whether or not there is a significant difference in the magnitude of the different relationships. This would give further insight into the possibility that the music-language relationship is of particular importance over and above the influence of Working Memory or general intelligence.

Finally, it is hoped that future work in the area of music-language research might employ a reading span test, in order to investigate Working Memory capacity, and a standardised test of Phonological Working Memory span. Results of these tests could be considered in relation to music and language aptitude test scores in order

to more fully determine the role of Working Memory in accounting for correlations between the two.

6.7 Summary

This study shows a moderate to large correlation between music and language aptitude in both receptive and productive domains. Phases I and II of the study are compatible with an underlying correlation of between r=0.22 and r=0.50 between music and language aptitude. Part of this correlation is likely due to the influence of intelligence and Working Memory abilities. However, this issue is addressed in the calculation of partial correlations where it was shown that the correlation remains significant, although somewhat smaller, when intelligence is factored out. The study therefore highlights the fact that a link between music and language aptitude does indeed exist. It is not simply training in music which has an effect on language development but even basic aptitude in music may be helpful for the language learner.

Bibliography

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Abrahamsson, N., & Hyltenstam, K. (2004). The role of language aptitude for near-native L2 proficiency, *EUROSLA*. Bilbao, Spain.

Adamson, L., Tomasello, M., & Benbisty, L. (1984). *An "expressive" infant's communication development.* Paper presented at the International Conference on Infant Studies, New York.

Aitchison, J. (1998). *The articulate mammal* (4th ed.). London: Routledge.

Altman, D. G. (1999). *Practical statistics for medical research*. London: Chapman & Hall/CRC.

Altman, H. (1980). Foreign language teaching: Focus on the learner. In H. Altman & C. V. James (Eds.), *Foreign language teaching: Meeting individuals' needs*. Oxford: Pergamon.

Ardila, A. (2003). Language representation and working memory with bilinguals. *Journal of Communication Disorders, 36*, 233 - 240.

Ardila, A., Rosselli, M., & Puente, A. (1994). *Neuropsychological evaluation of the Spanish speaker*. New York: Plenum Press.

Artés, J. S. (1985). *Historias para conversar*. Madrid: Sociedad General Española de Librería, S. A.

Astésano, C., Besson, M., & Alter, K. (2004). Brain potentials during semantic and prosodic processing in French. *Cognitive Brain Research, 18*, 172 - 184.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 89 - 195). New York: Academic Press.

Avenell, A., Campbell, M. K., Hannaford, P. C., Kilonzo, M. M., McNeill, G., Milne, A. C., et al. (2005). Effect of multivitamin and multimineral supplements on morbidity from infections in older people (MAVIS trial): Pragmatic, randomised, double blind, placebo controlled trial. *British Medical Journal, 331*, 324 - 329.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*, 417 - 423.

Baddeley, A. D. (2003). Working memory and language: An overview. *Journal of Communication Disorders, 36*(3), 189 - 208.

Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*(1), 158 - 173.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47 - 89). New York: Academic Press.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory - mechanisms of active maintenance and executive control* (pp. 28 - 61). Cambridge: Cambridge University Press.

Baddeley, A. D., Papagno, C., & Vallar, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory and Language, 27*, 586 - 595.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of STM. *Journal of Verbal Learning and Verbal Behaviour, 14*, 575 - 589.

Balkwill, L.-L., & Thompson, W. F. (1999). A cross-cultural investigation of the perception of emotion in music: Psychophysical and cultural cues. *Music Perception, 17*(1), 43 - 64.

Barber, E. (1980). Language acquisition and applied linguistics. *ADFL Bull., 12*, 26-32.

Bates, E., Bretherton, I., & Snyder, L. (1988). *From first words to grammar: Individual differences and dissociable mechanisms*. New York: Academic Press.

Benson, D. F., Dobkin, B., Gonzalez-Roth, L. J., Helman-Estabrook, N., & Kertesz, A. (1994). Assessment: Melodic intonation therapy. *Neurology, 44*, 566 - 568.

Bentley, A. (1966a). Measures of music abilities. Berkshire: NFER-Nelson Publishing Company.

Bentley, A. (1966b). *Musical ability in children*. London: Harrap & Co.

Berz, W. L. (1995). Working memory in music: A theoretical model. *Music Perception, 12*(3), 353 - 364.

Besson, M., & Schön, D. (2003). Comparison between language and music. In I. Peretz & R. J. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 269 - 293). New York: Oxford University Press.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research* (pp. 171 - 206). Timonium, MD: York Press Inc.

Bever, T., & Chiarello, R. (1974). Cerebral dominance in musicians and non-musicians. *Science, 185*, 537 - 539.

Beye, C. R. (1972). *The Iliad, the Odyssey and the epic tradition*. Gloucester, Mass.: Peter Smith.

Bialystok, E. (1981). The role of conscious strategies in second language proficiency. *Modern Language Journal, 65*, 24 - 35.

Bialystok, E. (2004). Bilingualism: Cognitive effects across the lifespan, *EUROSLA*. San Sebastian, Spain.

Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging, 19*, 290 - 303.

Bialystok, E., & Martin, M. M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental Science, 7*, 325 - 339.

Binet, A., & Simon, T. (1905a). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *L'Année Psychologique, 11*, 191 - 244.

Binet, A., & Simon, T. (1905b). Sur la nécessité d'établir un diagnostique scientifique des états inférieurs de l'intelligence. *Année Psychologique, 11*, 163 - 190.

Blacking, J. (1973). *How musical is man?* Seattle: University of Washinton Press.

Blecha, P. (2004). *Taboo tunes: A history of banned bands & censored songs*. San Francisco, CA: Backbeat Books.

Blickenstaff, C. B. (1963). Musical talents and foreign language learning ability. *Modern Language Journal, 47*, 359 - 363.

Block, R. A., & Grosfield, G. A. (2000). The Mozart effect: A meta-analytic review of musically enhanced spatial ability: Manuscript submitted for publication.

Blonder, L. X., Bowers, D., & Heilman, K. M. (1991). The role of the right hemisphere in emotional communication. *Brain and Cognition, 114*, 1115 - 1127.

Bod, R. (2002). A unified model of structural organization in language and music. *Journal of Artificial Intelligence Research, 17*, 289 - 308.

Boersma, P., & Weenink, D. (1992 - 2003). Praat (Version 4.1.14). SIL Encore Fonts: Summer Institute of Linguistics.

Bongaerts, T., Planken, B., & Schils, E. (1995). Can late learners attain a native accent in a foreign language? A test of the critical period hypothesis. In D. Singleton & Z. Lengyel (Eds.), *The age factor in second language acquisition* (pp. 30-50). Clevedon: Multilingual Matters.

Bradley, L., & Bryant, P. (1983). Categorising sounds and learning to read: A causal connection. *Nature, 301*, 419 - 421.

Bradshaw, A. H. (1988). *Pentatonic folksongs - folk music for Irish wee folk.* Dublin: Bradshaw Music Education.

Bradshaw, E., & McHenry, M. A. (in press). Pitch discrimination and pitch matching abilities of adults who sing inaccurately. *Journal of Voice.*

Brain facts. (13 June 2005). Retrieved 13.06.2005, from http://apu.sfn.org/content/Publications/BrainFacts/brainfacts.pdf

Broca, P. (1861). Remarques sur le siège de la faculté de langage articulé, suivis d'une observation d'aphémie (perte de la parole). *Bulletins de la Societé Anatomique, 6*, 330 - 357.

Brody, N. (1992). *Intelligence* (2nd ed.). San Diego, CA: Academic Press.

Bruner, J. S., Goodnow, J., & Austin, G. A. (1956). *A study of thinking*. New York: Wiley.

Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Attapaiboon, C., Lacerda, F., et al. (1996). Perception of lexical tone across languages: Evidence for a linguistic mode of processing, *4th International Conference on Spoken Language Processing*. Philadelphia, PA.

Butt, T. (2004). *Understanding people*. New York: Palgrave Macmillan.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioural and Brain Sciences, 22*(1), 77 -94.

Carroll, J. B. (1962). The prediction of success in intensive foreign language training. In R. Glaser (Ed.), *Training research & education* (pp. 87 - 136). Pittsburgh: University of Pittsburgh Press.

Carroll, J. B. (1973). Implications of aptitude test research and psycholinguistic theory for foreign language teaching. *International Journal of Psycholinguistics, 2*, 5 - 14.

Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. In K. Diller (Ed.), *Individual differences and universals in language learning*. Rowley Mass.: Newbury House.

Carroll, J. B. (1990). Cognitive abilities in foreign language aptitude: Then and now. In T. S. Parry & C. W. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 11 - 29). New Jersey: Prentice Hall Regents.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*: Cambridge University Press.

Carroll, J. B., & Sapon, S. (1955). *Modern language aptitude test*. New York: The Psychological Corporation.

Carstairs-McCarthy, A. (1999). *The origins of complex language: An inquiry into the evolutionary beginnings of sentences, syllables, and truth*. Oxford: Oxford University Press.

Carstens, C. B., Huskins, E., & Hounshell, G. W. (1995). Listening to Mozart may not enhance performance on the revised Minnesota paper form board test. *Psychological Reports, 77*, 111 - 114.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40*, 153 - 193.

Cattell, R. B. (1965). *The scientific analysis of personality*. Harmondsworth: Penguin.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston: Houghton Mifflin.

Chan, A. S., Ho, Y.-C., & Cheung, M.-C. (1998). Music training improves verbal memory. *Nature, 396*, 128.

Chapelle, C. (1995). Field-dependence/field-independence in the second language classroom. In J. Reid (Ed.), *Learning styles in the ESL/EFL classroom*. Boston: Heinle and Heinle.

Chomsky, N. (1964). *Current issues in linguistic theory*. The Hague: Mouton.

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

Colom, R., Rebollo, I., Palacios, A., Juan-Espinosa, M., & Kyllonen, P. C. (2004). Working memory is (almost) perfectly predicted by g. *Intelligence, 32*, 277 - 296.

Cook, V. (1996). *Second language learning and language teaching* (2nd ed.). London: Arnold.

Cooper, C. (1998). *Individual differences*. New York: Arnold.

Cooper, C. (1999). *Intelligence and abilities*. London: Routledge.

Copland, A. (1957). *What to listen for in music*. New York: McGraw-Hill Book Company.

Costa-Giomi, E. (1997). The McGill piano project: Effects of piano instruction on children's cognitive abilities, *Third Triennial ESCOM conference*. Uppsala, Sweden.

Costa-Giomi, E. (1999). The effects of three years of piano instruction on children's cognitive development. *Journal of Research in Music Education, 47*(5), 198 - 212.

Costa, P., & McCrae, R. (1985). *The NEO personality inventory manual*. Odessa, FL.: Psychology Assessment Resources.

Costa, P., & McCrae, R. (1992). *Revised NEO personality inventory: NEO PI and NEO five-factor inventory (NEO FFI) professional manual*. Odessa, FL.: Psychological Assessment Resources.

Coulter, G. R., & Anderson, S. R. (1993). Introduction. In *Phonetics and phonology: Current issues in ASL phonology*. San Diego, CA: Academic Press.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin, 104*, 163 - 191.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory - mechanisms of*

active maintenance and executive control (pp. 62 - 101). Cambridge: Cambridge University Press.

Cowan, N., & Kail, R. (1996). Covert processes and their development in short-term memory. In S. Gathercole (Ed.), *Models of short-term memory* (pp. 29 - 50). Hove, England: Psychology Press.

Crocker, T. (1999). *Intelligence*. Milton Keynes, UK.: The National Association for Gifted Children.

Cummins, J. (1979). Cognitive/academic language proficiency, linguistic interdependence, the optimal age question and some other matters. *Working papers on bilingualism, 19*, 197-205.

Curtiss, S. (1988). Abnormal language acquisition and modularity of language. In F. J. Newmeyer (Ed.), *Linguistics: The Cambridge survey: Vol. 2. Linguistic theory: Extensions and implications* (Vol. 2, pp. 90-116). Cambridge: Cambridge University Press.

Cutler, A., Mehler, J., Norris, D. G., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language, 25*, 385 - 400.

Dale, H. C. A., & Baddeley, A. D. (1966). Remembering a list of two-digit numbers. *Quarterly Journal of Experimental Psychology, 18*, 212 - 219.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behaviour, 19*, 450 - 466.

Daneman, M., & Carpenter, P. A. (1983). Individual differences in integrating information between and within sentences. *Journal of Experimental Psychology. Learning, 9*, 561 - 584.

Darwin, C. (1871). *The descent of man and selection in relation to sex*. London: John Murray.

Dauer, R. M. (1983). Stress-timing and syllable timing re-analysed. *Journal of Phonetics, 11*, 51 - 62.

Dauer, R. M. (1987). *Phonetic and phonological components of language rhythm.* Paper presented at the XIth International Congress of Phonetic Sciences, Talinn, Estonia.

Deary, I. J. (2001). *Intelligence - a very short introduction*. New York: Oxford University Press.

Deary, I. J., Egan, V., Gibson, G. J., Austin, E. J., Brand, C. R., & Kellaghan, T. (1996). Intelligence and the differentiation hypothesis. *Intelligence, 23*, 105 - 132.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition, 22*, 499 - 533.

DeVellis, R. F. (1991). *Scale development: Theory and applications*. Newbury Park, CA: Sage.

Devitt, A. (2005). *Methods for meaningful text representation and comparison.* Unpublished Ph.D. thesis, Trinity College, Dublin.

Dewaele, J.-M. (2001). Tongue-tied in a foreign language. *The Linguist, 40*(5), 153 - 155.

Dexter, E. S., & Omwake, K. T. (1934). The relation between pitch discrimination and accent in modern languages. *Journal of Applied Psychology, 18*, 267 - 271.

Dinklage, K. (1971). Inability to learn a foreign language. In G. Blain & C. McArthur (Eds.), *Emotional problems of the student*. New York: Appleton-Century-Crofts.

Dörnyei, Z. (1994). Motivation and motivating in the foreign language classroom. *Modern Language Journal, 78*, 273 - 284.

Dörnyei, Z. (2002). The motivational basis of language learning tasks. In P. Robinson (Ed.), *Individual differences and instructed language learning* (Vol. 2, pp. 137 - 158). Amsterdam/Philadelphia: John Benjamins Publishing Co.

Dörnyei, Z., & Skehan, P. (2003). Individual differences in second language learning. In C. J. Doughty & M. H. Long (Eds.), *The handbook of second language acquisition* (pp. 589 - 630). Oxford: Blackwell.

Douglas, S., & Willatts, P. (1994). The relationship between musical ability and literacy skills. *Journal of Research in Reading, 17*(2), 99 - 107.

Ehrman, M. E. (1997). Field independence, field dependence, and field sensitivity. In J. Reid (Ed.), *Understanding learning styles in the second language classroom* (pp. 62 - 70). Englewood, NJ.: Regents Prentice Hall.

Ehrman, M. E., & Leaver, B. L. (2003). Cognitive styles in the service of language learning. *System, 31*, 393 - 415.

Ehrman, M. E., Leaver, B. L., & Oxford, R. L. (2003). A brief overview of individual differences in second language learning. *System, 31*, 313 - 330.

Ehrman, M. E., & Oxford, R. L. (1990). Adult language learning styles and strategies in an intensive training setting. *Modern Language Journal, 74*, 311 - 327.

Eisenstein, M. (1982). A study of social variation in adult second language acquisition. *Language Learning, 32*, 367 - 391.

Ekstrand, L. (1977). *Social and individual frame factors in L2 learning: Comparative aspects.* Paper presented at the Papers from the First Nordic Conference on Bilingualism, Universitet Helsingfors.

Elbert, T., Pantev, C., Wienbruch, C., Rockstrub, B., & Taub, E. (1995). Increased cortical representation of the fingers of the left hand in string players. *Science, 270*, 305 - 307.

Ellis, N., & Beaton, A. (1993). Factors affecting the learning of foreign language vocabulary: Imagery keyword mediators and phonological short-term memory. *Quarterly Journal of Experimental Psychology, 46A*, 533 - 558.

Ellis, N. C. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 33 - 68). Cambridge: Cambridge University Press.

Ellis, R. (1994). *The study of second language acquisition.* Oxford: Oxford University Press.

Emmorey, K., Kosslyn, S. M., & Bellugi, U. (1993). Visual imagery and visual-spatial language: Enhanced imagery abilities in deaf and hearing ASL signers. *Cognition, 46*, 139 - 181.

Engle, R. W., Kane, M. J., & Tuholski, S. W. (1999a). Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In A. Miyake & P. Shah (Eds.), *Models of working memory - mechanisms of active maintenance and executive control* (pp. 102 - 134). Cambridge: Cambridge University Press.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999b). Working memory, short-term memory, and general fluid intelligence: A latent variable approach. *Journal of Experimental Psychology. General, 128*, 309 - 331.

Engler, B. (2003). *Personality theories* (6th ed.). Boston/New York: Houghton Mifflin Company.

Erdelyi, M. H. (1985). *Psychoanalysis: Freud's cognitive psychology*. New York, NY.: Freeman.

Eterno, J. A. (1961). Foreign language pronunciation and musical aptitude. *Modern Language Journal, 45*, 168 - 170.

Evans, W. E., & Bastian, J. (1969). Marine mammal communication: Social and ecological factors. In H. T. Andersen (Ed.), *The biology of marine mammals*. New York: Academic Press.

Ewers, D. W. F. (1950). Relationship between auditory abilities and reading abilities: A problem in psychometrics. *Journal of Experimental Education, 18*, 239 - 263.

Eysenck, H. (1973). Personality, learning and anxiety. In H. Eysenck (Ed.), *Handbook of abnormal psychology* (pp. 390 - 419). London: Pitman.

Eysenck, H., & Levey, A. (1972). Conditioning, introversion-extroversion and the strength of the nervous system. In V. Nebylitsyn & J. Gray (Eds.), *Biological basis of individual behaviour* (pp. 206 - 220). New York: Academic Press.

Eysenck, H. J., & Eysenck, S. B. G. (1976). *Psychoticism as a dimension of personality*. New York: Crane, Russak.

Farhady, H. (1982). Measures of language proficiency from the learner's perspective. *TESOL Quarterly, 16*, 43 - 59.

Feld, S., & Keil, C. (1994). *Music grooves: Essays and dialogues*. Chicago: Chicago University Press.

Fisher, S., & Greenberg, R. P. (1996). *Freud scientifically appraised*. New York, NY.: John Wiley.

Fodor, J. (1983). *The modularity of the mind*. Cambridge, MA.: M.I.T. Press.

Fox, L. H. (1981). Identification of the academically gifted. *American Psychologist, 36*, 1103 - 1111.

Frackowiak, R. S. (1994). Frontal mapping of verbal memory and language. *Trends in Neuroscience, 17*, 109 - 115.

Fraisse, P. (1984). Perception and estimation of time. *Annual Review of Psychology, 35*, 1 - 36.

Freud, S. (1917/1964). *Introductory lectures on psychoanalysis*. Harmondsworth: Penguin.

Freud, S. (1932). *New introductory lectures on psychoanalysis*. Harmondsworth: Penguin.

Freud, S. (1957). *Five lectures on psychoanalysis*. Harmondsworth: Penguin.

Fries, W., & Swihart, A. A. (1990). Disturbance of rhythm sense following right hemisphere damage. *Neuropsychologia, 28*(12), 1317 - 1323.

Fromkin, V. A. (1997). Some thoughts about the brain/mind/language interface. *Lingua, 100*, 3 - 27.

Gall, F. J. (1791). *Philosophisch-medicinische untersuchungen*. Vienna: Graffer.

Gall, F. J., & Spurzheim, J. C. (1810). *Anatomie et physiologie du système nerveux*. Paris: Schoell.

Ganschow, L., & Sparks, R. L. (1986). Learning disabilities and foreign language difficulties: Deficit in listening skills? *Journal of Reading, Writing, and Learning Disabilities International, 2*, 116 - 123.

Garcia, T., & Pintrich, P. (1995). Assessing students' motivation and learning strategies: The motivated strategies for learning questionnaire, *Annual Meeting of the American Educational Research Association*. San Francisco.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Gardner, H. (1993). *Multiple intelligences: The theory in practice*. New York: Basic Books.

Gardner, H. (1997). *Is musical intelligence special?* Paper presented at the Ithaca Conference '96: Music as Intelligence, Ithaca, New York.

Gardner, H. (1999). Are there additional intelligences? The case for naturalistic, spiritual, and existential intelligences. In J. Kane (Ed.), *Education, information, and transformation* (pp. 111 - 131). Upper Saddle River, NJ.: Prentice Hall.

Gardner, K. (1997). *Sounding the inner landscape: Music as medicine*. Rockport, Mass.: Shaftesbury.

Gardner, R. C. (2001). Integrative motivation: Past, present and future, *Distinguished Lecturer Series, Tokyo*. Temple University Japan.

Gardner, R. C., & Lambert, W. E. (1965). *Attitudes and motivation in second-language learning*. Rowley, MA.: Newbury House.

Gardner, R. C., & MacIntyre, P. D. (1992). A student's contribution to second language learning: Part I, cognitive factors. *Language Teaching, 25*, 211 - 220.

Gass, S., & Selinker, L. (1994). *Second language acquisition: An introductory course*. Hillsdale: Erlbaum.

Gathercole, S., & Adams, A. (1993). Phonological working memory in very young children. *Developmental Psychology, 29*, 770 - 778.

Gathercole, S., & Adams, A. (1994). Children's phonological working memory: Contributions of long-term knowledge and rehearsal. *Journal of Memory and Language, 33*, 672 - 688.

Gathercole, S., & Baddeley, A. D. (1989). Development of vocabulary in children and short-term phonological memory. *Journal of Memory and Language, 28*, 200 - 213.

Gathercole, S., & Hitch, G. J. (1993). Developmental changes in short-term memory: A revised working memory perspective. In A. Collins, S. E. Gathercole, M. A. Conway & P. E. Morris (Eds.), *Theories of memory* (pp. 189 - 210). Hillsdale, NJ: Erlbaum.

Gathercole, S., Hitch, G. J., Service, E., & Martin, A. J. (1997). Phonological short-term memory and new word learning in children. *Developmental Psychology, 33*, 966 - 979.

Gathercole, S., Willis, C., Emslie, H., & Baddeley, A. D. (1992). Phonological memory and vocabulary development during the early school years: A longitudinal study. *Developmental Psychology, 28*, 887 - 898.

Gathercole, S., Willis, C. S., & Baddeley, A. D. (1991). Differentiating phonological memory and awareness of rhyme: Reading and vocabulary development in children. *British Journal of Psychology, 82*, 387 - 406.

Gedamke, J., Costa, D. P., & Dunstan, A. (2001). Localization and visual verification of a complex minke whale vocalization. *Journal of the Acoustical Society of America, 109*(6), 3038 - 3047.

Genesse, F. (1976). The role of intelligence in second language learning. *Language Learning, 26*, 267-280.

Geva, E., & Ryan, E. (1993). Linguistic and cognitive correlates of academic skills in first and second languages. *Language Learning, 43*, 5 - 42.

Ghazanfar, A. A., & Logothetis, N. K. (2003). Facial expressions linked to monkey calls. *Nature, 423*, 937 - 938.

Goldberg, L. R. (1981). Language and individual differences: The search for universals in personality lexicons. In L. R. Wheeler (Ed.), *Review of personality and social psychology* (Vol. 2, pp. 141 - 165). Beverly Hills, CA.: Sage.

Goleman, D. (1991, December 24th). Hope emerges as key to success in life. *New York Times,* p. C1.

Goleman, D. (1996). *Emotional intelligence*. London: Bloomsbury.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In *Papers in laboratory phonology* (Vol. 7): Mouton.

Great Books of the Western World. (Vol. 7)(1987). Chicago: Encyclopedia Britannica.

Gregersen, T., & Horwitz, E. K. (2002). Language learning and perfectionism: Anxious and non-anxious language learners' reactions to their own oral performance. *The Modern Language Journal, 86*(4), 562 - 570.

Grigorenko, E. L. (2002). Foreign language acquisition and language-based learning disabilities. In P. Robinson (Ed.), *Individual differences and instructed language learning* (Vol. 2, pp. 95 - 112). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Grigorenko, E. L., Sternberg, R. J., & Ehrman, M. E. (2000). A theory-based approach to the measurement of foreign language learning ability: The canal-f theory and test. *The Modern Language Journal, 84*(3), 390 - 405.

Grimshaw, G. M., Adelstein, A., Bryden, M. P., & MacKinnon, G. E. (1998). First-language acquisition in adolescence: Evidence for a critical period for verbal language development. *Brain and Language, 63*(2), 237-255.

Grogan, S. C. (1995). Which cognitive abilities at age four are the best predictors of reading ability at age seven? *Journal of Research in Reading, 18*(1), 24 - 31.

Grove, W. M., Andreasen, N. C., McDonald-Scott, P., Keller, M. B., & Shapiro, R. W. (1981). Reliability studies of psychiatric diagnosis. *Archives of General Psychiatry, 38*, 408 - 413.

Grubb, W. L., & McDaniel, M. A. (2000). Working memory, short-term memory and intelligence: A meta-analysis, *First Annual Conference of the International Society for Intelligence Research*. Cleveland, OH.

Gruhn, W. (1997). Neurobiological foundations and educational implications. *Research Studies in Music Education, 9*, 36 - 47.

Gruhn, W. (1998). Respondent 6: Responses to Katie Overy's paper "can music really 'improve' the mind?" *Psychology of Music, 26*, 208 - 210.

Gruhn, W., & Rauscher, F. H. (2002). The neurobiology of music cognition and learning. In R. Colwell & C. Richardson (Eds.), *Second handbook on music teaching and learning* (pp. 445 - 460). New York: Oxford University Press.

Gustafsson, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8*, 179 - 203.

Hahn, M., & Hwang, I. (1999). Effects of tempo and familiarity of background music on message processing in TV advertising: A resource-matching perspective. *Psychology & Marketing, 16*(8), 659 - 675.

Hallam, S. (2000). The effects of listening to music on children's spatial task performance. *British Psychological Society Education Review, 25*(2), 22 - 26.

Harley, B., & Hart, D. (1997). Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition, 19*, 379 - 400.

Harrington, M., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skill. *Studies in Second Language Acquisition, 14*, 25 - 38.

Hatch, E. (1974). Second language learning - universals? *Working papers on bilingualism, 3*, 1 - 17.

Hauser, M. D., Evans, C. S., & Marler, P. (1993). The role of articulations in the production of rhesus monkey, Macaca mulatta, vocalizations. *Animal Behaviour, 45*, 423 - 433.

Hauser, M. D., & McDermott, J. (2003). The evolution of the music faculty: A comparative perspective. *Nature Neuroscience, 6*(7), 663 - 668.

Helweg, D. A., Frankel, A. S., Mobley, J. R., & Herman, L. M. (1992). Humpback whale song: Our current understanding. In J. A. Thomas, R. A. Kastelein & A. Y. Supin (Eds.), *Marine mammal sensory systems* (pp. 459 - 483). New York: Plenum.

Hetland, L. (2000a). Learning to make music enhances spatial reasoning. *Journal of Aesthetic Education, 34*, 179 - 238.

Hetland, L. (2000b). Listening to music enhances spatial-temporal reasoning: Evidence for the "Mozart effect". *Journal of Aesthetic Education, 34*(3-4), 105 - 148.

Hinde, R. A., & Rowell, T. (1962). Communications by postures and facial expression in the Rhesus monkey (Macaca mulatta). *Proceedings of the Zoological Society of London, 138*, 1 - 21.

Hockett, C. (1958). *A course in modern linguistics*. New York: MacMillan.

Hockett, C., & Altman, S. (1968). A note on design features. In T. Sebeok (Ed.), *Animal communication: Techniques of study and results of research*. Bloomington: Indiana University Press.

Holec, H. (1981). *Autonomy in foreign language learning*. Oxford: Pergamon.

Holmes, J. A. (1954). A substrata analysis of spelling ability for elements of auditory images. *Journal of Experimental Education, 22*, 329 - 349.

Hopkins, W. G. (2000). A scale of magnitudes for effect statistics. In *A new view of statistics*. Internet Society for Sport Science: http://www.sportsci.org/resource/stats/effectmag.html (last verified 17.06.05).

Horn, J. L., & Cattell, R. B. (1966). Refinement and test of the theory of fluid and crystallised intelligence. *Journal of Educational Psychology, 57*, 253 - 270.

Hutchinson, S. P. (1973). *An objective index of the English-Spanish pronunciation dimension.* University of Texas, Austin, TX.

Ivanov, V. K., & Geake, J. G. (2003). The Mozart effect and primary school children. *Psychology of Music, 31*(4), 405 - 413.

Jakobson, L. S., Cuddy, L. L., & Kilgour, A. R. (2003). Time tagging: A key to musicians' superior memory. *Music Perception, 20*(3), 307 - 313.

Jensen, A. R. (1997). The psychophysiology of g. In C. Cooper & V. Varma (Eds.), *Process models of individual differences*. London: Routledge.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.

Jurden, F. H. (1995). Individual differences in working memory and complex cognition. *Journal of Educational Psychology, 87*, 93 - 102.

Kallinen, K. (2002). Reading news from a pocket computer in a distracting environment: Effects of the tempo of background music. *Computers in Human Behaviour, 18*, 537 - 551.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*(2), 189 - 217.

Karimer, L. (1984). Can Southeast Asian students learn to discriminate between English phonemes more quickly with the aid of music and rhythm? *Bilingual Education, V(12).*

Karlin, J. E. (1942). *The factorial isolation of the primary auditory abilities.* University of Chicago.

Kelly, G. (1955). *The psychology of personal constructs*. New York: Norton.

Kelly, G. (1963). *The autobiography of a theory.* Ohio State University, Columbus, OH.

Kim, J.-M., Wang, C., Peabody, M., & Seneff, S. (2004). An interactive English pronunciation dictionary for Korean learners, *8th International Conference on Spoken Language Processing*. Jeju Island, Korea.

Kim, Y.-W. (2005). *Master key to English hearing and pronunciation*: Mesa Publishing Co.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30*, 580 - 602.

Kitano, K. (2001). Anxiety in the college Japanese language classroom. *The Modern Language Journal, 85*(4), 549 - 566.

Kline, P. (1981). *Fact and fantasy in Freudian theory*. London: Methuen.

Koren, S. (1993). *Relevance of musicality to the acquisition of foreign language pronunciation.* Unpublished Ph.D. thesis, Bar Ilan University.

Krashen, S. (1981). Aptitude and attitude in relation to second language acquisition and learning. In K. C. Diller (Ed.), *Individual differences and universals in foreign language aptitude*. Rowley, Mass.: Newbury House.

Krashen, S. (1982). Accounting for child-adult differences in second language rate and attainment. In S. Krashen, R. Scarcella & M. Long (Eds.), *Child adult differences in second language acquisition* (pp. 202-226). Rowley, Mass.: Newbury House.

Krashen, S. (1983). The din in the head, input, and language acquisition device. *Foreign Language Ann., 16*, 41-44.

Kreppner, K. (1992). William L. Stern 1871-1938: A neglected founder of developmental psychology. *Developmental Psychology, 28*, 539 - 547.

Kwalwasser, J., & Dykema, P. W. (1930). *Kwalwasser-Dykema music tests*. New York: Carl Fischer.

Kyllonen, P. C., & Christal, R. (1990). Reasoning ability is (little more than) working memory capacity? *Intelligence, 14*, 389 - 433.

Kyllonen, P. C., & Stephens, D. L. (1990). Cognitive abilities as determinants of success in acquiring logic skill. *Learning and Individual Differences, 2*, 129 - 160.

Lamont, A. (1998). Respondent 3 - responses to Katie Overy's paper, "can music really 'improve' the mind?" *Psychology of Music, 26*, 201 - 204.

Lansdell, H. (1962). A sex difference in effect of temporal lobe neurosurgery on design preference. *Nature, 194*, 852 - 854.

Larsen-Freeman, D., & Long, M. H. (1991). *An introduction to second language acquisition research*. London & New York: Longman.

Lehman, P. R. (1968). *Tests and measurements in music*. Englewood Cliffs, NJ.: Prentice-Hall.

Leith, G. (1974). Individual differences in learning: Interaction of personality and teaching methods. In *Personality and academic process* (pp. 14 - 25). London: Association of Educational Psychologists.

Leng, X., & Shaw, G. L. (1991). Toward a neural theory of higher brain function using music as a window. *Concepts in Neuroscience, 2*, 229 - 258.

Lenneberg, E. (1967). *Biological foundations of language*. New York: Wiley & Sons.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA.: The MIT Press.

Leutenegger, R. R., Mueller, T. H., & Wershow, I. R. (1965). Auditory factors in foreign language acquisition. *Modern Language Journal, 49*, 22 - 31.

Likert, R., & Quasha, W. (1948). *The Revised Minnesota Paper Form Board Test*. New York: Psychological Corp.

Little, D. (1991). *Learner autonomy. 1, definitions, issues and problems*. Dublin: Authentik.

Little, D., Ridley, J., & Ushioda, E. (Eds.). (2003). *Learner autonomy in the foreign language classroom: Teacher, learner, curriculum and assessment*. Dublin: Authentik.

Logie, R. H., & Edworthy, J. (1986). Shared mechanisms in the processing of verbal and musical material. In D. G. Russell, D. F. Marks & J. T. E. Richardson (Eds.), *Imagery 2* (pp. 32 - 37). New Zealand: Human Performance Associates.

Loring, D. W., Meador, K., Lee, G., Murro, A., Smith, J., Flanigan, H., et al. (1990). Cerebral language lateralization: Evidence from intracarotid amobarbitol testing. *Neuropsychologia, 28*, 831 - 838.

Low, E. E. L., & Grabe, E. (1995, 13 - 19 August). *Prosodic patterns in Singapore English.* Paper presented at the Intonational Congress of Phonetic Sciences, Stockholm.

Low, E. E. L., Grabe, E., & Nolan, F. (2001). Quantitative characteristics of speech rhythm: 'syllable-timing' in Singapore English. *Language and Speech, 43*(4), 377 - 401.

Lowe, A. (1998). Integration of music and French: A successful story. *International Journal of Music Education, 32*, 33 - 52.

MacIntyre, P. D. (2002). Motivation, anxiety and emotion in second language acquisition. In P. Robinson (Ed.), *Individual differences and instructed language learning* (Vol. 2, pp. 45 - 68). Amsterdam: John Benjamins Publishing Co.

Mackintosh, N. J. (1998). *IQ and human intelligence*. Oxford: Oxford University Press.

Mackintosh, N. J., & Bennett, E. S. (2003). The fractionation of working memory maps onto different components of intelligence. *Intelligence, 31*, 519 - 531.

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience, 4*(5), 540 - 545.

Marthohardjono, G., & Flynn, S. (1995). Is there an age factor for universal grammar? In D. Singleton & Z. Lengyel (Eds.), *The age factor in second language acquisition* (pp. 135-153). Clevedon: Multilingual Matters Ltd.

Marthohardjono, G., & Gair, J. (1993). Apparent UG inaccessibility in second language acquisition: Misapplied principles or principled misapplications?

In F. Eckman (Ed.), *Confluence: Linguistics, L2 acquisition and speech pathology* (pp. 79-103). Amsterdam: John Benjamins.

Mayer, J. D., & Salovey, P. (1993). The intelligence of emotional intelligence. *Intelligence, 17*, 433 - 442.

McCrae, R. R., & Costa, P. T. (1989). Validation of the five-factor model of personality across instruments and observers. *Journal of Personality and Social Psychology, 52*, 81 - 90.

McCrae, R. R., & Costa, P. T. (1990). *Personality in adulthood*. New York: Guilford Press.

McLaughlin, B. (1990). The relationship between first and second languages: Language-proficiency and language aptitude. In B. Harley, P. Allen, J. Cummins & M. Swain (Eds.), *The development of second language proficiency* (pp. 158 - 178). Cambridge: Cambridge University Press.

Meschyan, G., & Hernandez, A. (2002). Is native-language decoding skill related to second-language learning? *Journal of Educational Psychology, 94*(1), 14 - 22.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review, 63*, 81 - 97.

Miller, G. F. (2000). *The mating mind*. New York: Doubleday.

Miller, M. (1984). On the perception of rhythm. *Journal of Phonetics, 12*, 75 - 83.

Milner, B. (1962). Laterality effects in audition. In V. Mountcastle (Ed.), *Interhemispheric relations and cerebral dominance*. Baltimore: John Hopkins University Press.

Müller, R.-A., & Basho, S. (in press). Are nonlinguistic functions in "Broca's area" prerequisites for language acquisition? FMRI findings from an ontogenetic viewpoint. *Brain and Language*.

Muñoz, C. (2003). Variation in oral skills development and age of onset. In M. Del Pilar Garcia Mayo & M. L. Garcia Lecumberri (Eds.), *Age and the acquisition of English as a foreign language*. Clevedon: Multilingual Matters.

Murayama, J., Kashiwagi, T., Kashiwagi, A., & Mimura, M. (2004). Impaired pitch production and preserved rhythm production in a right brain-damaged patient with amusia. *Brain and Cognition, 56*, 36-42.

Murphey, T. (1990). The song stuck in my head phenomenon: A melodic din in the lad? *System, 18*(1), 53 - 64.

Nager, W., Kohlmetz, C., Altenmüller, E., Rodriguez-Fornells, A., & Münte, T. F. (2003). The fate of sounds in conductors' brains: An ERP study. *Cognitive Brain Research, 17*, 83 - 93.

Nantais, K. M., & Schellenberg, E. G. (1999). The Mozart effect: An artifact of preference. *Psychological Science, 10*(4), 370 - 373.

Nazzi, T., Jusczyk, P. W., & Johnson, E. K. (2000). Language discrimination by English-learning 5-month-olds: Effects of rhythm and familiarity. *Journal of Memory and Language, 43*, 1 - 19.

Nell, V. (2000). *Cross-cultural neuropsychological assessment: Theory and practics*. Hillsdale, NJ: Erlbaum.

Neufeld, G. (1978). A theoretical perspective on the nature of linguistic aptitude. *International Review of Applied Linguistics, 16*, 15 - 26.

Norman, W. T. (1963). Toward an adequate taxonomy of personality attributes. *Journal of Abnormal and Social Psychology, 66*, 574 - 583.

Núñez-Peña, M. I., & Honrubia-Serrano, M. L. (2004). P600 related to rule violation in an arithmetic task. *Cognitive Brain Research, 18*, 130 - 141.

Oller, J., & Perkins, K. (1978). Intelligence and language proficiency as sources of variance in self-reported affective variables. *Language Learning, 28*, 85-97.

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language, 32*, 258 - 278.

Overy, K. (1998). Discussion note: Can music really "improve" the mind? *Psychology of Music, 26*, 97 - 99.

Oxford, R. L. (1990). Styles, strategies and aptitude: Connections for language learning. In T. S. Parry & C. W. Stansfield (Eds.), *Language aptitude reconsidered* (pp. 67 - 125). New Jersey: Prentice Hall Regents.

Pantev, C., Oostenveld, R., Engellen, A., Ross, B., Roberts, L., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature, 392*, 811 - 814.

Papagno, C., Valentine, T., & Baddeley, A. D. (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language, 30*, 331 - 347.

Papagno, C., & Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. *Quarterly Journal of Experimental Psychology, 48*, 98 - 107.

Park, J. (1990). *The experimental phonetic study of rhythm -- Korean words in view of duration.* Unpublished Ph.D., Seoul National University, Seoul, Korea.

Parsons, L., Hodges, D., & Fox, P. T. (1998). Neural basis of the comprehension of musical harmony, melody and rhythm, *Cognitive Neuroscience Society Meeting*. San Francisco.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6*(7), 674 - 681.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience, 10*(6), 717 - 733.

Patel, A. D., & Peretz, I. (1997). Is music autonomous from language? A neuropsychological appraisal. In I. Deliège & J. Sloboda (Eds.), *Perception and cognition of music* (pp. 191 - 209). East Sussex: Psychology Press.

Payne, R. S., & Webb, D. (1971). Orientation by means of long range acoustic signalling in baleen whales. *Annals of the New York Academy of Sciences, 188*, 110 - 141.

Penfield, W., & Roberts, L. (Eds.). (1959). *Speech and brain mechanisms*. New York: Atheneum Press.

Perani, D., Paulesu, E., Sebastian Galles, N., Dupoux, E., Dehaene, S., Bettinardi, V., et al. (1998). The bilingual brain: Proficiency and age of acquisition of the second language. *Brain, 121*, 1841-1852.

Peterson, G. E., & Lehiste, I. (1960). Duration of syllable nuclei in English. *The Journal of the Acoustical Society of America, 32*(6), 693 - 703.

Pettigrew, T. W. (1958). The measurement and correlates of category width as a cognitive variable. *Journal of Personality, 26*, 532 - 544.

Piaget, J. (1923/1955). *The language and thought of the child* (M. Gabain, Trans.). New York: Meridian.

Piaget, J. (1970). Piaget's theory. In P. H. Mussen (Ed.), *Carmichael's manual of child psychology* (3rd ed., Vol. I, pp. 703 - 732). New York: Wiley.

Piaget, J. (1972). *The psychology of intelligence*. Totowa, NJ: Littlefield Adams.

Pike, K. L. (1945). *Intonation of American English*. Ann Arbor: University of Michigan.

Pimsleur, P. (1966). *The Pimsleur language aptitude battery*. New York: Harcourt Brace Jovanovice.

Pimsleur, P., Stockwell, R. P., & Comrey, A. L. (1962). Foreign language learning ability. *Journal of Educational Psychology, 53*(1), 15 - 26.

Pine, J. M. (1995). Variation in vocabulary development as a function of birth order. *Child Development, 66*, 272 - 281.

Pinker, S. (1994). *The language instinct*. New York: W. Morrow and Co.

Pinker, S. (1997). *How the mind works*. New York: Norton.

Plato. (360 B.C.E.). *The Republic* (B. Jowett, Trans.): http://classics.mit.edu/Plato/republic.html (last verified 16.08.2005).

Plomin, R., Campos, J., & Corley, R. (1990). Individual differences within the second year of life: The MacArthur longitudinal twin study. In J. Columbo & J. Fagan (Eds.), *Individual differences in infancy: Reliability, stability and predictability*. Hillsdale, NJ: Lawrence Erlbaum.

Pointon, G. E. (1995). Rhythm and duration in Spanish. In J. W. Lewis (Ed.), *Studies in general and English phonetics: Essays in honour of professor j.D. O'Connor* (pp. 266 - 269). London and New York: Routledge.

Politzer, R. L. (1983). An exploratory study of self-reported language learning behaviours and their relation to achievement. *Studies in Second Language Acquisition, 6*(1), 54 - 68.

Politzer, R. L., & Weiss, L. (1969). An experiment in improving achievement in foreign language learning through learning of selected skills associated with language aptitude. Stanford, CA.: Stanford University, Stanford, CA.

Pollard, C., & Sag, I. A. (1994). *Head-driven phrase structure grammar*. Chicago: University of Chicago Press.

Prior, M., Kinsella, G., & Giese, J. (1990). Assessment of musical processing in brain-damaged patients: Implications for laterality of music. *Journal of Clinical and Experimental Neuropsychology, 12*, 301 - 312.

Pullum, G. K. (1991). The great Eskimo vocabulary hoax. In G. K. Pullum (Ed.), *The great Eskimo vocabulary hoax and other irreverent essays on the study of language* (pp. 159 - 171). Chicago and London: University of Chicago Press.

Ramus, F., Hauser, M. D., Miller, C. T., Morris, D., & Mehler, J. (2000). Language discrimination by human newborns and cotton-top tamarins. *Science, 288*, 349 - 351.

Rauscher, F. H. (1999). Enhancing abstract reasoning in disadvantaged children (unpublished raw data).

Rauscher, F. H. (2000). Is the "Mozart effect" debunked? *Bi-annual meeting of the International Conference on Music Perception and Cognition*. Keele University, Keele, UK.

Rauscher, F. H. (2002). Mozart and the mind: Factual and fictional effects of musical enrichment. In J. Aronson (Ed.), *Improving academic achievement: Impact of psychological factors on education* (pp. 269 - 278). New York: Academic Press.

Rauscher, F. H., & Shaw, G. L. (1998). Key components of the Mozart effect. *Perceptual and Motor Skills, 86*, 835 - 841.

Rauscher, F. H., Shaw, G. L., & Ky, K. N. (1993). Music and spatial task performance. *Nature, 365*, 611.

Rauscher, F. H., Shaw, G. L., & Ky, K. N. (1995). Listening to Mozart enhances spatial-temporal reasoning: Towards a neurophysiological basis. *Neuroscience Letters, 185*, 44 - 47.

Rauscher, F. H., Shaw, G. L., Levine, L. J., Wright, E. L., Dennis, W. R., & Newcomb, R. L. (1997). Music training causes long-term enhancement of preschool children's spatial-termporal reasoning. *Neurological Research, 19*, 2 - 8.

Rauscher, F. H., & Sparr, M. (2000). Experimenter expectancy and the "Mozart effect": Unpublished raw data.

Rauscher, F. H., & Zupan, M. A. (2000). Classroom keyboard instruction improves kindergarten children's spatial-temporal performance: A field experiment. *Early Childhood Research Quarterly, 15*(2), 215 - 228.

Raven, J. C., Court, J. H., & Raven, J. (1998). Raven manual section 3 (1998 ed.). Oxford: Oxford Psychologists Press.

Reves, T. (1983). *What makes a good language learner?* Hebrew University.

Rideout, B. E., Dougherty, S., & Wernert, L. (1998). Effect of music on spatial performance: A test of generality. *Perceptual and Motor Skills, 86*, 512 - 514.

Robinson, P. (2001). Individual differences, cognitive abilities, aptitude complexes and learning conditions in second language acquisition. *Second Language Research, 17*(4), 368 - 392.

Robinson, P. (2002a). Learning conditions, aptitude complexes and SLA. In P. Robinson (Ed.), *Individual differences in instructed language learning* (Vol. 2, pp. 113 - 136). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Robinson, P. (2003). Attention and memory during SLA. In C. J. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 631 - 678). Oxford: Blackwell.

Robinson, P. (Ed.). (2002b). *Individual differences and instructed language learning* (Vol. 2). Amsterdam: John Benjamins Publishing Co.

Rogers, C. R. (1959). A theory of therapy, personality and interpersonal relationships, as developed in the client-centered framework. In S. Koch (Ed.), *Psychology: A study of science* (pp. 184 - 256). New York: McGraw-Hill.

Rogers, C. R. (1967). *On becoming a person*. London: Constable.

Rogers, C. R., Kirschenbaum, H., & Henderson, V. (Eds.). (1989). *The Carl Rogers reader*. Boston, MA: Houghton Mifflin.

Romani, C., Ward, J., & Olson, A. (1999). Developmental surface dysgraphia: What is the underlying cognitive impairment. *Quarterly Journal of Experimental Psychology: Section A, 52*(1), 97 - 128.

Rönnberg, J., Rudner, M., & Ingvar, M. (2004). Neural correlates of working memory for sign language. *Cognitive Brain Research, 20*, 165 - 182.

Rossier, R. (1976). *Extraversion-introversion as a significant variable in the learning of oral English.* Unpublished Ph.D. thesis, University of Southern California.

Rousseau, J.-J. (1998). *Essay on the origin of languages and writings related to music* (J. T. Scott, Trans. Vol. 7). London: University Press of New England.

Rowell, T. E., & Hinde, R. A. (1962). Vocal communication by the rhesus monkey (Macaca mulatta). *Proceedings of the Zoological Society of London, 138*, 279 - 294.

Rymer, R. (1993). *Genie: An abused child's flight from silence*. New York: Harper Collins.

Şakar, A. (2003). *Exploring the relationships among learning styles, annotation use and reading comprehension for foreign language reading in a hypermedia environment.* Unpublished Master of Arts, Boğaziçi University, Istanbul, Turkey.

Sarnthein, J., vonStein, A., Rappelsberger, P., Petsche, H., Rauscher, F. H., & Shaw, G. L. (1997). Persistent patterns of brain activity: An EEG coherence study of the positive effect of music on spatial-temporal reasoning. *Neurological Research, 19*, 107 - 116.

Sasaki, M. (1993). Relationships among second language proficiency, foreign language aptitude, and intelligence: A structural equation modeling approach. *Language Learning, 43*(3), 313 - 344.

Sasaki, M. (1996). *Second language proficiency, foreign language aptitude, and intelligence*. Baltimore: Peter Lang.

Schellenberg, E. G. (2003). Does exposure to music have beneficial side effects? In I. Peretz & R. J. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 430 - 448). New York: Oxford University Press.

Schlaug, G., Jancke, L., Huang, Y., & Steinmetz, H. (1994). *In vivo morphometry of interhemispheric asymmetry and connectivity in musicians.* Paper presented at the 3rd International Conference for Music Perception and Cognition, Liège, Belgium.

Schlaug, G., Jancke, L., Huang, Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science, 267*, 699 - 701.

Schlenker, B., & Leary, M. (1982). Social anxiety and self-presentation: A conceptualization and model. *Psychological Bulletin, 92*, 641 - 669.

Schmeck, R. (1988). *Learning strategies and learning styles*. New York: Plenum.

Sdorow, L. M., & Rickabaugh, C. A. (2002). *Psychology* (5th ed.). New York: McGraw Hill.

Seashore, C. E., Lewis, D., & Saetveit, J. (1939, 1960). *Seashore Measures of Musical Talents*. New York: The Psychological Corporation.

Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics, 10*, 209 - 231.

Service, E. (1992). Phonology, working memory, and foreign-language learning. *Quarterly Journal of Experimental Psychology, 45A*, 21 - 50.

Sevilla, J. M. (1996). Involuntary rehearsal of second language at the elementary level: Do elementary school children experience the din in the head? *System, 24*(1), 101-105.

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*, 4 - 27.

Sharp, C. (1998). *The effects of teaching and learning in the arts*. London: Qualifications and Curriculum Authority/National Foundation for Educational Research.

Shore, C. M. (1995). *Individual differences in language development* (Vol. 7). London, New Delhi: Thousand Oaks.

Singleton, D. (1998). *Lexical processing and the "language module"* (CLCS Occasional Paper No. 53). Dublin, Ireland: Centre for Language and Communication Studies, Trinity College Dublin.

Singleton, D. (2001). Age and second language acquisition. *Annual Review of Applied Linguistics, 21*, 77-89.

Singleton, D. (2003). Critical period or general age factor(s)? In M. Del Pilar Garcia Mayo & M. L. Garcia Lecumberri (Eds.), *Age and the acquisition of English as a foreign language*. Clevedon: Multilingual Matters.

Singleton, D. (2004). Will the real critical period hypothesis please stand up? *EUROSLA*. Bilbao, Spain.

Skehan, P. (1980). Memory, language aptitude, and second language performance. *Polyglot 2* (Fiche 3).

Skehan, P. (1982). *Memory and motivation in language aptitude testing*. University of London.

Skehan, P. (1985). Where does language aptitude come from? *BAAL 1985 Seminar*. University of Edinburgh.

Skehan, P. (1986). Where does language aptitude come from? In P. Meara (Ed.), *Spoken language*. London: Centre for Information on Language Teaching.

Skehan, P. (1989). *Individual differences in second language learning*. London: Edward Arnold.

Skehan, P. (1998). *A cognitive approach to language learning*. Oxford: Oxford University Press.

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (Vol. 2, pp. 69 - 94). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Skehan, P., & Ducroquet, L. (1988). A comparison of first and foreign language learning ability (Vol. Working Documents No. 8). ESOL Department: Institute of Education, London University.

Smith Salcedo, C. (2002). *The effects of song in the foreign language classroom on text recall and involuntary mental rehearsal.* Louisiana State University and Agricultural and Mechanical College.

Snow, R. E. (1987). Aptitude complexes. In R. E. Snow & M. J. Farr (Eds.), *Aptitude, learning and instruction*. Hillsdale, NJ: Erlbaum.

Snow, R. E. (1994). Abilities in academic tasks. In R. J. Sternberg & W. R. K. (Eds.), *Mind in context: Interactionist perspectives on human intelligence* (pp. 3 - 37). New York: Cambridge University Press.

Sparks, R. L., Ganschow, L., & Pohlman, J. (1989). Linguistic coding deficits in foreign language learners. *Annals of Dyslexia, 39*, 179 - 195.

Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology, 15*, 201 - 293.

Spearman, C. (1927). *The abilities of man: Their nature and measurement*. New York: Macmillan.

Spiro, J. (2003). Music and the brain. *Nature Neuroscience, 6*(7), 661.

Springer, S. P., & Deutsch, G. (1998). *Left brain right brain* (5th ed.). New York: W.H. Freemand and Company.

SPSS Inc. (2001). SPSS for Windows (Version 11.0.1).

Spychiger, M. (1993). *Music makes the school*: Die Blaue Eule.

Spychiger, M. (1998). Responses to Katie Overy's paper, "can music really improve the mind?" *Psychology of Music, 26*, 197 - 210.

Stankov, L., & Horn, J. L. (1980). Human abilities revealed through auditory tests. *Journal of Educational Psychology, 72*, 21 - 44.

Steele, K. M., Bass, K. E., & Crook, M. D. (1999). 'the mystery of the Mozart effect'. *Psychological Science, 10*(4), 366 - 368.

Stern, W. (1912). *Die psychologische methoden der intelligenzprüfung*. Leipzig: Barth.

Sternberg, R. J. (1985). *Beyond IQ: A triarchic theory of intelligence*. New York: Cambridge University Press.

Sternberg, R. J. (1990). *Metaphors of mind: Conceptions of the nature of intelligence*. New York: Cambridge University Press.

Sternberg, R. J. (1997). *Successful intelligence*. New York: Plume.

Sternberg, R. J. (2002). The theory of successful intelligence and its implications for language aptitude testing. In P. Robinson (Ed.), *Individual differences and instructed language learning* (Vol. 2, pp. 13 - 44). Amsterdam: John Benjamins Publishing Co.

Stevenson, R. (1999). *Is there an association between musical and linguistic aptitude? An empirical study.* Unpublished M.Phil, Trinity College, Dublin.

Sutcliffe, T. (in progress). Syntactic structures in music: http://www.harmony.org.uk/book_chord_progressions.htm (last verified 11.08.2005).

Sutton, J. (1995). The sound-world of speech and language impaired children. In A. Gilroy & C. Lea (Eds.), *Art and music: Therapy and research* (pp. 152 - 163). London and New York: Routledge.

Swinscow, T. D. V. (1997). Statistics at square one: BMJ Publishing Group.

Thorndike, R., Hagen, E., & Jerome, M. (1986). *The Stanford-Binet intelligence scale*. Chicago, IL.: Riverside Publishing.

Thurstone, L. L. (1931). Multiple factor analysis. *Psychological Review, 38*, 406 - 427.

Thurstone, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.

Thurstone, L. L. (1940). *The vectors of mind*. Chicago: University of Chicago Press.

Trainor, L. J., & Heinmiller, B. M. (1998). The development of evaluative responses to music: Infants prefer to listen to consonance over dissonance. *Inf. Behav. Dev., 21*, 77 - 88.

Trehub, S. E. (2000). Human processing predispositions and musical universals. In N. L. Wallin, B. Merker & S. Brown (Eds.), *The origins of music* (pp. 427 - 448). Cambridge, MA: MIT Press.

Trehub, S. E. (2001). Musical predispositions in infancy. *Annals of the New York Academy of Sciences, 930*, 1 - 16.

Trehub, S. E. (2003). The developmental origins of musicality. *Nature Neuroscience, 6*(7), 669 - 673.

Tucker, A. (2000). Relationships among FL aptitude, music aptitude, and L2 proficiency. In S. Cornwell & P. Robinson (Eds.), *Individual differences in FLL: Effects of aptitude, intelligence and motivation* (pp. 49-65). Tokyo: Japanese Association for Language Teaching.

Tuholski, S. W. (1997). *Examining the relationships among attentional resources, working memory, and fluid intelligence.* University of Southern Carolina, Columbia, SC.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*, 127 - 154.

von Frisch, K. (1950). *Bees: Their vision, chemical sense and language*. Ithaca, NY.: Cornell University Press.

von Frisch, K. (1954). *The dancing bees*. London: Methuen.

Vygotsky, L. S. (1962). *Thought and language*. Cambridge, MA.: MIT Press.

Walker, R. (1996). Open peer commentary: Can we understand the music of another culture? *Psychology of Music, 24*, 103 - 130.

Waters, A. J. (1998). Respondent 5: Responses to Katie Overy's paper "can music really 'improve' the mind?" *Psychology of Music, 26*, 205 - 208.

Watkins, W. A., Tyack, P., & Moore, K. E. (1987). The 20-Hz signals of finback whales (Balaenoptera physalus). *Journal of the Acoustical Society of America, 82*, 1901 - 1912.

Wechsler, D. (1958). *Measurement and appraisal of adult intelligence* (4th ed.). Baltimore: Williams and Wilkins.

Weinberger, N. M. (1995). The musician's brain, *MuSICA Research Notes*.

Weinberger, N. M. (1998a). Brain, behaviour, biology and music: Some research findings and their implications for educational policy. *Arts Education Policy Review, 99*, 28 - 36.

Weinberger, N. M. (1998b). Creating creativity with music, *MuSICA Research Notes* (Vol. 2).

Weintraub, S., Mesulam, M. M., & Kramer, L. (1981). Disturbances in prosody: A right hemisphere contribution to language. *Archives of Neurology, 38*, 742 - 744.

Wernicke, C. (1874). *The aphasia symptom-complex: A psychological study on an anatomical basis.*

Wesche, M. (1981). Language aptitude measures in streaming, matching students with methods, and diagnosis of learning problems. In K. Diller (Ed.), *Individual differences and universals in language learning* (pp. 83-118). Rowley Mass.: Newbury House.

Wheeler, L. R., & Wheeler, V. D. (1954). A study of the relationship of auditory discrimination to silent reading abilities. *Journal of Educational Research, 48*, 103 - 113.

Whorf, B. L. (1940). Science and linguistics. *Technology Review (MIT), 42*(6), 229 - 231, 247 - 248.

Williams, J. (1999). Memory, attention and inductive learning. *Studies in Second Language Acquisition, 21*, 1 - 48.

Williamson, A., & Egner, T. (2004). Memory structures for encoding and retrieving a piece of music: An ERP investigation. *Cognitive Brain Research, 22*, 36 - 44.

Williamson, A., & Valentine, E. (2002). The role of retrieval structures in memorizing music. *Cognitive Psychology, 44*, 1 - 32.

Wilson, M., Bettger, J. G., Niculae, I., & Klima, E. S. (1997). Modality of language shapes working memory: Evidence from digit span and spatial span in ASL signers. *Journal of Deaf Studies and Deaf Education, 2*(3), 150 - 160.

Wilson, M., & Emmorey, K. (1997a). A visual-spatial 'phonological loop' in working memory: Evidence from American sign language. *Memory & Cognition, 25*(3), 313 - 320.

Wilson, M., & Emmorey, K. (1997b). Working memory for sign language: A window into the architecture of the working memory system. *Journal of Deaf Studies and Deaf Education, 2*(3q), 121 - 130.

Wilson, M., & Emmorey, K. (1998). A word length effect for sign language: Further evidence on the role of language in structuring working memory. *Memory & Cognition, 26*(3), 584 - 590.

Wilson, M., & Emmorey, K. (2003). The effect of irrelevant visual input on working memory for sign language. *Journal of Deaf Studies and Deaf Education, 8*(2), 97 - 103.

Wilson, R. C. (1983). The Louisville twin study: Developmental synchronies in behaviour. *Child Development, 54*, 298 - 316.

Wing, H. (1968). *Tests of musical ability and appreciation* (2nd ed.). Cambridge: Cambridge University Press.

Wingfield, A., & Byrnes, D. (1981). *The psychology of human memory*. New York.

Wisbey, A. (1980). *Music as the source of learning*. Lancaster: M.T.P. Press.

Witkin, H., Oltman, P., Raskin, E., & Karp, S. (1971). *A manual for the embedded figures test*. Palo Alto, CA.: Consulting Psychologists Press.

Wright, A. A., Rivera, J. J., Hulse, S. H., Shyan, M., & Neiworth, J. J. (2000). Music perception and octave generalization in rhesus monkeys. *Journal of Experimental Psychology General, 129*, 291 - 307.

Zatorre, R. J. (1979). Recognition of dichotic melodies by musicians and non-musicians. *Neuropsychologia, 17*.

Zatorre, R. J. (2001). Neural specializations for tonal processing. *Annals of the New York Academy of Sciences, 930*, 193 - 210.

Appendix 1

Correlations

Pearson Correlations

Correlations

		PITCH	TUNES	CHORDS	RHYTHM	Music (Recep)	SOUND	CHINESE	Numerical	GRAMMAR	Language (Recep)	RAVEN
PITCH	Pearson Correlation	1	.412**	.199*	.228**	.724**	-.022	.093	.101	.140	.128	.167*
	Sig. (2-tailed)	.	.000	.015	.005	.000	.790	.259	.219	.088	.119	.042
	N	149	149	149	149	149	149	149	149	149	149	149
TUNES	Pearson Correlation	.412**	1	.293**	.346**	.705**	.118	.166*	.187*	.244**	.281**	.209*
	Sig. (2-tailed)	.000	.	.000	.000	.000	.152	.043	.023	.003	.001	.011
	N	149	149	149	149	149	149	149	149	149	149	149
CHORDS	Pearson Correlation	.199*	.293**	1	.254**	.683**	.206*	.088	.053	.201*	.224**	.286**
	Sig. (2-tailed)	.015	.000	.	.002	.000	.012	.283	.524	.014	.006	.000
	N	149	149	149	149	149	149	149	149	149	149	149
RHYTHM	Pearson Correlation	.228**	.346**	.254**	1	.597**	.311**	.068	.284**	.444**	.473**	.269**
	Sig. (2-tailed)	.005	.000	.002	.	.000	.000	.411	.000	.000	.000	.001
	N	149	149	149	149	149	149	149	149	149	149	149
Music (Recep)	Pearson Correlation	.724**	.705**	.683**	.597**	1	.202*	.148	.204*	.348**	.371**	.337**
	Sig. (2-tailed)	.000	.000	.000	.000	.	.013	.071	.013	.000	.000	.000
	N	149	149	149	149	149	149	149	149	149	149	149
SOUND	Pearson Correlation	-.022	.118	.206*	.311**	.202*	1	.054	.311**	.333**	.648**	.244**
	Sig. (2-tailed)	.790	.152	.012	.000	.013	.	.511	.000	.000	.000	.003
	N	149	149	149	149	149	149	149	149	149	149	149
CHINESE	Pearson Correlation	.093	.166*	.088	.068	.148	.054	1	.217**	.016	.323**	-.033
	Sig. (2-tailed)	.259	.043	.283	.411	.071	.511	.	.008	.845	.000	.686
	N	149	149	149	149	149	149	149	149	149	149	149
Numerical	Pearson Correlation	.101	.187*	.053	.284**	.204*	.311**	.217**	1	.366**	.701**	.242**
	Sig. (2-tailed)	.219	.023	.524	.000	.013	.000	.008		.000	.000	.003
	N	149	149	149	149	149	149	149	149	149	149	149
GRAMMAR	Pearson Correlation	.140	.244**	.201*	.444**	.348**	.333**	.016	.366**	1	.818**	.412**
	Sig. (2-tailed)	.088	.003	.014	.000	.000	.000	.845	.000	.	.000	.000
	N	149	149	149	149	149	149	149	149	149	149	149
Language (Recep)	Pearson Correlation	.128	.281**	.224**	.473**	.371**	.648**	.323**	.701**	.818**	1	.395**
	Sig. (2-tailed)	.119	.001	.006	.000	.000	.000	.000	.000	.000	.	.000
	N	149	149	149	149	149	149	149	149	149	149	149
RAVEN	Pearson Correlation	.167*	.209*	.286**	.269**	.337**	.244**	-.033	.242**	.412**	.395**	1
	Sig. (2-tailed)	.042	.011	.000	.001	.000	.003	.686	.003	.000	.000	.
	N	149	149	149	149	149	149	149	149	149	149	149

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

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

Spearman's Rho

Correlations

			PITCH	TUNES	CHORDS	RHYTHM	Music (Recep)	SOUND	CHINESE	Numerical	GRAMMAR	Language (Recep)	RAVEN
Spearman's rho	PITCH	Correlation Coefficient	1.000	.415**	.217**	.245**	.736**	-.002	.094	.102	.144	.158	.190*
		Sig. (2-tailed)	.	.000	.008	.003	.000	.979	.253	.214	.080	.054	.020
		N	149	149	149	149	149	149	149	149	149	149	149
	TUNES	Correlation Coefficient	.415**	1.000	.291**	.367**	.712**	.110	.167*	.189*	.237**	.293**	.224**
		Sig. (2-tailed)	.000	.	.000	.000	.000	.182	.042	.021	.004	.000	.006
		N	149	149	149	149	149	149	149	149	149	149	149
	CHORDS	Correlation Coefficient	.217**	.291**	1.000	.252**	.643**	.189*	.048	.036	.203*	.196*	.296**
		Sig. (2-tailed)	.008	.000	.	.002	.000	.021	.559	.660	.013	.017	.000
		N	149	149	149	149	149	149	149	149	149	149	149
	RHYTHM	Correlation Coefficient	.245**	.367**	.252**	1.000	.607**	.271**	.063	.285**	.406**	.475**	.274**
		Sig. (2-tailed)	.003	.000	.002	.	.000	.001	.442	.000	.000	.000	.001
		N	149	149	149	149	149	149	149	149	149	149	149
	Music (Recep)	Correlation Coefficient	.736**	.712**	.643**	.607**	1.000	.187*	.139	.198*	.334**	.381**	.348**
		Sig. (2-tailed)	.000	.000	.000	.000	.	.022	.091	.016	.000	.000	.000
		N	149	149	149	149	149	149	149	149	149	149	149
	SOUND	Correlation Coefficient	-.002	.110	.189*	.271**	.187*	1.000	.011	.293**	.261**	.578**	.271**
		Sig. (2-tailed)	.979	.182	.021	.001	.022	.	.898	.000	.001	.000	.001
		N	149	149	149	149	149	149	149	149	149	149	149
	CHINESE	Correlation Coefficient	.094	.167*	.048	.063	.139	.011	1.000	.200*	.017	.301**	-.025
		Sig. (2-tailed)	.253	.042	.559	.442	.091	.898	.	.014	.839	.000	.763
		N	149	149	149	149	149	149	149	149	149	149	149
	Numerical	Correlation Coefficient	.102	.189*	.036	.285**	.198*	.293**	.200*	1.000	.345**	.682**	.276**
		Sig. (2-tailed)	.214	.021	.660	.000	.016	.000	.014	.	.000	.000	.001
		N	149	149	149	149	149	149	149	149	149	149	149
	GRAMMAR	Correlation Coefficient	.144	.237**	.203*	.406**	.334**	.261**	.017	.345**	1.000	.811**	.435**
		Sig. (2-tailed)	.080	.004	.013	.000	.000	.001	.839	.000	.	.000	.000
		N	149	149	149	149	149	149	149	149	149	149	149
	Language (Recep)	Correlation Coefficient	.158	.293**	.196*	.475**	.381**	.578**	.301**	.682**	.811**	1.000	.409**
		Sig. (2-tailed)	.054	.000	.017	.000	.000	.000	.000	.000	.000	.	.000
		N	149	149	149	149	149	149	149	149	149	149	149
	RAVEN	Correlation Coefficient	.190*	.224**	.296**	.274**	.348**	.271**	-.025	.276**	.435**	.409**	1.000
		Sig. (2-tailed)	.020	.006	.000	.001	.000	.001	.763	.001	.000	.000	.
		N	149	149	149	149	149	149	149	149	149	149	149

**. Correlation is significant at the .01 level (2-tailed).

*. Correlation is significant at the .05 level (2-tailed).

Groups divided by Raven scores

Sample (n=149) divided into two groups – group AAS where subjects scored average, above average or superior and group BAI where subjects scored below average or impaired.

Group Statistics

	Raven Category	N	Mean	Std. Deviation	Std. Error Mean
Music (Recep)	AAS	86	35.65	7.423	.800
	BAI	63	31.30	7.553	.952
Language (Recep)	AAS	86	31.81	5.834	.629
	BAI	63	26.83	6.859	.864

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Music (Recep)	Equal variances assumed	.068	.794	3.507	147	.001	4.35	1.240	1.899	6.800
	Equal variances not assumed			3.498	132.417	.001	4.35	1.244	1.890	6.809
Language (Recep)	Equal variances assumed	2.160	.144	4.785	147	.000	4.99	1.043	2.928	7.049
	Equal variances not assumed			4.667	120.450	.000	4.99	1.069	2.872	7.105

Correlations

Raven Category			Music (Recep)	Language (Recep)	Raven	Rhythm
AAS	Music (Recep)	Pearson Correlation	1	.181	.203	.512**
		Sig. (2-tailed)	.	.096	.060	.000
		N	86	86	86	86
	Language (Recep)	Pearson Correlation	.181	1	.126	.350**
		Sig. (2-tailed)	.096	.	.248	.001
		N	86	86	86	86
	Raven	Pearson Correlation	.203	.126	1	-.017
		Sig. (2-tailed)	.060	.248	.	.876
		N	86	86	86	86
	Rhythm	Pearson Correlation	.512**	.350**	-.017	1
		Sig. (2-tailed)	.000	.001	.876	.
		N	86	86	86	86
BAI	Music (Recep)	Pearson Correlation	1	.442**	.258*	.616**
		Sig. (2-tailed)	.	.000	.041	.000
		N	63	63	63	63
	Language (Recep)	Pearson Correlation	.442**	1	.403**	.461**
		Sig. (2-tailed)	.000	.	.001	.000
		N	63	63	63	63
	Raven	Pearson Correlation	.258*	.403**	1	.097
		Sig. (2-tailed)	.041	.001	.	.451
		N	63	63	63	63
	Rhythm	Pearson Correlation	.616**	.461**	.097	1
		Sig. (2-tailed)	.000	.000	.451	.
		N	63	63	63	63

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

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

Groups divided by gender

Group Statistics

	0=female, 1=male	N	Mean	Std. Deviation	Std. Error Mean
Music (Recep)	.00	67	34.61	7.534	.920
	1.00	82	33.16	7.923	.875
Language (Recep)	.00	67	32.27	6.059	.740
	1.00	82	27.61	6.565	.725
RAVEN	.00	67	39.67	22.306	2.725
	1.00	82	45.23	31.171	3.442

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Music (Recep	Equal variance assumed	.191	.662	1.139	147	.257	1.45	1.276	-1.069	3.976
	Equal variance not assumed			1.144	143.620	.254	1.45	1.270	-1.057	3.964
Language (Recep)	Equal variance assumed	.278	.599	4.460	147	.000	4.66	1.045	2.595	6.723
	Equal variance not assumed			4.496	144.793	.000	4.66	1.036	2.611	6.707
RAVEN	Equal variance assumed	18.228	.000	-1.224	147	.223	-5.55	4.536	-14.519	3.411
	Equal variance not assumed			-1.265	144.628	.208	-5.55	4.390	-14.232	3.124

Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	RAVEN, RHYTHM[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: Language (Recep)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.549[a]	.301	.291	5.670

a. Predictors: (Constant), RAVEN, RHYTHM

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2021.449	2	1010.724	31.440	.000[a]
	Residual	4693.558	146	32.148		
	Total	6715.007	148			

a. Predictors: (Constant), RAVEN, RHYTHM

b. Dependent Variable: Language (Recep)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	18.481	1.550		11.925	.000
	RHYTHM	1.218	.221	.396	5.506	.000
	RAVEN	7.043E-02	.018	.288	4.016	.000

a. Dependent Variable: Language (Recep)

Outliers omitted

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	RAVEN, RHYTHM[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: Language (Recep)

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.580[a]	.336	.327	5.421

a. Predictors: (Constant), RAVEN, RHYTHM

b. Dependent Variable: Language (Recep)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2145.392	2	1072.696	36.505	.000[a]
	Residual	4231.438	144	29.385		
	Total	6376.830	146			

a. Predictors: (Constant), RAVEN, RHYTHM

b. Dependent Variable: Language (Recep)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	17.978	1.487		12.089	.000
	RHYTHM	1.312	.213	.435	6.157	.000
	RAVEN	6.817E-02	.017	.282	3.992	.000

a. Dependent Variable: Language (Recep)

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	21.15	37.58	29.70	3.833	147
Std. Predicted Value	-2.231	2.055	.000	1.000	147
Standard Error of Predicted Value	.451	1.156	.752	.185	147
Adjusted Predicted Value	20.91	37.85	29.70	3.836	147
Residual	-12.39	15.02	.00	5.384	147
Std. Residual	-2.285	2.770	.000	.993	147
Stud. Residual	-2.297	2.782	.000	1.003	147
Deleted Residual	-12.52	15.15	.00	5.487	147
Stud. Deleted Residual	-2.332	2.850	.000	1.009	147
Mahal. Distance	.019	5.648	1.986	1.440	147
Cook's Distance	.000	.049	.006	.009	147
Centered Leverage Value	.000	.039	.014	.010	147

a. Dependent Variable: Language (Recep)

Gender as predictor

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Gender, Raven,Rhythm[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: Language (Recep)

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.655[a]	.429	.417	5.046

a. Predictors: (Constant), Gender, Raven, Rhythm

b. Dependent Variable: Language (Recep)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2736.106	3	912.035	35.823	.000[a]
	Residual	3640.723	143	25.460		
	Total	6376.830	146			

a. Predictors: (Constant), Gender, Raven, Rhythm

b. Dependent Variable: Language (Recep)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	21.273	1.544		13.778	.000
	Raven	7.951E-02	.016	.329	4.948	.000
	Rhythm	1.088	.204	.360	5.336	.000
	Gender	-4.149	.861	-.314	-4.817	.000

a. Dependent Variable: Language (Recep)

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	19.94	39.07	29.70	4.329	147
Std. Predicted Value	-2.256	2.164	.000	1.000	147
Standard Error of Predicted Value	.569	1.204	.820	.144	147
Adjusted Predicted Value	19.64	39.03	29.70	4.330	147
Residual	-10.44	12.66	.00	4.994	147
Std. Residual	-2.069	2.509	.000	.990	147
Stud. Residual	-2.087	2.532	.000	1.002	147
Deleted Residual	-10.62	12.89	.00	5.122	147
Stud. Deleted Residual	-2.112	2.582	.001	1.008	147
Mahal. Distance	.864	7.325	2.980	1.415	147
Cook's Distance	.000	.046	.006	.008	147
Centered Leverage Value	.006	.050	.020	.010	147

a. Dependent Variable: Language (Recep)

Stepwise regression

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Rhythm	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
2	Raven	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).
3	Gender	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

a. Dependent Variable: Language (Recep)

Model Summary[d]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.513[a]	.263	.258	5.693
2	.580[b]	.336	.327	5.421
3	.655[c]	.429	.417	5.046

a. Predictors: (Constant), Rhythm

b. Predictors: (Constant), Rhythm, Raven

c. Predictors: (Constant), Rhythm, Raven, Gender

d. Dependent Variable: Language (Recep)

ANOVA[d]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1677.069	1	1677.069	51.742	.000[a]
	Residual	4699.761	145	32.412		
	Total	6376.830	146			
2	Regression	2145.392	2	1072.696	36.505	.000[b]
	Residual	4231.438	144	29.385		
	Total	6376.830	146			
3	Regression	2736.106	3	912.035	35.823	.000[c]
	Residual	3640.723	143	25.460		
	Total	6376.830	146			

a. Predictors: (Constant), Rhythm

b. Predictors: (Constant), Rhythm, Raven

c. Predictors: (Constant), Rhythm, Raven, Gender

d. Dependent Variable: Language (Recep)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	19.267	1.525		12.638	.000
	Rhythm	1.548	.215	.513	7.193	.000
2	(Constant)	17.978	1.487		12.089	.000
	Rhythm	1.312	.213	.435	6.157	.000
	Raven	6.817E-02	.017	.282	3.992	.000
3	(Constant)	21.273	1.544		13.778	.000
	Rhythm	1.088	.204	.360	5.336	.000
	Raven	7.951E-02	.016	.329	4.948	.000
	Gender	-4.149	.861	-.314	-4.817	.000

a. Dependent Variable: Language (Recep)

Excluded Variables[c]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
1	Raven	.282[a]	3.992	.000	.316	.924
	Gender	-.267[a]	-3.835	.000	-.304	.962
2	Gender	-.314[b]	-4.817	.000	-.374	.941

a. Predictors in the Model: (Constant), Rhythm

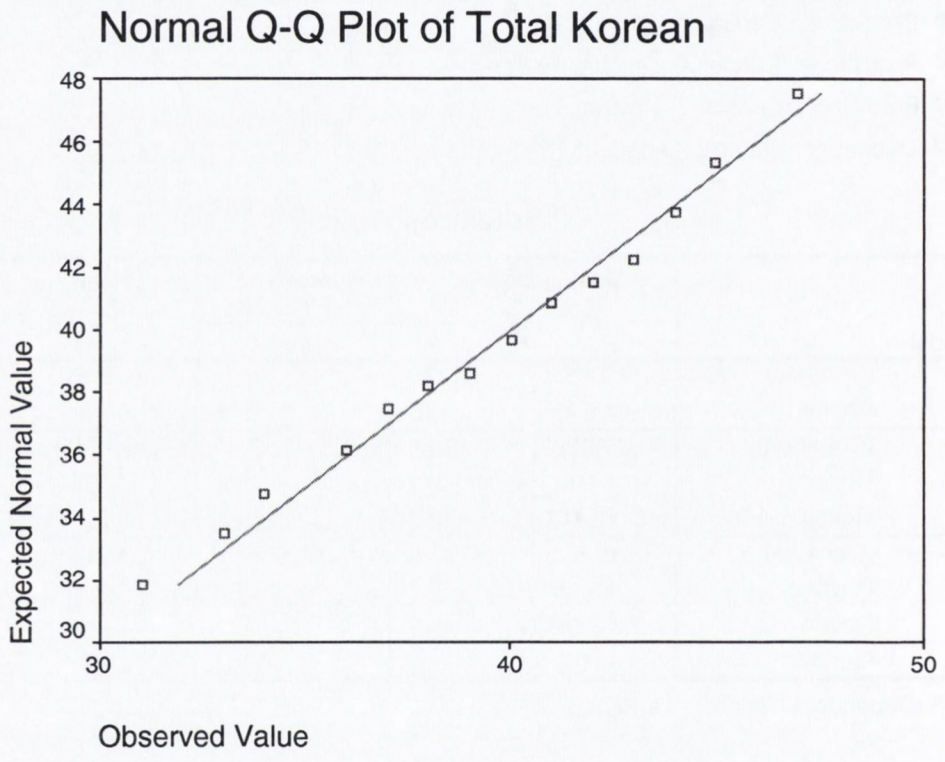
b. Predictors in the Model: (Constant), Rhythm, Raven

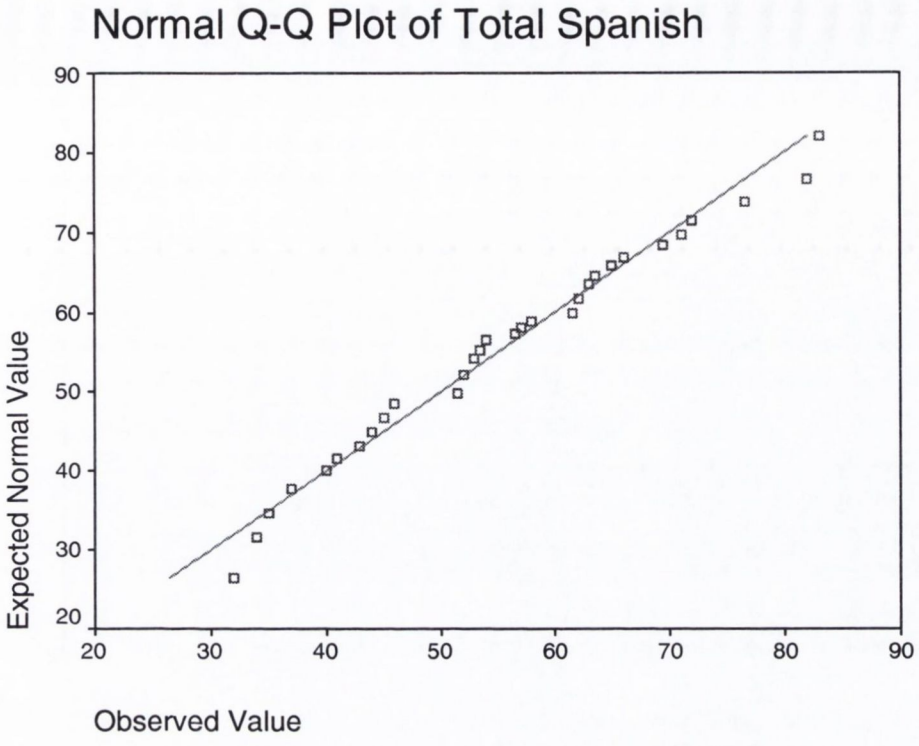
c. Dependent Variable: Language (Recep)

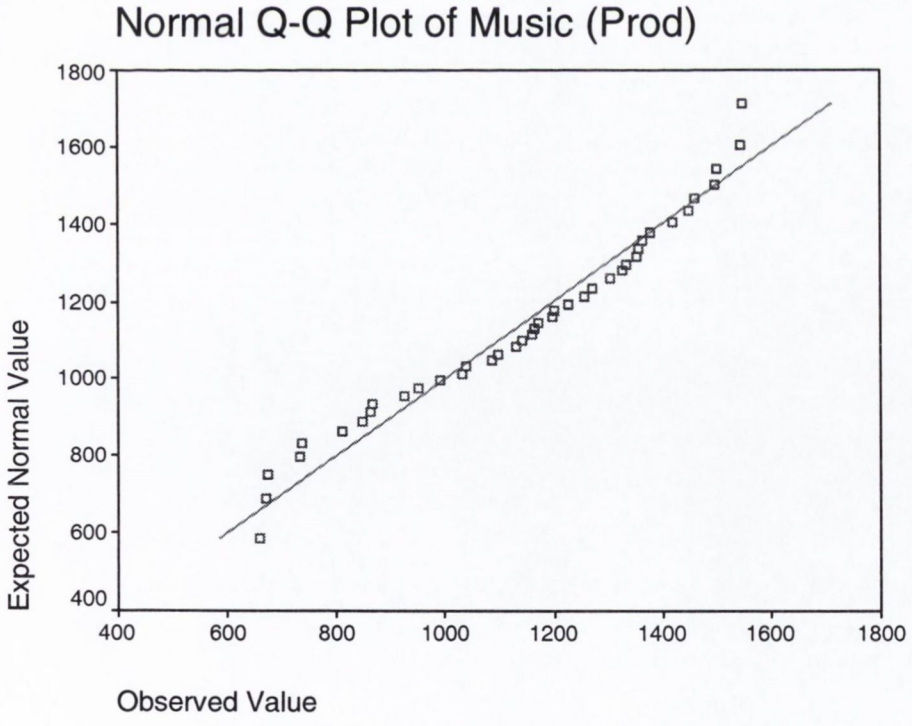
Appendix 2

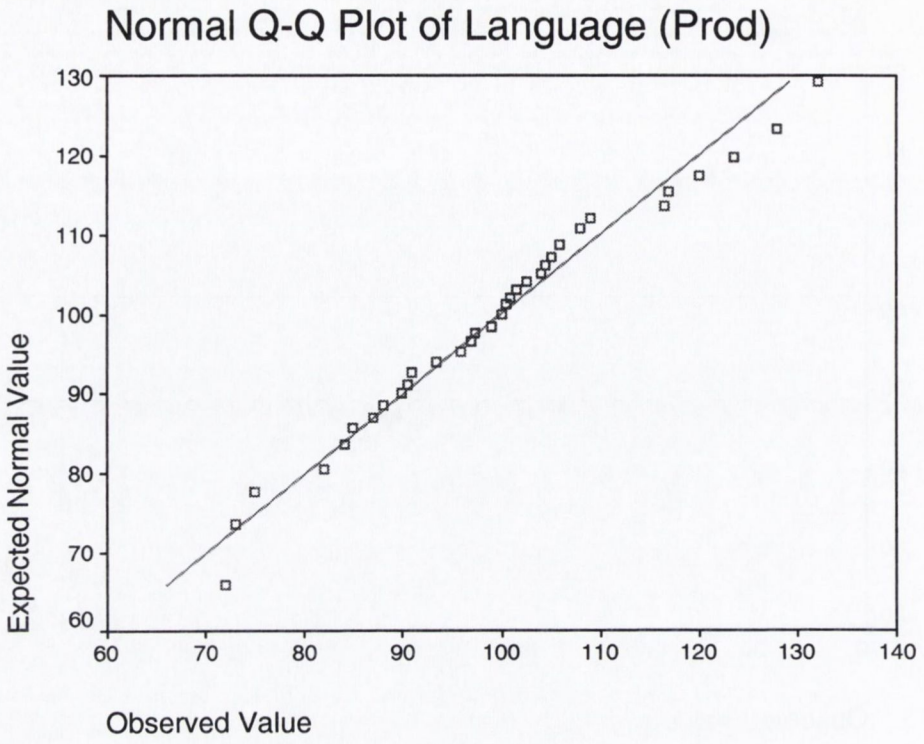
Normal Q-Q Plots

Normal Q-Q plots are provided for the main variables examined in Phase II, i.e. Korean, Spanish, Total Music (Prod), and Total Language (Prod). A Normal Q-Q plot examines how the observed values in the data compare to the values expected by the normal distribution. If observed values follow expected values exactly, points lie on a straight line.




Normal Q-Q Plot of Total Spanish
Expected Normal Value
Observed Value
90
80
70
60
50
40
30
20
20
30
40
50
60
70
80
90


Normal Q-Q Plot of Music (Prod)
Expected Normal Value
Observed Value
1800
1600
1400
1200
1000
800
600
400
400
600
800
1000
1200
1400
1600
1800


Normal Q-Q Plot of Language (Prod)
Expected Normal Value
Observed Value
130
120
110
100
90
80
70
60
60
70
80
90
100
110
120
130
140

Correlations

Pearson Correlations

		Korean	Spanish	Language (Prod)	Music (Prod)	Raven
Korean	Pearson Correlation	1	.283	.511**	.225	-.084
	Sig. (2-tailed)	.	.073	.001	.157	.603
	N	41	41	41	41	41
Spanish	Pearson Correlation	.283	1	.964**	.554**	.262
	Sig. (2-tailed)	.073	.	.000	.000	.098
	N	41	41	41	41	41
Language (Prod)	Pearson Correlation	.511**	.964**	1	.553**	.230
	Sig. (2-tailed)	.001	.000	.	.000	.148
	N	41	41	41	41	41
Music (Prod)	Pearson Correlation	.225	.554**	.553**	1	.174
	Sig. (2-tailed)	.157	.000	.000	.	.278
	N	41	41	41	41	41
Raven	Pearson Correlation	-.084	.262	.230	.174	1
	Sig. (2-tailed)	.603	.098	.148	.278	.
	N	41	41	41	41	41

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

Correlations

		Spanish Rhythm	Spanish pronunciation	Music (Prod)
Spanish Rhythm	Pearson Correlation	1	.609**	.496**
	Sig. (2-tailed)	.	.000	.001
	N	41	41	41
Spanish pronunciation	Pearson Correlation	.609**	1	.521**
	Sig. (2-tailed)	.000	.	.000
	N	41	41	41
Music (Prod)	Pearson Correlation	.496**	.521**	1
	Sig. (2-tailed)	.001	.000	.
	N	41	41	41

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

Correlations

			Korean	Spanish	Language (Prod)	Music (Prod)	Raven
Spearman's rho	Korean	Correlation Coefficient	1.000	.241	.484**	.178	.028
		Sig. (2-tailed)	.	.129	.001	.266	.864
		N	41	41	41	41	41
	Spanish	Correlation Coefficient	.241	1.000	.950**	.538**	.272
		Sig. (2-tailed)	.129	.	.000	.000	.086
		N	41	41	41	41	41
	Language (Prod)	Correlation Coefficient	.484**	.950**	1.000	.490**	.211
		Sig. (2-tailed)	.001	.000	.	.001	.185
		N	41	41	41	41	41
	Music (Prod)	Correlation Coefficient	.178	.538**	.490**	1.000	.145
		Sig. (2-tailed)	.266	.000	.001	.	.366
		N	41	41	41	41	41
	Raven	Correlation Coefficient	-.028	.272	.211	.145	1.000
		Sig. (2-tailed)	.864	.086	.185	.366	.
		N	41	41	41	41	41

**. Correlation is significant at the .01 level (2-tailed).

Correlations

			Spanish Rhythm	Spanish pronunciation	Music (Prod)
Spearman's rh	Spanish Rhythm	Correlation Coeffici	1.000	.619**	.469**
		Sig. (2-tailed)	.	.000	.002
		N	41	41	41
	Spanish pronunciation	Correlation Coeffici	.619**	1.000	.533**
		Sig. (2-tailed)	.000	.	.000
		N	41	41	41
	Music (Prod)	Correlation Coeffici	.469**	.533**	1.000
		Sig. (2-tailed)	.002	.000	.
		N	41	41	41

**. Correlation is significant at the .01 level (2-tailed).

Group divided by Raven scores

Group Statistics

	AAS vs. BAI	N	Mean	Std. Deviation	Std. Error Mean
Korean	BAI	18	41.17	3.808	.898
	AAS	23	40.00	4.134	.862
Total Spanish	BAI	18	48.917	8.3776	1.9746
	AAS	23	58.348	14.2609	2.9736
Language (Prod)	BAI	18	92.417	9.3593	2.2060
	AAS	23	101.565	16.7612	3.4950
Music (Prod)	BAI	18	1082.6389	303.02298	71.42320
	AAS	23	1197.0109	214.65114	44.75786

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Korean	Equal variances assumed	.021	.885	.928	39	.359	1.17	1.257	-1.376	3.710
	Equal variances not assumed			.938	37.905	.354	1.17	1.244	-1.353	3.686
Total Spanish	Equal variances assumed	5.538	.024	-2.486	39	.017	-9.431	3.7936	-17.1044	-1.7579
	Equal variances not assumed			-2.642	36.496	.012	-9.431	3.5695	-16.6671	-2.1953
Language (Prod)	Equal variances assumed	5.108	.029	-2.073	39	.045	-9.149	4.4132	-18.0750	-.2221
	Equal variances not assumed			-2.214	35.691	.033	-9.149	4.1329	-17.5331	-.7640
Music (Prod)	Equal variances assumed	4.847	.034	-1.414	39	.165	-114.3720	80.85705	-277.921	49.17682
	Equal variances not assumed			-1.357	29.463	.185	-114.3720	84.28843	-286.644	57.89972

Correlations

Raven Category			Korean	Total Spanish	Language (Prod)	Music (Prod)	Raven
BAI	Korean	Pearson (r)	1	.027	.458	-.104	-.076
		Sig. (2-tailed)	.	.915	.056	.681	.763
		N	18	18	18	18	18
	Total Spanish	Pearson (r)	.027	1	.899**	.581*	-.029
		Sig. (2-tailed)	.915	.	.000	.011	.910
		N	18	18	18	18	18
	Language (Prod)	Pearson (r)	.458	.899**	1	.440	-.055
		Sig. (2-tailed)	.056	.000	.	.067	.829
		N	18	18	18	18	18
	Music (Prod)	Pearson (r)	-.104	.581*	.440	1	.133
		Sig. (2-tailed)	.681	.011	.067	.	.598
		N	18	18	18	18	18
	Raven	Pearson (r)	-.076	-.029	-.055	.133	1
		Sig. (2-tailed)	.763	.910	.829	.598	.
		N	18	18	18	18	18
AAS	Korean	Pearson (r)	1	.520*	.669**	.652**	.105
		Sig. (2-tailed)	.	.011	.000	.001	.635
		N	23	23	23	23	23
	Total Spanish	Pearson (r)	.520*	1	.977**	.562**	-.087
		Sig. (2-tailed)	.011	.	.000	.005	.694
		N	23	23	23	23	23
	Language (Prod)	Pearson (r)	.669**	.977**	1	.659**	-.053
		Sig. (2-tailed)	.000	.000	.	.001	.811
		N	23	23	23	23	23
	Music (Prod)	Pearson (r)	.652**	.562**	.659**	1	-.069
		Sig. (2-tailed)	.001	.005	.001	.	.753
		N	23	23	23	23	23
	Raven	Pearson (r)	.105	-.087	-.053	-.069	1
		Sig. (2-tailed)	.635	.694	.811	.753	.
		N	23	23	23	23	23

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

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

Regression

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Raven, Music (Prod)[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: Language (Prod)

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.570[a]	.325	.289	12.3020

a. Predictors: (Constant), Raven, Music (Prod)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2764.019	2	1382.009	9.132	.001[a]
	Residual	5750.884	38	151.339		
	Total	8514.902	40			

a. Predictors: (Constant), Raven, Music (Prod)

b. Dependent Variable: Language (Prod)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	60.204	8.955		6.723	.000
	Music (Prod)	2.968E-02	.008	.529	3.910	.000
	Raven	7.778E-02	.076	.138	1.021	.314

a. Dependent Variable: Language (Prod)

Outliers omitted

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	Raven, Music (Prod)[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: Language (Prod)

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.573[a]	.329	.290	11.3100

a. Predictors: (Constant), Raven, Music (Prod)

b. Dependent Variable: Language (Prod)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	2191.212	2	1095.606	8.565	.001[a]
	Residual	4477.031	35	127.915		
	Total	6668.243	37			

a. Predictors: (Constant), Raven, Music (Prod)

b. Dependent Variable: Language (Prod)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	61.645	8.925		6.907	.000
	Music (Prod)	2.756E-02	.008	.510	3.645	.001
	Raven	.114	.080	.199	1.422	.164

a. Dependent Variable: Language (Prod)

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	80.933	109.448	97.487	7.6956	38
Std. Predicted Value	-2.151	1.554	.000	1.000	38
Standard Error of Predicted Value	2.0374	5.2211	3.0692	.8348	38
Adjusted Predicted Value	80.140	107.577	97.581	7.8401	38
Residual	-19.688	22.438	.000	11.0000	38
Std. Residual	-1.741	1.984	.000	.973	38
Stud. Residual	-1.827	2.038	-.004	1.015	38
Deleted Residual	-21.688	23.689	-.095	11.9846	38
Stud. Deleted Residual	-1.893	2.140	-.003	1.032	38
Mahal. Distance	.227	6.911	1.947	1.666	38
Cook's Distance	.000	.128	.030	.035	38
Centered Leverage Value	.006	.187	.053	.045	38

a. Dependent Variable: Language (Prod)

Stepwise regression

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	Music (Prod)	.	Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100).

a. Dependent Variable: Language (Prod)

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.538[a]	.290	.270	11.4695

a. Predictors: (Constant), Music (Prod)

b. Dependent Variable: Language (Prod)

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1932.505	1	1932.505	14.690	.000[a]
	Residual	4735.739	36	131.548		
	Total	6668.243	37			

a. Predictors: (Constant), Music (Prod)

b. Dependent Variable: Language (Prod)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	64.352	8.843		7.277	.000
	Music (Prod)	2.909E-02	.008	.538	3.833	.000

a. Dependent Variable: Language (Prod)

Excluded Variables[b]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
1	Raven	.199[a]	1.422	.164	.234	.980

a. Predictors in the Model: (Constant), Music (Prod)

b. Dependent Variable: Language (Prod)

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	83.501	109.377	97.487	7.2270	38
Std. Predicted Value	-1.935	1.645	.000	1.000	38
Standard Error of Predicted Value	1.8607	4.0959	2.5463	.6722	38
Adjusted Predicted Value	82.681	110.523	97.530	7.2883	38
Residual	-18.832	25.031	.000	11.3134	38
Std. Residual	-1.642	2.182	.000	.986	38
Stud. Residual	-1.704	2.212	-.002	1.014	38
Deleted Residual	-20.282	25.721	-.043	11.9515	38
Stud. Deleted Residual	-1.752	2.347	.002	1.032	38
Mahal. Distance	.000	3.745	.974	1.083	38
Cook's Distance	.000	.152	.028	.035	38
Centered Leverage Value	.000	.101	.026	.029	38

a. Dependent Variable: Language (Prod)

Appendix 3

Correlations

Pearson Correlations

Correlations

		Language (Prod)	Music (Prod)	Music (Recep)	Language (Recep)
Language (Prod)	Pearson Correlation	1	.553**	.181	-.062
	Sig. (2-tailed)	.	.000	.257	.701
	N	41	41	41	41
Music (Prod)	Pearson Correlation	.553**	1	.536**	.349*
	Sig. (2-tailed)	.000	.	.000	.025
	N	41	41	41	41
Music (Recep)	Pearson Correlation	.181	.536**	1	.592**
	Sig. (2-tailed)	.257	.000	.	.000
	N	41	41	41	41
Language (Recep)	Pearson Correlation	-.062	.349*	.592**	1
	Sig. (2-tailed)	.701	.025	.000	.
	N	41	41	41	41

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

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

Correlations

			Language (Prod)	Music (Prod)	Music (Recep)	Language (Recep)
Spearman's rho	Language (Prod)	Correlation Coefficien	1.000	.490**	.110	-.118
		Sig. (2-tailed)	.	.001	.492	.464
		N	41	41	41	41
	Music (Prod)	Correlation Coefficien	.490**	1.000	.527**	.385*
		Sig. (2-tailed)	.001	.	.000	.013
		N	41	41	41	41
	Music (Recep)	Correlation Coefficien	.110	.527**	1.000	.556**
		Sig. (2-tailed)	.492	.000	.	.000
		N	41	41	41	41
	Language (Recep)	Correlation Coefficien	-.118	.385*	.556**	1.000
		Sig. (2-tailed)	.464	.013	.000	.
		N	41	41	41	41

**. Correlation is significant at the .01 level (2-tailed).

*. Correlation is significant at the .05 level (2-tailed).

Group divided by Raven scores

Correlations

AAS vs. BAI			Music (Prod)	Music (Recep)	Language (Prod)	Language (Recep)
BAI	Music (Prod)	Pearson (r)	1	.756**	.440	.594**
		Sig. (2-tailed)	.	.000	.067	.009
		N	18	18	18	18
	Music (Recep)	Pearson (r)	.756**	1	.497*	.546*
		Sig. (2-tailed)	.000	.	.036	.019
		N	18	18	18	18
	Language (Prod)	Pearson (r)	.440	.497*	1	.016
		Sig. (2-tailed)	.067	.036	.	.949
		N	18	18	18	18
	Language (Recep)	Pearson (r)	.594**	.546*	.016	1
		Sig. (2-tailed)	.009	.019	.949	.
		N	18	18	18	18
AAS	Music (Prod)	Pearson (r)	1	.190	.659**	.028
		Sig. (2-tailed)	.	.384	.001	.899
		N	23	23	23	23
	Music (Recep)	Pearson (r)	.190	1	-.040	.622**
		Sig. (2-tailed)	.384	.	.856	.002
		N	23	23	23	23
	Language (Prod)	Pearson (r)	.659**	-.040	1	-.201
		Sig. (2-tailed)	.001	.856	.	.359
		N	23	23	23	23
	Language (Recep)	Pearson (r)	.028	.622**	-.201	1
		Sig. (2-tailed)	.899	.002	.359	.
		N	23	23	23	23

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

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

Appendix 4

Bentley Measures of Musical Aptitude

Reproduced from Bentley (1966, p.76*ff.*).

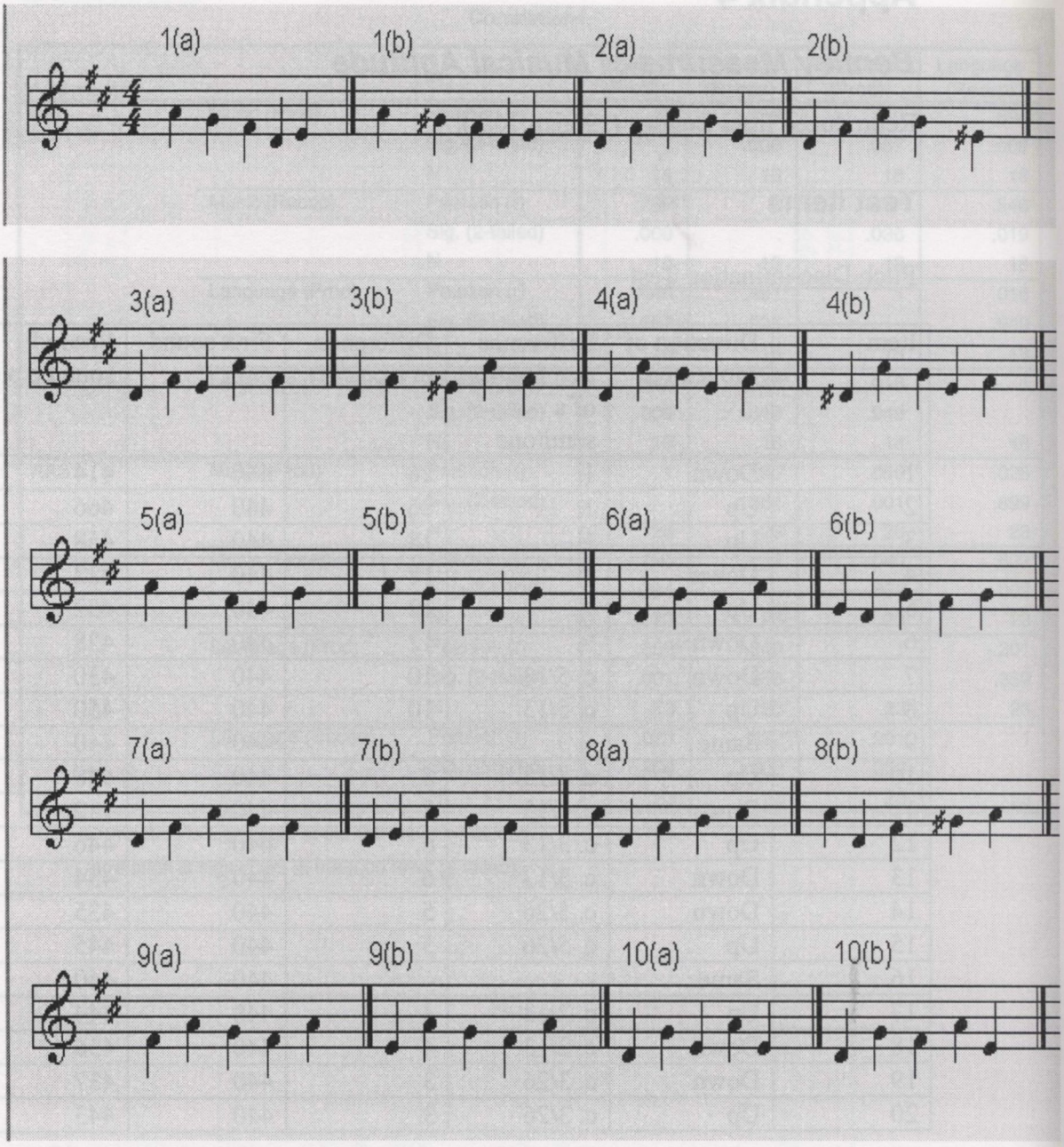
Test items

Pitch Discrimination Test

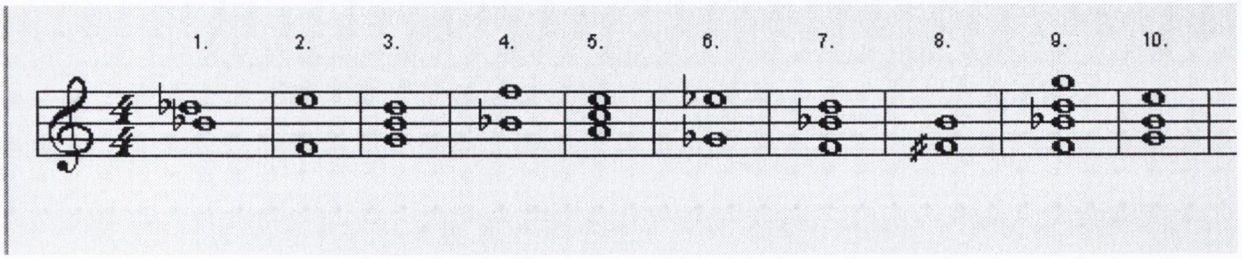
Item	Direction of Movement	Difference as a fraction of a semitone	Difference in c.p.s.[28]	First sound c.p.s	Second sound c.p.s.
1	Down	1	26	440	414
2	Up	1	26	440	466
3	Up	¾	18	440	458
4	Down	¾	18	440	422
5	Up	½	12	440	452
6	Down	½	12	440	428
7	Down	c. 5/13	10	440	430
8	Up	c. 5/13	10	440	450
9	Same	-	-	440	440
10	Up	c. 4/13	8	440	448
11	Down	c. 4/13	8	440	432
12	Up	c. 3/13	6	440	446
13	Down	c. 3/13	6	440	434
14	Down	c. 5/26	5	440	435
15	Up	c. 5/26	5	440	445
16	Same	-	-	440	440
17	Up	c. 2/13	4	440	444
18	Down	c. 2/13	4	440	436
19	Down	c. 3/26	3	440	437
20	Up	c. 3/26	3	440	443

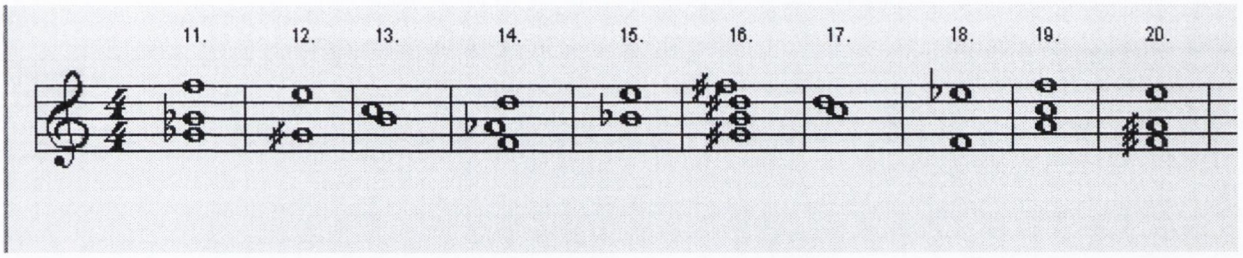
[28] c.p.s. is Cycles per Second

Tonal Memory Test

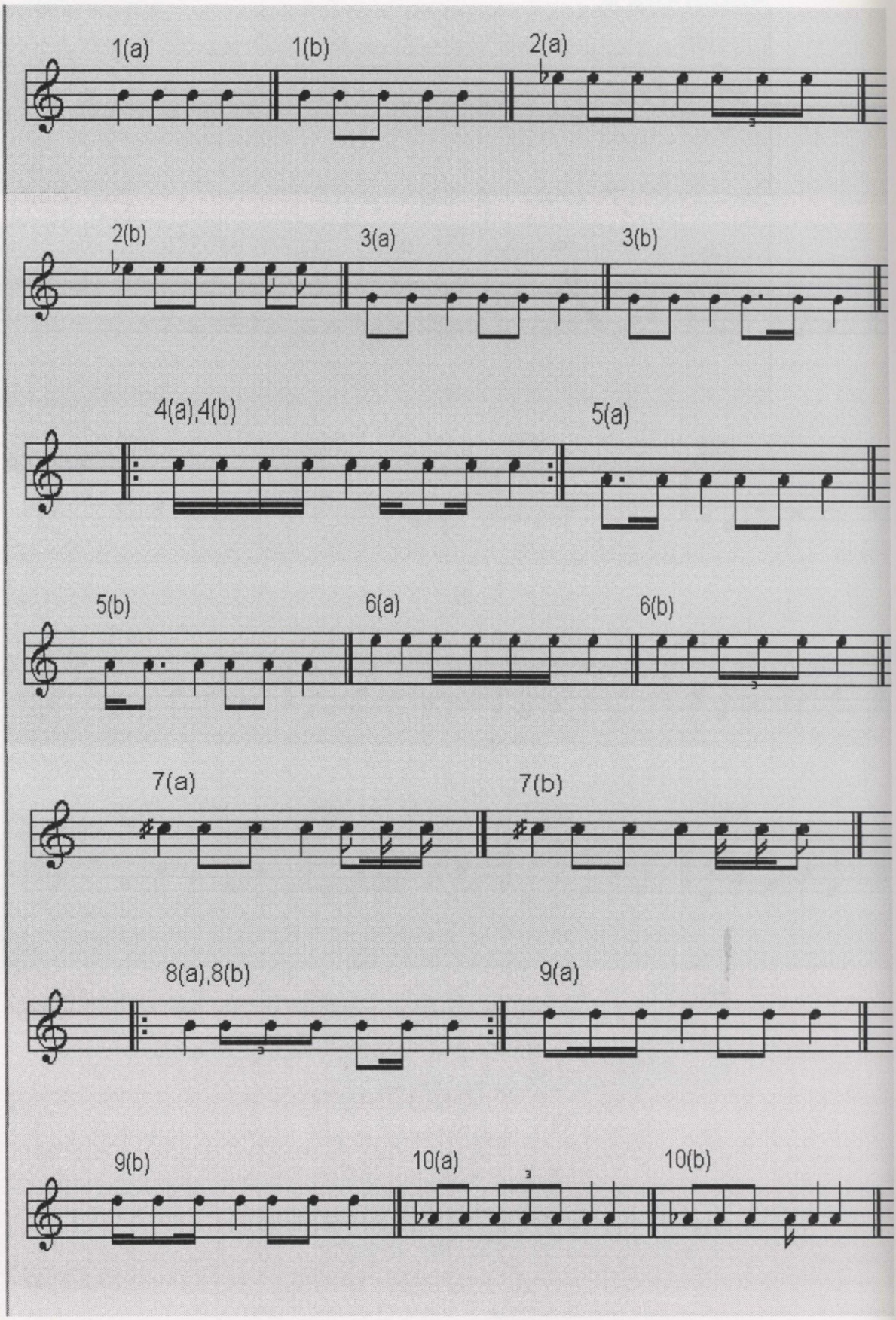


Chord Analysis Test





Rhythmic Memory Test



Answer sheet

Measures of Musical Abilities

Name _____

Age _____ years _____ months

Circle as appropriate

Gender m f

Year 1st 2nd 3rd

Left-handed Right-handed

1. PITCH	
1.	
2.	
3.	
4.	
5.	
6.	
7.	
8.	
9.	
10.	
11.	
12.	
13.	
14.	
15.	
16.	
17.	
18.	
19.	
20.	

2. TUNES	
1.	
2.	
3.	
4.	
5.	
6.	
7.	
8.	
9.	
10.	

3. CHORDS	
1.	
2.	
3.	
4.	
5.	
6.	
7.	
8.	
9.	
10.	
11.	
12.	
13.	
14.	
15.	
16.	
17.	
18.	
19.	
20.	

4. RHYTHM	
1.	
2.	
3.	
4.	
5.	
6.	
7.	
8.	
9.	
10.	

Transcript of Instructions

Taken from Bentley (1966, p.72*ff.*).

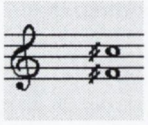
Pitch Discrimination Test

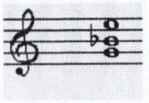
"Test number one – pitch. Listen to these two sounds" (item 2 – semitone up); "the second sound is higher than the first and has moved 'up'. Listen to these two sounds" (item 1 – semitone down); "now the second sound is lower than the first; it has moved 'down'. The next two sounds" (item 9 – 'same') "are the same. Some of the sounds you will hear are much closer together than you might expect. Listen to these" (item 12 – 6 c.p.s. difference up); "the second sound goes 'up' from the first. Now listen to one that goes 'down'" (item 13 – 6 c.p.s. difference down). "So; if the second sound goes up, write 'U'; if the second sound goes down, write 'D'. Is that clear? 'S' for same; 'U' for up; 'D' for down. I shall call out each number as we come to it."

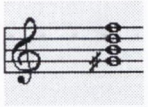
Tonal Memory Test

"Test number two – tunes. For each item two tunes will be played, like this" (second half of item 4, repeated). "If the second tune is the same as the first, as that was, write 'S'. If the second tune is not the same as the first, one note will have been changed. Listen to this example, and count the notes as they are played" (item 10). "In the second tune the third note is changed, and you would write the figure '3'. Listen to it again, and don't forget to count" (item 10). "So; if the third note is changed you will write the figure '3'; if the fourth note is changed you will write the figure '4'; if the second note is changed you will write the figure '2'; and so on. All the tunes have five notes; count them as they are played."

Chord Analysis Test

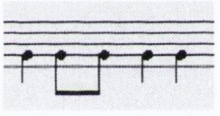
"Test number three – chords. You will hear chords; that means groups of notes played together. For example, here is a chord  containing two notes. Listen to the two notes played separately" (example); "and again, together, as a

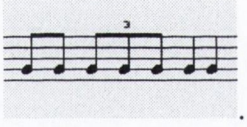
chord" (chord repeated). "Here is another chord  containing three notes. Listen to the three notes played separately" (example); "and together, as a chord" (chord repeated).

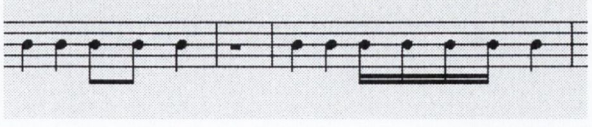
"Now listen to a chord containing four notes  ; here are the four notes played separately" (example); "and together, as a chord" (chord repeated). "In the test the notes will not be played separately; they will be played together, as chords. Listen carefully, and write down the number of notes you hear in each chord."

Rhythmic Memory Test

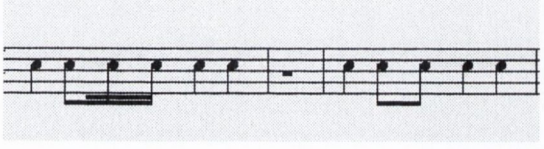
"Test number four – rhythm. You will hear two patterns of notes. Each pattern has four beats, or pulses, like this:

One two three four  ; or like this:

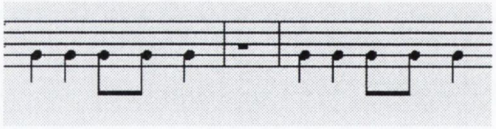
One two three four  . If the second pattern is the same as the first, write 'S' for 'same'; if the second pattern is different from the first, write down the number of the beat or pulse that is changed. Listen to this example, and see if you can decide which beat is changed:

One two three four 

Yes, the third beat was changed. Here is another:

One two three four 

There the second beat was changed. And another:

One two three four 

Those were the same. Your answer will be either 1 or 2 or 3 or 4 if there is a change, of S if the two patterns are the same. Now here is the test."

Receptive Language Aptitude Test

Test items

Sounds

1. BYLI	were (pl.)	BÍLÍ	white
2. VILA	villa	VÍLA	fairy
3. MÁMA	mother	NÁNA	silly girl
4. CHODÍVAT	to go	PODÍVAT	to look
5. DENNÍ	daily	DĚNÍ	events
6. BADATEL	researcher	PAKATEL	"for nothing"
7. NEROZUMÍ	doesn't understand	NEROZUMNÍ	not clever
8. CHRÁM	cathedral	KRÁM	little shop
9. VINA	fault	VINNÁ	wine
10. VLK	wolf	VLEK	ski-lift
11. ŽÍT	live	ŠÍT	knit
12. PŘIKROČIT	to approach	PŘEKROČIT	to climb over
13. BLÁTO	mud	DLÁTO	chisel
14. ZODPOVÍDAT	to be responsible for	ODPOVÍDAT	correspond to
15. VLKY	wolf	PLKY	gossip
16. VZPÍRAT	to weight-lift	SBÍRAT	to collect

Chinese

wū 屋 (house, room, dwelling)

wú 无 (without, un-, in-)

wǔ 五 (five)

wù 物 (object, thing, matter)

1. 这间屋子很小，我要搬到另一个地方住。

zhè jìan wūzǐ hěn xiǎo, wǒ yào bān dào lìng yī gè dìfāng zhù

this MEA room very small, me want move to other one MEA place live

`This room is too small. I want to move somewhere else.'

2. 他是无辜的，你别处罚他。

tā shì wúgū de, nǐ bíe chǔfá tā

him be innocent MOD, you don't punish him

`He's innocent, don't punish him.'

3. 水是一种物体。

shǔi shì yī zhǒng wùtǐ

water be one kind substance

`Water is a kind of substance'

4. 昨天我在学校里找到了五块钱。

zúotīan wǒ zài xúexiào lǐ zhǎodào le wǔ kuài qían

yesterday me at school in find PRF five MEA money

`I found 5 quid/bucks/pounds/dollars at school yesterday'

5. 他在家里无聊时，会上网玩电脑游戏。

tā zài jīalǐ wúliáo shí, hùi shàng wǎng wán dìannǎo yóuxì

him at home bored time, can on net play computer game

`If he's bored at home, he sometimes goes on the internet to play computer games.'

6. 为了庆祝新年，我们决定去动物园玩一天。

wèile qìngzhù xīnnían, wǒmen júedìg qù dòngwùyúan wán yī tīan

in-order-to celebrate new-year, we decide go zoo play one day

`We decided to go to the zoo for the day to celebrate the New Year's.'

7. 由于他家里的屋顶看起来要倒，所以他要去请人修理。

yóuyú tā jīal de wūdǐng kànqǐlái yào dǎosǔoyǐ tā yào qù qǐng rén xīulǐ

due-to him home MOD root look-like will collapse, so him want go request person repair

`He wants to get someone to repair his roof, because it looks like it will collapse.'

8. 他在大学读医科读了五年突然退学，真是半途而废。

tā zài dàxúe dú yīkē dú le wǔ nían tūrán tùixúe, zhēn shì bàntúerfèi

him at university read medicine read PRF five year sudden drop-out, really be do-something-by-halves

`He studied medicine for five years before dropping out all of a sudden. That's really doing something by halves.'

9. 我们今年有五个科目要学，真的很难应付。

wǒmen jīn nían yǒu wǔ gè kēmù yào xúe, zhēnde hěn nán yìngfu

we this year have five MEA subject have-to learn, really very hard cope-with

`We have five subjects to study this year – it's a lot to cope with.'

10. 他整天无所事事，在家里玩电脑游戏。

tā zhěng tīan wúsǔoshìshì, zài jīalǐ wán dìannǎo yóuxì

him whole day at-loose-end, at home play computer games

`He had nothing to do all day, and spent it playing computer games.'

MEA – measure word (aka classifier)

MOD – modifying particle

PRF – perfective particle

Numerical

1. dvacet (20)
2. čtyři (4)
3. pět set (500)
4. dvě stě (200)
5. tři (3)
6. šedesát (60)
7. tři sta (300)
8. pět (5)
9. třicet (30)
10. padesát (50)
11. čtyři sta (400)
12. dva / dvě (2)
13. šest set (600)
14. šest (6)
15. čtyřicet (40)

Grammatical Functions of Words in Sentences

Look at the following two sentences:

i. **Mary** likes to go to school.
ii. **(A) He** likes to go **(B) fishing** in **(C) Maine.**

In sentence ii., the word 'He' has the same grammatical function as the word 'Mary' in sentence i. They both act as the subject of the sentence – Mary likes, He likes.

You must decide which of the words in bold in sentence 2. fulfills the same grammatical function as the underlined word in sentence 1.

On the answer sheet provided, put a tick in one of the boxes marked A, B, C or D. Please attempt all questions.

1. i. **Sad** people cry.

   ii. **(A) Happy** people are **(B) often** seen **(C) laughing** and **(D) smiling.**

2. i. She hopes to buy a car **when** she gets the money.

   ii. **(A) After** you left at 6.00 **(B) most** of the group remained **(C) in** the **(D) pub**.

3. i. His friend bought **her** a new car.

   ii. Why won't **(A) you** tell **(B) me (C) something (D) yourself**?

4. i. The **basic** rules of the game are not hard to learn.

   ii. **(A) She** had great **(B) difficulty** in finding the **(C) proper** exit she should to **(D) through**.

5. i. It's not **to** be taken away from here.

   ii. She talked **(A) to** me **(B) about** how I need **(C) to** realise matters have gone **(D) too** far.

6. i. A **number** of people arrived at the theatre late.

ii. **(A) I** have experienced **(B) many** strange **(C) things** in **(D) my** life.

7. i. My foot became **sore** from the infection.
   ii. The plant **(A) quickly** grew **(B) strong** in the **(C) warmth** from the **(D) sun.**

8. i. The woman fell down and hurt **herself**.
   ii. **(A) You** know **(B) yourself** that she will win **(C) prizes** if she does **(D) your** work.

9. i. The playground was full of **screaming** children.
   ii. **(A) Usually** constant practice is the best **(B) method** for **(C) fast (D) learning**.

10. i. Jane has gone **to** make a telephone call.
    ii. **(A) Two** people will be needed **(B) to** carry this bed because it is **(C) too** heavy **(D) for** one.

11. i. **She** calls her mother Margaret.
    ii. **(A) Mary** telephones **(B) him** every **(C) day** according to **(D) her** sister.

12. i. I can see **you** from my front window.
    ii. **(A) You** can hear **(B) her (C) from** outside the **(D) door**.

13. i. **Money** seems to make her happy.
    ii. **(A) Years** ago most **(B) farming** was **(C) done** by **(D) hand**.

14. i. James sold **Liam** his bike.
    ii. **(A) I** will guarantee **(B) them** a **(C) huge (D) bonus**.

15. i. **She** was late.
    ii. Because of the **(A) huge** demand for the **(B) product** the

(C) manager is ordering more **(D) stock**.

16. i. I wrote **Paul** a letter.
    ii. **(A) She** promised **(B) them (C) immediately** that **(D) she** would be there.

17. i. She can run **fast** when she tries.
    ii. The **(A) girl** is **(B) upset** that she cannot see at all **(C) well** in her **(D) new** contact lenses.

18. i. The moon did **appear** after all.
    ii. When you are **(A) next** after some **(B) advice**, I might **(C) advise** you **(D) against** it.

19. i. **Rarely** does he go out these days.
    ii. I should like **(A) to** take the **(B) opportunity to (C) thank** you most **(D) sincerely**.

20. i. **Many** people saw the film.
    ii. **(A) Several (B) children** liked the **(C) new (D) book**.

Answer sheet

Sound Discrimination

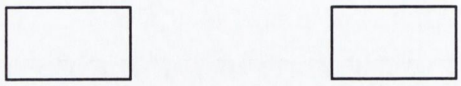
Name: _____

You will hear on tape a series of words in a language which is unfamiliar. These words are said in pairs. The first word you hear is called sound A and the second word you hear is called sound B. After a pause of a few seconds one of these words is repeated. If you think that the repeated word is A, put a tick in the box under the heading A. If you think that the repeated word is B, put a cross in the box under the heading B. There is a total of 16 pairs of words.

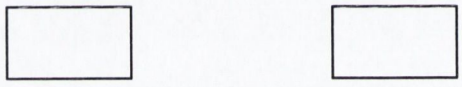
	A.	B.
1.	☐	☐
2.	☐	☐
3.	☐	☐
4.	☐	☐
5.	☐	☐
6.	☐	☐
7.	☐	☐
8.	☐	☐

9. 

10. 

11. 

12. 

13. 

14. 

15. 

16. 

Chinese Test

Listen to the instructions on the tape. Fill in your answers below.

1. _____
2. _____
3. _____
4. _____
5. _____
6. _____
7. _____
8. _____
9. _____
10. _____

Numerical Test

Listen to the instructions on the tape. Fill in your answers below.

1.	
2.	
3.	
4.	
5.	

6.	
7.	
8.	
9.	
10.	

11.	
12.	
13.	
14.	
15.	

Grammatical Functions of Words in Sentences

	A.	B.	C.	D.
1.				
2.				
3.				
4.				
5.				
6.				
7.				
8.				
9.				
10.				
11.				
12.				
13.				
14.				
15.				
16.				
17.				
18.				
19.				
20.				

Transcript of Instructions

Numerical

I'm going to teach you fifteen numbers in a language that is totally unfamiliar to you. You will learn by listening to me, repeating the numbers in English and the unfamiliar language. You are not allowed to write anything as you listen. Just make a mental note as we go along. I shall tell you when it is time for you to write anything down. First you will hear the numbers from two to six twice.

Two dva
Three tři
Four čtyři
Five pět
Six šest

The same numbers again.

Two dva
Three tři
Four čtyři
Five pět
Six šest

Two	dva	Twenty	dvacet
Three	tři	Thirty	třicet
Four	čtyři	Forty	čtyřicet
Five	pět	Fifty	padesát
Six	šest	Sixty	šedesát
Twenty	dvacet	Two Hundred	dvě stě
Thirty	třicet	Three Hundred	tři sta
Forty	čtyřicet	Four Hundred	čtyři sta
Fifty	padesát	Five Hundred	pět set
Sixty	šedesát	Six Hundred	šest set

Now I'm going to give you the foreign numbers jumbled up. Try to write the digits down on the paper in the boxes provided.

Chinese

I'm going to teach you four words in a language that is completely unfamiliar to you. You will learn by listening to me, repeating the words in English and in the unfamiliar language.

You are not allowed to write anything as you listen, just make a mental note as we go along. I shall tell you when it is time for you to write anything down.

Listen to the words:

The same words again:

Now you will hear the foreign words in a sentence. Try to write the English equivalent of the foreign word you hear.

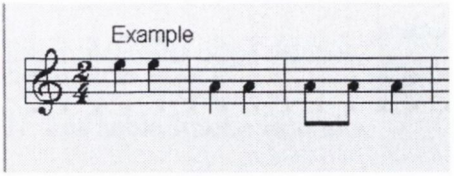
For example:

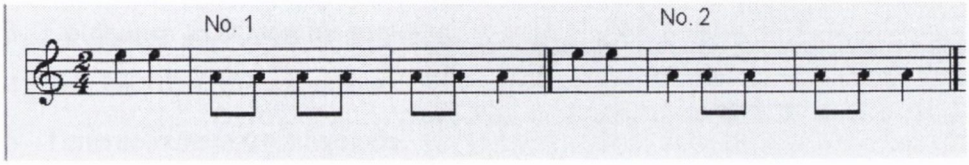
In this sentence you heard the word which means 'house' so you write 'house'.

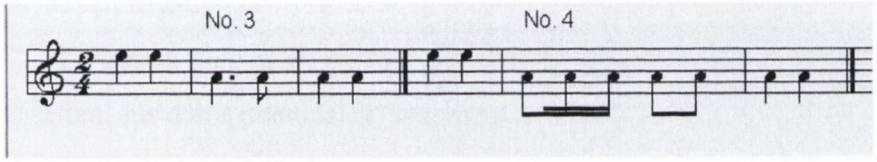
There are 10 sentences starting now.

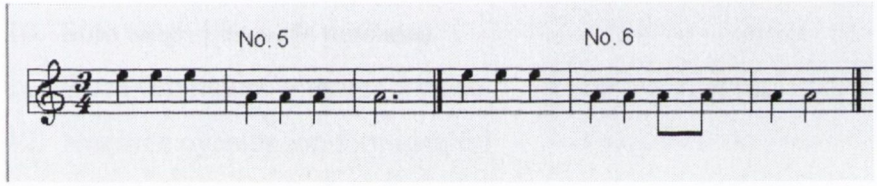
Appendix 5

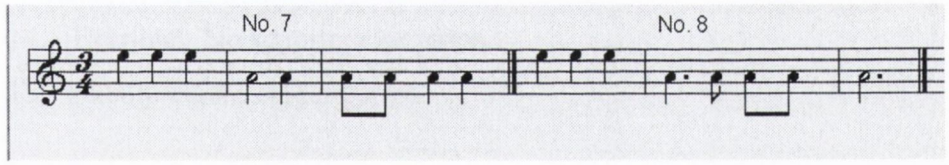
Music productive test

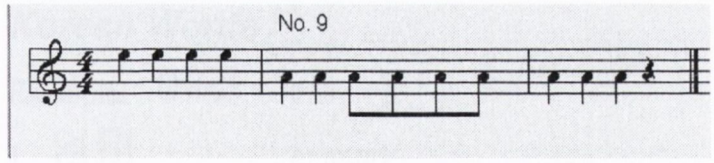


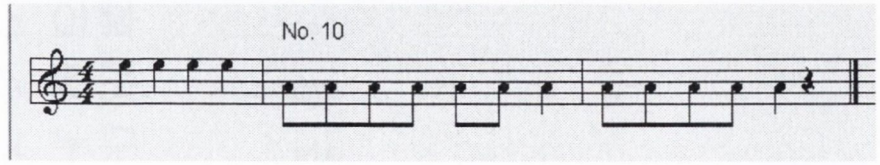


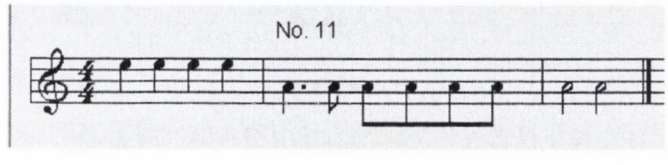


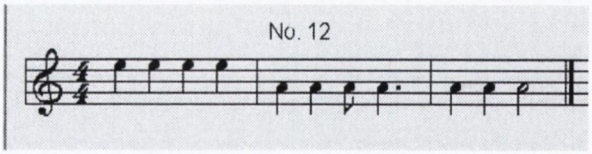


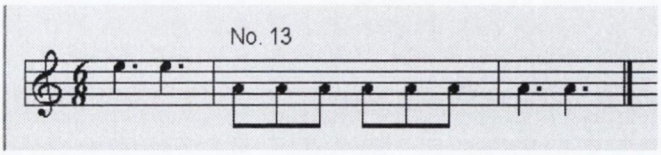


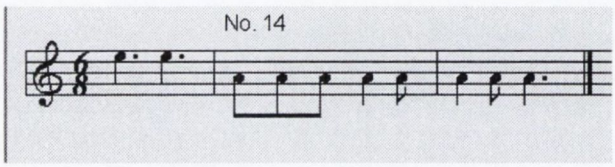


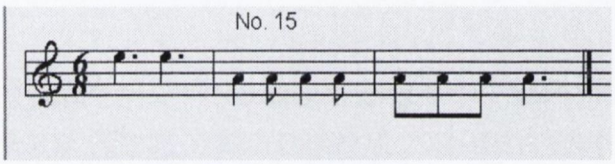


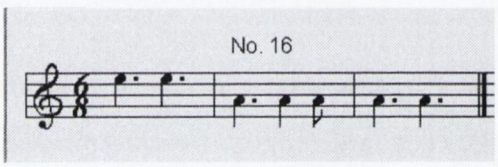



No. 12


No. 13


No. 14


No. 15


No. 16

Spanish sentences

Taken from Busco Empleo (Artés, 1985, p.16)

Example: Una edad estupenda.

1. Les habla Nacho Buendía.
2. Aquí Radio Enlace.
3. Colchones LUX son los mejores.
4. Son las 10,30 de la mañana.
5. Tenemos una nueva llamada.
6. A ver, dígame.
7. ¿Le gustan mucho los perros?
8. Bien, sus datos personales, por favor.
9. Me llamo Julia Lozano.
10. Sólo tengo libres las mañanas.
11. Entre veinticinco y treinta y cinco.
12. Nuestros oyentes son formidables!
13. ¿Qué le parece esto, Julia?
14. ¡Horrible! No soporto a los niños.
15. Yo solo quiero cuidar perros.

Korean Words

Example: 이것	this
1. 엄마	mummy
2. 아빠	daddy
3. 저것	that
4. 누구	who
5. 책	book
6. 옷	clothes
7. 의자	chair
8. 수건	towel
9. 오리	duck

10. 냉장고 fridge
11. 지하철 subway
12. 창문 window
13. 깡총 hop

Appendix 6

Korean rater instructions

Each CD has a letter on it, i.e. CD-H, CD-I, CD-J, CD-K, CD-L, CD-M, CD-N.

Each answer sheet has a corresponding letter in the top left hand corner, i.e. CD-H, CD-I, CD-J, CD-K, CD-L, CD-M, CD-N.

Each answer sheet lists all the tracks that are on the corresponding CD. The tracks are labelled 1 to 98, except in the case of CD-N where there are only 18 tracks.

A single track consists of one student uttering one Korean word twice, i.e. the student said the word once and then it was copied so that you may hear it twice before making a judgement.

There are 14 words in total. Before a new word is introduced, you will hear the native speaker pronounce it. This is marked on the answer sheet as No1word1, No2word2, No3word3, etc. After hearing the native speaker say the word twice, you hear each of the students say the word twice.

Your task is to give a score to each track based on the following:

Pronunciation Score

5. *Perfect repetition of word as native Korean speaker would have pronounced it.*

4. *Near-perfect repetition of the word.*

3. *A reasonable repetition of word.*

2. *Less than half the word correct.*

1. *The repetition bore little resemblance to the original.*

Thank you!

The following sentences were added for Rater 2: "*You may use half scores if necessary. You can use 1.5, 2.5, 3.5, 4.5 if necessary*".

Spanish rater instructions

The track number is the number that will appear on your CD player. Please ensure that when you give a score on the answer sheet, the track number on the CD player is the same as the track number on the answer sheet.

The subject number is simply an identification number assigned to each subject.

Pronunciation Score

Please give a score for pronunciation. You may give a mark between 1 and 5.

5. *Perfect repetition of sentence as native Spanish speaker would have pronounced it.*

4. *Near-perfect repetition of the sentence.*

3. *A reasonable repetition of sentence –more than half correct but part incorrect, foreign accent.*

2. *Less than half the sentence correct. Many of the vowels and consonants poorly pronounced.*

1. *The repetition bore little resemblance to the original.*

Rhythm Score

Please give a score for rhythm. Does the sentence sound like a Spanish sentence, i.e. even if the words are wrong, the overall rhythm could resemble Spanish.

5. *Perfect native-speaker-like rhythm.*

4. *Near-perfect rhythm.*

3. *Reasonable attempt to imitate rhythm.*

2. *Some attempt to imitate rhythm.*

1. *Little resemblance to Spanish native-speaker rhythm.*

Thank you!

The following sentences were added for Rater 2: "*You may use half scores if necessary. You can use 1.5, 2.5, 3.5, 4.5 if necessary.*

Appendix 7

Calculation of Kappa statistic for Korean raters

Korean rater 1

Algorithm

The following algorithm, written in Perl, was used to calculate Krippendorff's alpha for Korean Rater (Devitt, 2005). It is modified slightly for Rater 2 to take account of the fact that Rater 2 had the possibility of using 0.5 divisions of the scale.

```
#!usr/bin/perl -w

use strict;

my ($infile, $outfile);

if($#ARGV!=1){
    die "Usage: perl $0 <input file> <outfile>\n";
}

else {
    $infile=$ARGV[0];
    $outfile=$ARGV[1];
}

open(FULL, ">$outfile.all") or die "Couldn't open $outfile.all:
$!\n";
open(OUT, ">$outfile.metric") or die "Couldn't open
$outfile.metric: $!\n";

my @fulltable=();
## 1=2, 3=4, 5=4, 7=6, 9=8
#my @evenValsTable=();
# for(my $i=-1;$i<=4;$i++){
#     for(my $j=-1;$j<=4;$j++){
#     $fulltable[$i+1][$j+1]=0;
#     }
# }
for(my $i=0;$i<=4;$i++){
    for(my $j=0;$j<=4;$j++){
      $fulltable[$i][$j]=0;
    }
}


## READ FILES ##
open(SIM, "$infile")  or die "Couldn't open results file $infile:
$!\n";
my $sentence=0;
    ## $grouptable[story][user]=value
my @grouptable=();
my $numberOfRespondentsPerStory=2;
while (defined (my $line = <SIM>)) {
    chomp $line;
    my @vals=split(/\s+/ , $line);
    for(my $user=0;$user<=$#vals;$user++){
```

```
      $grouptable[$sentence][$user]=$vals[$user];
    }
    $sentence++;
}
close SIM;
## foreach sentence
for(my $i=0;$i<=$#grouptable;$i++){
    ## for each rater, note their agreement with every other rater
    for(my $j=0;$j<=$#{ $grouptable[$i] };$j++){
      for(my $k=0;$k<=$#{ $grouptable[$i] };$k++){
          if($j!=$k){
            my $val1 = $grouptable[$i][$j];
            my $val2 = $grouptable[$i][$k];
            my $frac = 1/$numberOfRespondentsPerStory;
            ## IGNORE MISSING VALUES
            if($val1!=-1 && $val2!=-1){

$fulltable[$val1][$val2]=$fulltable[$val1][$val2]+$frac;
            }
          }
      }
    }
}

my @totals=();
my $fullTotal=0;
for(my $i=0;$i<=$#fulltable;$i++){
    if($i==0){
      print FULL "\t0\t1\t2\t3\t4\n";
    }
    for(my $j=0;$j<=$#fulltable;$j++){
      if($j==0){
          print FULL "$i";
      }
      print FULL "\t$fulltable[$i][$j]";
      if($i<=$#fulltable && $j<=$#fulltable){
          $totals[$i]+=$fulltable[$i][$j];
      }
    }
    print FULL "\n";
    if($i<=$#fulltable){
      $fullTotal+=$totals[$i];
    }
}

#close FULL;

#open(OUT, ">$outfile.metric") or die "Couldn't open
$outfile.metric: $!\n";

my @outMatrix=();
for(my $i=0; $i<=$#fulltable; $i++){
    for(my $j=0; $j<=$#{ $fulltable[$i] }; $j++){
      if($j<=$i){
          $outMatrix[$i][$j]=0;
      }
      else {
          my $ans=($totals[$i]/2)+($totals[$j]/2);
          for(my $k=$i+1;$k<$j;$k++){
            $ans+=$totals[$k];
          }
```

```
            $ans=$ans*$ans;
            $outMatrix[$i][$j]=$ans;
        }
    }
}

### print coincidence matrix
print OUT "COINCIDENCE MATRIX\n\n";
for(my $i=0;$i<=$#fulltable;$i++){
    if($i==0){
      print OUT "\t0\t1\t2\t3\t4\n";
    }
    for(my $j=0;$j<=$#{ $fulltable[$i] };$j++){
      if($j==0){
          my $numb=$i;
          print OUT "$numb";
      }
      printf OUT "\t%.2f", $fulltable[$i][$j];
    }
    print OUT "\n";
}

## print matrix of differenc metrics for each value
print OUT "MATRIX OF DIFFERENCE METRICS\n\n";
for(my $i=0;$i<=$#outMatrix;$i++){
    if($i==0){
      print OUT "\t0\t1\t2\t3\t4\n";
    }
    for(my $j=0;$j<=$#outMatrix;$j++){
      if($j==0){
          my $numb=$i+1;
          print OUT "$numb";
      }
      if($j<$i){
          printf OUT "\t%.2f", sqrt($outMatrix[$j][$i]);
      }
      else {
          printf OUT "\t%.2f", $outMatrix[$i][$j];
      }
    }
    print OUT "\n";
}

print OUT "\n\n";


my $calc=0;
my $top=0;
my $bottom=0;
for(my $i=0; $i<=4; $i++){
    for(my $j=$i; $j<=4; $j++){
      if($outMatrix[$i][$j]!=0){
          $top=$top + ($fulltable[$i][$j]*$outMatrix[$i][$j]);
#         print STDERR "TOP $i $j: $fulltable[$i][$j] *
$outMatrix[$i][$j]\n";

$bottom=$bottom+($totals[$i]*$totals[$j]*$outMatrix[$i][$j]);
#         print STDERR "BOT $i $j: $totals[$i] * $totals[$j] *
$outMatrix[$i][$j]\n";
      }
#     else {
```

```
#          print STDERR "skipped $i $j\n";
#      }
    }
}
print OUT "\nKAPPA COEFFICIENT CALCULATIONS\n";
print OUT "\n **** 1- ((n-1)(Sum of (values * difference
metric))****\n";
print OUT " ****    ----------------------------------------
****\n";
print OUT " ****         Sum of (totals*total*diffMetric)
****\n";

print OUT "full total (n): $fullTotal\n";
print OUT "top / bottom: $top / $bottom\n";
$calc=$fullTotal-1;
$calc=$calc*$top;
print OUT "(n-1) * top = $calc\n";
$calc=$calc/$bottom;
print OUT "(n-1) * top / bottom = $calc\n";
$calc=1-$calc;
printf OUT "Kappa: %.4f\n", $calc;

close FULL;
close OUT;
```

Items examined

Korean Rater1 examined 13 items twice. The following is the list of items which were examined twice. The item code consists of the subject code (e.g. a117) followed by the word in question (e.g. s2 for word 2). Time1 is the score awarded on the first occasion while Time2 is the score awarded on the second occasion.

Item	time1	time2
a117s1	3.00	3.00
a117s2	2.00	2.00
a117s3	3.00	3.00
a117s4	3.00	3.00
a117s5	4.00	3.00
a117s6	2.00	2.00
a117s7	3.00	2.00
a117s8	4.00	4.00
a117s9	3.00	3.00
a117s10	4.00	3.00
a117s11	3.00	2.00
a117s12	2.00	2.00
a117s14	5.00	2.00

Output of algorithm

Adding the numbers in the Coincidence Matrix should give the number of items examined, i.e. 3 + 1 + 0.50 + 1 + 4 + 1 + 1 + 1 + 0.50 = 13. This is correct as it is shown above that 13 items were examined twice.

```
COINCIDENCE MATRIX

      0     1     2     3     4     5     6     7     8
0     0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
1     0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
2     0.00  0.00  3.00  1.00  0.00  0.50  0.00  0.00  0.00
3     0.00  0.00  1.00  4.00  1.00  0.00  0.00  0.00  0.00
4     0.00  0.00  0.00  1.00  1.00  0.00  0.00  0.00  0.00
5     0.00  0.00  0.50  0.00  0.00  0.00  0.00  0.00  0.00
6     0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
7     0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
8     0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
MATRIX OF DIFFERENCE METRICS

      1     1.5   2     2.5   3     3.5   4     4.5   5
1     0     0     5.06  56.25 132.25      162.56 169   169   169
2     0     0     5.06  56.25 132.25      162.56       169   169
      169
3     2.25  2.25  0.00  27.56 85.56 110.25      115.56      115.56
      115.56
4     7.50  7.50  5.25  0.00  16.00 27.56 30.25 30.25 30.25
5     11.50 11.50 9.25  4.00  0.00  1.56  2.25  2.25  2.25
6     12.75 12.75 10.50 5.25  1.25  0.00  0.06  0.06  0.06
7     13.00 13.00 10.75 5.50  1.50  0.25  0.00  0.00  0.00
8     13.00 13.00 10.75 5.50  1.50  0.25  0.00  0.00  0.00
9     13.00 13.00 10.75 5.50  1.50  0.25  0.00  0.00  0.00


KAPPA COEFFICIENT CALCULATIONS

 **** 1- ((n-1)(Sum of (values * difference metric))****
 ****    ---------------------------------------- ****
 ****         Sum of (totals*total*diffMetric)        ****
full total (n): 13
top / bottom: 98.6875 / 2038.5625
(n-1) * top = 1184.25
(n-1) * top / bottom = 0.580924058006561
Kappa: 0.4191
```

Korean rater 2

Algorithm

The algorithm takes account of 10 values rather than 5 above as Rater 2 had the possibility of using 0.5 divisions of the scale. Thus rather than having 5 values - 1 through 5, there are now 10 values, i.e. 1, 1.5, 2, 2.5, 3, 3.5, … 10. Before running the program, the Rater's scores were multiplied by 2 in order to convert them to whole numbers for convenience, e.g. 1.5 becomes 3. Kappa was then calculated on the whole number values.

```
#!usr/bin/perl -w

use strict;

my ($infile, $outfile);

if($#ARGV!=1){
```

```
    die "Usage: perl $0 <input file> <outfile>\n";
}

else {
    $infile=$ARGV[0];
    $outfile=$ARGV[1];
}

open(FULL, ">$outfile.all") or die "Couldn't open $outfile.all:
$!\n";
open(OUT, ">$outfile.metric") or die "Couldn't open
$outfile.metric: $!\n";

my @fulltable=();
## 1=2, 3=4, 5=4, 7=6, 9=8
#my @evenValsTable=();
# for(my $i=-1;$i<=4;$i++){
#     for(my $j=-1;$j<=4;$j++){
#     $fulltable[$i+1][$j+1]=0;
#     }
# }
for(my $i=0;$i<=9;$i++){
    for(my $j=0;$j<=9;$j++){
      $fulltable[$i][$j]=0;
    }
}


## READ FILES ##
open(SIM, "$infile")  or die "Couldn't open results file $infile:
$!\n";
my $sentence=0;
    ## $grouptable[story][user]=value
my @grouptable=();
my $numberOfRespondentsPerStory=2;
while (defined (my $line = <SIM>)) {
    chomp $line;
    my @vals=split(/\s+/ , $line);
    for(my $user=0;$user<=$#vals;$user++){
      $grouptable[$sentence][$user]=$vals[$user];
    }
    $sentence++;
}
close SIM;
## foreach sentence
for(my $i=0;$i<=$#grouptable;$i++){
    ## for each rater, note their agreement with every other rater
    for(my $j=0;$j<=$#{ $grouptable[$i] };$j++){
      for(my $k=0;$k<=$#{ $grouptable[$i] };$k++){
          if($j!=$k){
            my $val1 = $grouptable[$i][$j];
            my $val2 = $grouptable[$i][$k];
            my $frac = 1/$numberOfRespondentsPerStory;
            ## IGNORE MISSING VALUES
            if($val1!=-1 && $val2!=-1){

$fulltable[$val1][$val2]=$fulltable[$val1][$val2]+$frac;
            }
          }
      }
    }
```

```
}

my @totals=();
my $fullTotal=0;
for(my $i=0;$i<=$#fulltable;$i++){
    if($i==0){
      print FULL "\t0\t1\t2\t3\t4\t5\t6\t7\t8\t9\n";
    }
    for(my $j=0;$j<=$#fulltable;$j++){
      if($j==0){
          print FULL "$i";
      }
      print FULL "\t$fulltable[$i][$j]";
      if($i<=$#fulltable && $j<=$#fulltable){
          $totals[$i]+=$fulltable[$i][$j];
      }
    }
    print FULL "\n";
    if($i<=$#fulltable){
      $fullTotal+=$totals[$i];
    }
}

#close FULL;

#open(OUT, ">$outfile.metric") or die "Couldn't open
$outfile.metric: $!\n";

my @outMatrix=();
for(my $i=0; $i<=$#fulltable; $i++){
    for(my $j=0; $j<=$#{ $fulltable[$i] }; $j++){
      if($j<=$i){
          $outMatrix[$i][$j]=0;
      }
      else {
          my $ans=($totals[$i]/2)+($totals[$j]/2);
          for(my $k=$i+1;$k<$j;$k++){
            $ans+=$totals[$k];
          }
          $ans=$ans*$ans;
          $outMatrix[$i][$j]=$ans;
      }
    }
}

### print coincidence matrix
print OUT "COINCIDENCE MATRIX\n\n";
for(my $i=0;$i<=$#fulltable;$i++){
    if($i==0){
      print OUT "\t0\t1\t2\t3\t4\t5\t6\t7\t8\t9\n";
    }
    for(my $j=0;$j<=$#{ $fulltable[$i] };$j++){
      if($j==0){
          my $numb=$i;
          print OUT "$numb";
      }
      printf OUT "\t%.2f", $fulltable[$i][$j];
    }
    print OUT "\n";
}
```

```
## print matrix of differenc metrics for each value
print OUT "MATRIX OF DIFFERENCE METRICS\n\n";
for(my $i=0;$i<=$#outMatrix;$i++){
    if($i==0){
      print OUT "\t0\t1\t2\t3\t4\t5\t6\t7\t8\t9\n";
    }
    for(my $j=0;$j<=$#outMatrix;$j++){
      if($j==0){
          my $numb=$i+1;
          print OUT "$numb";
      }
      if($j<$i){
          printf OUT "\t%.2f", sqrt($outMatrix[$j][$i]);
      }
      else {
          printf OUT "\t%.2f", $outMatrix[$i][$j];
      }
    }
    print OUT "\n";
}

print OUT "\n\n";


my $calc=0;
my $top=0;
my $bottom=0;
for(my $i=0; $i<=9; $i++){
    for(my $j=$i; $j<=9; $j++){
      if($outMatrix[$i][$j]!=0){
          $top=$top + ($fulltable[$i][$j]*$outMatrix[$i][$j]);
#         print STDERR "TOP $i $j: $fulltable[$i][$j] *
$outMatrix[$i][$j]\n";

$bottom=$bottom+($totals[$i]*$totals[$j]*$outMatrix[$i][$j]);
#         print STDERR "BOT $i $j: $totals[$i] * $totals[$j] *
$outMatrix[$i][$j]\n";
      }
#     else {
#         print STDERR "skipped $i $j\n";
#     }
    }
}
print OUT "\nKAPPA COEFFICIENT CALCULATIONS\n";
print OUT "\n **** 1- ((n-1)(Sum of (values * difference
metric))****\n";
print OUT " ****    ----------------------------------------
****\n";
print OUT " ****        Sum of (totals*total*diffMetric)
****\n";

print OUT "full total (n): $fullTotal\n";
print OUT "top / bottom: $top / $bottom\n";
$calc=$fullTotal-1;
$calc=$calc*$top;
print OUT "(n-1) * top = $calc\n";
$calc=$calc/$bottom;
print OUT "(n-1) * top / bottom = $calc\n";
$calc=1-$calc;
printf OUT "Kappa: %.4f\n", $calc;
```

```
close FULL;
close OUT;
```

Items examined

Korean Rater 2 examined 21 items twice.

Item	time1	time2
a117s1	3.00	2.50
a117s2	1.50	1.00
a117s3	2.00	3.00
a117s4	4.00	1.50
a117s5	3.50	2.00
a117s6	2.00	1.50
a117s7	2.50	1.00
a117s8	3.50	3.00
a117s9	2.00	2.00
a117s10	3.50	3.00
a117s11	1.50	1.50
a117s12	1.00	1.00
a117s14	2.00	1.50
br17w14	1.50	1.00
a14w14	3.50	2.00
a116w14	3.00	3.50
a15w14	2.00	1.00
a110w14	2.50	2.50
a118w14	1.00	1.00
br12w14	2.00	2.50
br16w14	2.00	2.00

Recoded as follows into whole numbers for calculation of Coincidence Matrix

```
6       5
3       2
4       6
8       3
7       4
4       3
5       2
7       6
4       4
7       6
3       3
2       2
4       3
3       2
7       4
6       7
4       2
5       5
2       2
4       5
4       4
```

Output

```
COINCIDENCE MATRIX

      0     1     2     3     4     5     6     7     8     9
0     0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
1     0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
2     0.00  0.00  2.00  1.00  0.50  0.50  0.00  0.00  0.00  0.00
3     0.00  0.00  1.00  1.00  1.00  0.00  0.00  0.00  0.50  0.00
4     0.00  0.00  0.50  1.00  2.00  0.50  0.50  1.00  0.00  0.00
5     0.00  0.00  0.50  0.00  0.50  1.00  0.50  0.00  0.00  0.00
6     0.00  0.00  0.00  0.00  0.50  0.50  0.00  1.50  0.00  0.00
7     0.00  0.00  0.00  0.00  1.00  0.00  1.50  0.00  0.00  0.00
8     0.00  0.00  0.00  0.50  0.00  0.00  0.00  0.00  0.00  0.00
9     0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
MATRIX OF DIFFERENCE METRICS

      0     1     2     3     4     5     6     7     8     9
1     0.00  0.00  4.00  33.06 105.06       203.06       280.56
      370.56      430.56      441.00
2     0.00  0.00  4.00  33.06 105.06       203.06       280.56
      370.56      430.56      441.00
3     2.00  2.00  0.00  14.06 68.06 150.06       217.56       297.56
      351.56      361.00
4     5.75  5.75  3.75  0.00  20.25 72.25 121.00       182.25
      225.00      232.56
5     10.25 10.25 8.25  4.50  0.00  16.00 42.25 81.00 110.25
      115.56
6     14.25 14.25 12.25 8.50  4.00  0.00  6.25  25.00 42.25 45.56
7     16.75 16.75 14.75 11.00 6.50  2.50  0.00  6.25  16.00 18.06
8     19.25 19.25 17.25 13.50 9.00  5.00  2.50  0.00  2.25  3.06
9     20.75 20.75 18.75 15.00 10.50 6.50  4.00  1.50  0.00  0.06
10    21.00 21.00 19.00 15.25 10.75 6.75  4.25  1.75  0.25  0.00


KAPPA COEFFICIENT CALCULATIONS

 **** 1- ((n-1)(Sum of (values * difference metric))****
 ****    ---------------------------------------- ****
 ****          Sum of (totals*total*diffMetric)        ****
full total (n): 21
top / bottom: 378.5 / 15646.3125
(n-1) * top = 7570
(n-1) * top / bottom = 0.483820069425304
Kappa: 0.5162
```

Spanish rater 1

The algorithm used is as for Korean but allows values up to 10 as maximum possible score in Spanish for each item is 10.

Items examined

Item	time1	time2
a117s1	5.00	5.00
a117s2	3.00	3.00
a117s3	2.00	2.00
a117s4	4.00	4.00

a117s5	2.00	2.00
a117s6	4.00	4.00
a117s7	3.00	3.00
a118s8	4.00	3.00
a117s9	3.00	3.00
a117s10	5.00	5.00
a117s11	4.00	4.00
a117s12	5.00	5.00
a117s13	3.00	3.00
a117s14	2.00	2.00
a117s15	4.00	4.00

Output

```
COINCIDENCE MATRIX

      0     1     2     3     4     5     6     7     8
0     0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
1     0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
2     0.00  0.00  3.00  0.00  0.00  0.00  0.00  0.00  0.00
3     0.00  0.00  0.00  4.00  0.50  0.00  0.00  0.00  0.00
4     0.00  0.00  0.00  0.50  4.00  0.00  0.00  0.00  0.00
5     0.00  0.00  0.00  0.00  0.00  3.00  0.00  0.00  0.00
6     0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
7     0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
8     0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
MATRIX OF DIFFERENCE METRICS

      1     1.5   2     2.5   3     3.5   4     4.5   5
1     0.00  0.00  2.25  27.56 95.06 182.25       225.00       225.00
      225.00
2     0.00  0.00  2.25  27.56 95.06 182.25       225.00       225.00
      225.00
3     1.50  1.50  0.00  14.06 68.06 144.00       182.25       182.25
      182.25
4     5.25  5.25  3.75  0.00  20.25 68.06 95.06 95.06 95.06
5     9.75  9.75  8.25  4.50  0.00  14.06 27.56 27.56 27.56
6     13.50 13.50 12.00 8.25  3.75  0.00  2.25  2.25  2.25
7     15.00 15.00 13.50 9.75  5.25  1.50  0.00  0.00  0.00
8     15.00 15.00 13.50 9.75  5.25  1.50  0.00  0.00  0.00
9     15.00 15.00 13.50 9.75  5.25  1.50  0.00  0.00  0.00


KAPPA COEFFICIENT CALCULATIONS

 **** 1- ((n-1)(Sum of (values * difference metric))****
 ****    ---------------------------------------- ****
 ****          Sum of (totals*total*diffMetric)       ****
full total (n): 15
top / bottom: 10.125 / 3923.4375
(n-1) * top = 141.75
(n-1) * top / bottom = 0.0361290322580645
Kappa: 0.9639
```

Spanish rater 2

Algorithm

As with Korean, the algorithm was modified to take into account the 0.5 divisions of scale.

```
#!usr/bin/perl -w

use strict;

my ($infile, $outfile);

if($#ARGV!=1){
    die "Usage: perl $0 <input file> <outfile>\n";
}

else {
    $infile=$ARGV[0];
    $outfile=$ARGV[1];
}

open(FULL, ">$outfile.all") or die "Couldn't open $outfile.all:
$!\n";
open(OUT, ">$outfile.metric") or die "Couldn't open
$outfile.metric: $!\n";

my @fulltable=();
## 1=2, 3=4, 5=4, 7=6, 9=8
#my @evenValsTable=();
# for(my $i=-1;$i<=4;$i++){
#     for(my $j=-1;$j<=4;$j++){
#     $fulltable[$i+1][$j+1]=0;
#     }
# }
for(my $i=0;$i<=21;$i++){
    for(my $j=0;$j<=21;$j++){
      $fulltable[$i][$j]=0;
    }
}


## READ FILES ##
open(SIM, "$infile")  or die "Couldn't open results file $infile:
$!\n";
my $sentence=0;
    ## $grouptable[story][user]=value
my @grouptable=();
my $numberOfRespondentsPerStory=2;
while (defined (my $line = <SIM>)) {
    chomp $line;
    my @vals=split(/\s+/ , $line);
    for(my $user=0;$user<=$#vals;$user++){
      $grouptable[$sentence][$user]=$vals[$user];
    }
    $sentence++;
}
close SIM;
## foreach sentence
for(my $i=0;$i<=$#grouptable;$i++){
```

```
    ## for each rater, note their agreement with every other rater
    for(my $j=0;$j<=$#{ $grouptable[$i] };$j++){
      for(my $k=0;$k<=$#{ $grouptable[$i] };$k++){
          if($j!=$k){
            my $val1 = $grouptable[$i][$j];
            my $val2 = $grouptable[$i][$k];
            my $frac = 1/$numberOfRespondentsPerStory;
            ## IGNORE MISSING VALUES
            if($val1!=-1 && $val2!=-1){

$fulltable[$val1][$val2]=$fulltable[$val1][$val2]+$frac;
            }
          }
      }
    }
}

my @totals=();
my $fullTotal=0;
for(my $i=0;$i<=$#fulltable;$i++){
    if($i==0){
      print FULL
"\t0\t1\t2\t3\t4\t5\t6\t7\t8\t9\t10\t11\t12\t13\t14\t15\t16\t17\t18
\t19\t20\t21\n";
    }
    for(my $j=0;$j<=$#fulltable;$j++){
      if($j==0){
          print FULL "$i";
      }
      print FULL "\t$fulltable[$i][$j]";
      if($i<=$#fulltable && $j<=$#fulltable){
          $totals[$i]+=$fulltable[$i][$j];
      }
    }
    print FULL "\n";
    if($i<=$#fulltable){
      $fullTotal+=$totals[$i];
    }
}

#close FULL;

#open(OUT, ">$outfile.metric") or die "Couldn't open
$outfile.metric: $!\n";

my @outMatrix=();
for(my $i=0; $i<=$#fulltable; $i++){
    for(my $j=0; $j<=$#{ $fulltable[$i] }; $j++){
      if($j<=$i){
          $outMatrix[$i][$j]=0;
      }
      else {
          my $ans=($totals[$i]/2)+($totals[$j]/2);
          for(my $k=$i+1;$k<$j;$k++){
            $ans+=$totals[$k];
          }
          $ans=$ans*$ans;
          $outMatrix[$i][$j]=$ans;
      }
    }
}
```

```
### print coincidence matrix
print OUT "COINCIDENCE MATRIX\n\n";
for(my $i=0;$i<=$#fulltable;$i++){
    if($i==0){
      print OUT
"\t0\t1\t2\t3\t4\t5\t6\t7\t8\t9\t10\t11\t12\t13\t14\t15\t16\t17\t18
\t19\t20\t21\n";
    }
    for(my $j=0;$j<=$#{ $fulltable[$i] };$j++){
      if($j==0){
          my $numb=$i;
          print OUT "$numb";
      }
      printf OUT "\t%.2f", $fulltable[$i][$j];
    }
    print OUT "\n";
}

## print matrix of differenc metrics for each value
print OUT "MATRIX OF DIFFERENCE METRICS\n\n";
for(my $i=0;$i<=$#outMatrix;$i++){
    if($i==0){
      print OUT
"\t0\t1\t2\t3\t4\t5\t6\t7\t8\t9\t10\t11\t12\t13\t14\t15\t16\t17\t18
\t19\t20\t21\n";
    }
    for(my $j=0;$j<=$#outMatrix;$j++){
      if($j==0){
          my $numb=$i+1;
          print OUT "$numb";
      }
      if($j<$i){
          printf OUT "\t%.2f", sqrt($outMatrix[$j][$i]);
      }
      else {
          printf OUT "\t%.2f", $outMatrix[$i][$j];
      }
    }
    print OUT "\n";
}

print OUT "\n\n";


my $calc=0;
my $top=0;
my $bottom=0;
for(my $i=0; $i<=21; $i++){
    for(my $j=$i; $j<=21; $j++){
      if($outMatrix[$i][$j]!=0){
          $top=$top + ($fulltable[$i][$j]*$outMatrix[$i][$j]);
#         print STDERR "TOP $i $j: $fulltable[$i][$j] *
$outMatrix[$i][$j]\n";

$bottom=$bottom+($totals[$i]*$totals[$j]*$outMatrix[$i][$j]);
#         print STDERR "BOT $i $j: $totals[$i] * $totals[$j] *
$outMatrix[$i][$j]\n";
      }
#     else {
#         print STDERR "skipped $i $j\n";
```

```
#      }
    }
}
print OUT "\nKAPPA COEFFICIENT CALCULATIONS\n";
print OUT "\n **** 1- ((n-1)(Sum of (values * difference
metric))****\n";
print OUT " ****    ----------------------------------------
****\n";
print OUT " ****        Sum of (totals*total*diffMetric)
****\n";

print OUT "full total (n): $fullTotal\n";
print OUT "top / bottom: $top / $bottom\n";
$calc=$fullTotal-1;
$calc=$calc*$top;
print OUT "(n-1) * top = $calc\n";
$calc=$calc/$bottom;
print OUT "(n-1) * top / bottom = $calc\n";
$calc=1-$calc;
printf OUT "Kappa: %.4f\n", $calc;

close FULL;
close OUT;
```

Items examined

Item	time1	time2
a117s1	10.00	8.00
a117s2	5.50	2.00
a117s3	3.00	2.00
a117s4	4.00	4.00
a117s5	3.00	2.00
a117s6	9.00	9.00
a117s7	5.00	3.00
a118s8	5.00	6.00
a117s9	6.00	5.00
a117s10	2.00	4.00
a117s11	4.00	3.00
a117s12	7.00	4.00
a117s13	5.00	2.50
a117s14	2.50	3.00
a117s15	6.00	4.00

Recoded into whole numbers as follows:

20.00	16.00
11.00	4.00
6.00	4.00
8.00	8.00
6.00	4.00
18.00	18.00
10.00	6.00
10.00	12.00
12.00	10.00
4.00	8.00

8.00 6.00
14.00 8.00
10.00 5.00
5.00 6.00
12.00 8.00

Output

```
COINCIDENCE MATRIX

      0     1     2     3     4     5     6     7     8     9
     10    11    12    13    14    15    16    17    18    19
     20    21
0    0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
     0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
     0.00  0.00
1    0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
     0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
     0.00  0.00
2    0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
     0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
     0.00  0.00
3    0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
     0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
     0.00  0.00
4    0.00  0.00  0.00  0.00  0.00  0.00  1.00  0.00  0.50  0.00
     0.00  0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
     0.00  0.00
5    0.00  0.00  0.00  0.00  0.00  0.00  0.50  0.00  0.00  0.00
     0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
     0.00  0.00
6    0.00  0.00  0.00  0.00  1.00  0.50  0.00  0.00  0.50  0.00
     0.50  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
     0.00  0.00
7    0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
     0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
     0.00  0.00
8    0.00  0.00  0.00  0.00  0.50  0.00  0.50  0.00  1.00  0.00
     0.00  0.00  0.50  0.00  0.50  0.00  0.00  0.00  0.00  0.00
     0.00  0.00
9    0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
     0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
     0.00  0.00
10   0.00  0.00  0.00  0.00  0.00  0.50  0.50  0.00  0.00  0.00
     0.00  0.00  1.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
     0.00  0.00
11   0.00  0.00  0.00  0.00  0.50  0.00  0.00  0.00  0.00  0.00
     0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
     0.00  0.00
12   0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.50  0.00
     1.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
     0.00  0.00
13   0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
     0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
     0.00  0.00
14   0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.50  0.00
     0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
     0.00  0.00
```

```
15    0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
      0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
      0.00  0.00
16    0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
      0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
      0.50  0.00
17    0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
      0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
      0.00  0.00
18    0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
      0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  1.00  0.00
      0.00  0.00
19    0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
      0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
      0.00  0.00
20    0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
      0.00  0.00  0.00  0.00  0.00  0.00  0.50  0.00  0.00  0.00
      0.00  0.00
21    0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
      0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00  0.00
      0.00  0.00
MATRIX OF DIFFERENCE METRICS

      0     1     2     3     4     5     6     7     8     9
      10    11    12    13    14    15    16    17    18    19
      20    21
1     0.00  0.00  0.00  0.00  1.00  6.25  18.06 30.25 49.00 72.25
      90.25 115.56      138.06      156.25      162.56      169.00
      175.56      182.25      196.00      210.25      217.56
      225.00
2     0.00  0.00  0.00  0.00  1.00  6.25  18.06 30.25 49.00 72.25
      90.25 115.56      138.06      156.25      162.56      169.00
      175.56      182.25      196.00      210.25      217.56
      225.00
3     0.00  0.00  0.00  0.00  1.00  6.25  18.06 30.25 49.00 72.25
      90.25 115.56      138.06      156.25      162.56      169.00
      175.56      182.25      196.00      210.25      217.56
      225.00
4     0.00  0.00  0.00  0.00  1.00  6.25  18.06 30.25 49.00 72.25
      90.25 115.56      138.06      156.25      162.56      169.00
      175.56      182.25      196.00      210.25      217.56
      225.00
5     1.00  1.00  1.00  1.00  0.00  2.25  10.56 20.25 36.00 56.25
      72.25 95.06 115.56      132.25      138.06      144.00
      150.06      156.25      169.00      182.25      189.06
      196.00
6     2.50  2.50  2.50  2.50  1.50  0.00  3.06  9.00  20.25 36.00
      49.00 68.06 85.56 100.00      105.06      110.25      115.56
      121.00      132.25      144.00      150.06      156.25
7     4.25  4.25  4.25  4.25  3.25  1.75  0.00  1.56  7.56  18.06
      27.56 42.25 56.25 68.06 72.25 76.56 81.00 85.56 95.06 105.06
      110.25      115.56
8     5.50  5.50  5.50  5.50  4.50  3.00  1.25  0.00  2.25  9.00
      16.00 27.56 39.06 49.00 52.56 56.25 60.06 64.00 72.25 81.00
      85.56 90.25
9     7.00  7.00  7.00  7.00  6.00  4.50  2.75  1.50  0.00  2.25
      6.25  14.06 22.56 30.25 33.06 36.00 39.06 42.25 49.00 56.25
      60.06 64.00
10    8.50  8.50  8.50  8.50  7.50  6.00  4.25  3.00  1.50  0.00
      1.00  5.06  10.56 16.00 18.06 20.25 22.56 25.00 30.25 36.00
      39.06 42.25
```

```
11    9.50  9.50  9.50  9.50  8.50  7.00  5.25  4.00  2.50  1.00
      0.00  1.56  5.06  9.00  10.56 12.25 14.06 16.00 20.25 25.00
      27.56 30.25
12    10.75 10.75 10.75 10.75 9.75  8.25  6.50  5.25  3.75  2.25
      1.25  0.00  1.00  3.06  4.00  5.06  6.25  7.56  10.56 14.06
      16.00 18.06
13    11.75 11.75 11.75 11.75 10.75 9.25  7.50  6.25  4.75  3.25
      2.25  1.00  0.00  0.56  1.00  1.56  2.25  3.06  5.06  7.56
      9.00  10.56
14    12.50 12.50 12.50 12.50 11.50 10.00 8.25  7.00  5.50  4.00
      3.00  1.75  0.75  0.00  0.06  0.25  0.56  1.00  2.25  4.00
      5.06  6.25
15    12.75 12.75 12.75 12.75 11.75 10.25 8.50  7.25  5.75  4.25
      3.25  2.00  1.00  0.25  0.00  0.06  0.25  0.56  1.56  3.06
      4.00  5.06
16    13.00 13.00 13.00 13.00 12.00 10.50 8.75  7.50  6.00  4.50
      3.50  2.25  1.25  0.50  0.25  0.00  0.06  0.25  1.00  2.25
      3.06  4.00
17    13.25 13.25 13.25 13.25 12.25 10.75 9.00  7.75  6.25  4.75
      3.75  2.50  1.50  0.75  0.50  0.25  0.00  0.06  0.56  1.56
      2.25  3.06
18    13.50 13.50 13.50 13.50 12.50 11.00 9.25  8.00  6.50  5.00
      4.00  2.75  1.75  1.00  0.75  0.50  0.25  0.00  0.25  1.00
      1.56  2.25
19    14.00 14.00 14.00 14.00 13.00 11.50 9.75  8.50  7.00  5.50
      4.50  3.25  2.25  1.50  1.25  1.00  0.75  0.50  0.00  0.25
      0.56  1.00
20    14.50 14.50 14.50 14.50 13.50 12.00 10.25 9.00  7.50  6.00
      5.00  3.75  2.75  2.00  1.75  1.50  1.25  1.00  0.50  0.00
      0.06  0.25
21    14.75 14.75 14.75 14.75 13.75 12.25 10.50 9.25  7.75  6.25
      5.25  4.00  3.00  2.25  2.00  1.75  1.50  1.25  0.75  0.25
      0.00  0.06
22    15.00 15.00 15.00 15.00 14.00 12.50 10.75 9.50  8.00  6.50
      5.50  4.25  3.25  2.50  2.25  2.00  1.75  1.50  1.00  0.50
      0.25  0.00


KAPPA COEFFICIENT CALCULATIONS

 **** 1- ((n-1)(Sum of (values * difference metric))****
 ****    ------------------------------------------ ****
 ****         Sum of (totals*total*diffMetric)       ****
full total (n): 15
top / bottom: 153.6875 / 4138.125
(n-1) * top = 2151.625
(n-1) * top / bottom = 0.519951668932185
Kappa: 0.4800
```

Appendix 8

Raw Data Phase I

SUBJ	GEND	PITCH	TUNE	CHORD	RHY	MUSIC RECEP	SOUND	CHIN	NUM	GRAM	LANG RECEP	RAVEN
A110	f	14	9	11	4	38	10	4	4	15	33	75
A114	m	14	4	4	7	29	13	3	4	16	36	98
A119	f	15	7	8	7	37	13	4	3	13	33	87
A210	f	9	7	7	9	32	12	4	5	18	39	79
A223	f	14	7	9	10	40	12	6	10	13	41	87
A310	f	17	5	11	8	41	14	2	2	14	32	75
A312	m	10	4	3	7	24	8	1	3	14	26	83
A313	f	12	3	11	7	33	15	2	7	19	43	85
A321	m	14	7	8	7	36	13	2	4	12	31	79
B2	m	15	7	18	8	48	10	5	4	14	33	95
B7	m	16	7	13	4	40	12	3	2	15	32	80
B9	m	13	6	6	9	34	14	2	4	8	28	97
B12	m	13	8	10	6	37	13	2	4	15	34	95
B14	m	17	7	11	8	43	13	2	2	16	33	83
B15	m	13	8	10	10	41	10	2	4	17	33	75
B17	m	14	6	13	9	42	10	4	5	13	32	80
B22	m	18	6	13	10	47	14	2	1	14	31	95
B23	m	9	3	9	6	27	12	1	4	14	31	75
B24	m	17	9	10	9	45	4	3	1	9	17	75
B27	m	9	6	8	5	28	14	2	9	18	43	92.5
C4	m	7	8	7	5	27	12	4	6	7	29	95
C7	m	13	7	9	4	33	13	1	7	11	32	79
C9	m	17	5	11	10	43	13	4	2	18	37	95
C11	m	12	8	15	8	43	11	2	3	15	31	92
C15	m	13	6	8	6	33	12	1	2	13	28	95
E12	f	14	7	15	10	46	12	3	8	17	40	87
A15	f	16	9	11	10	46	15	4	6	15	40	65
A17	f	17	7	6	7	37	11	0	4	18	33	50
A111	f	14	5	4	8	31	10	2	3	15	30	62.5
A113	m	10	2	9	7	28	13	3	3	7	26	60
A116	m	7	6	5	7	25	12	1	3	4	20	50
A117	f	15	8	11	8	42	10	4	11	18	43	50
A118	f	10	4	7	10	31	12	2	9	13	36	31
A121	f	15	7	12	5	39	13	2	1	15	31	70
A122	f	9	2	12	5	28	11	3	2	15	31	72.5
A125	f	4	3	14	7	28	13	5	5	10	33	60
A127	m	5	10	2	9	26	15	2	6	18	41	50
A128	f	16	6	12	10	44	12	5	7	15	39	55
A129	f	17	7	7	6	37	13	1	1	12	27	43
A130	m	14	3	10	3	30	13	2	3	6	24	60
A22	f	15	8	11	7	41	10	2	5	11	28	45
A23	m	15	6	8	6	35	8	1	2	11	22	35
A28	f	14	6	13	10	43	12	3	3	11	29	56
A29	m	13	6	8	8	35	9	2	5	13	29	50
A220	m	15	6	8	5	34	13	3	2	11	29	68
A224	f	12	4	10	6	32	10	4	4	9	27	50
A227	m	18	4	7	6	35	13	2	3	7	25	45
A228	f	13	4	8	7	32	11	2	3	13	29	50
A35	f	14	9	10	10	43	12	2	4	18	36	35
A37	f	15	5	6	6	32	16	3	8	13	40	50
A38	f	5	6	7	6	24	11	3	5	10	29	45
A316	f	11	8	13	7	39	14	8	7	17	46	56
A320	f	9	1	3	5	18	12	5	3	13	33	56
A324	f	17	8	10	9	44	14	1	7	10	32	50

A326	m	11	1	7	4	23	13	3	6	9	31	35
SUBJ	GEND	PITCH	TUNE	CHORD	RHY	MUSIC	SOUND	CHIN	NUM	GRAM	LANG	RAVEN
A327	m	7	2	1	7	17	10	2	5	17	34	37
A328	f	5	1	12	9	27	15	2	3	17	37	43
A329	m	13	4	12	5	34	11	2	5	11	29	68
D3	f	10	6	3	6	25	7	2	5	10	24	44.5
D20	m	9	6	10	8	33	13	2	1	9	25	55
D6	f	16	8	9	2	35	9	3	3	9	24	33
D19	m	12	4	5	3	24	11	6	3	4	24	70
B	m	10	3	11	8	32	14	2	3	13	32	37
B10	m	16	9	13	8	46	10	1	3	11	25	65
B13	m	13	5	7	9	34	12	2	6	16	36	50
B18	m	17	8	6	7	38	15	3	8	11	37	70
B19	m	17	6	13	9	45	13	2	4	14	33	70
B20	m	10	7	10	7	34	14	3	0	12	29	72.5
B21	m	18	8	8	9	43	11	4	2	13	30	70
B25	m	19	9	13	8	49	8	3	5	9	25	55
B26	m	10	6	11	3	30	9	2	0	14	25	41
B28	m	15	8	10	9	42	11	1	2	10	24	65
C5	m	17	9	6	8	40	12	5	6	9	32	43
C6	m	8	4	8	6	26	11	2	3	6	22	41
C10	m	19	10	13	8	50	11	5	4	17	37	55
C12	m	13	5	5	6	29	9	4	1	13	27	70
C14	m	14	8	11	7	40	14	3	6	16	39	72.5
E3	f	13	8	12	9	42	14	2	3	12	31	45
E4	f	16	9	14	10	49	11	3	5	13	32	62
E6	f	14	7	8	9	38	12	2	2	14	30	40
E7	f	13	6	6	10	35	12	5	3	15	35	30
E10	f	15	7	13	9	44	12	5	8	15	40	45
E11	f	7	8	7	8	30	10	3	1	12	26	40
E13	f	13	7	9	8	37	13	3	7	17	40	56
E17	f	13	7	5	9	34	11	4	7	16	38	35
E19	f	11	6	8	10	35	16	6	5	10	37	40
A12	m	10	0	6	5	21	7	4	3	3	17	7
A13	f	12	4	9	6	31	12	4	2	9	27	8
A14	m	13	6	14	7	40	12	3	1	7	23	10
A16	m	18	7	12	6	43	11	1	6	15	33	22.5
A18	f	8	2	5	5	20	12	4	3	15	34	25
A19	m	5	4	14	3	26	12	3	2	5	22	22.5
A123	f	14	8	7	7	36	11	1	1	11	24	15
A126	f	9	2	5	5	21	10	2	2	9	23	12.5
A131	m	10	6	9	5	30	11	2	2	7	22	17.5
A25	m	18	7	15	9	49	8	3	1	9	21	24
A26	f	7	4	7	5	23	10	2	3	12	27	18
A27	m	12	10	9	6	37	10	4	1	10	25	22
A211	f	18	6	3	4	31	5	5	4	6	20	23
A214	f	7	3	9	4	23	9	1	0	13	23	23
A215	m	17	9	3	4	33	11	4	5	11	31	12.5
A216	f	9	8	11	7	35	12	2	1	13	28	22.5
A217	f	7	6	6	9	28	16	3	3	15	37	23.5
A218	f	14	6	9	2	31	12	4	4	9	29	25
A219	f	17	5	7	4	33	10	0	3	10	23	22.5
A221	m	9	2	1	2	14	7	4	3	3	17	8
A222	m	11	1	10	6	28	8	3	2	9	22	10
A225	m	14	7	5	4	30	10	5	2	15	32	22
A32	m	8	5	6	4	23	13	3	1	8	25	15
A33	f	7	4	12	4	27	13	2	4	10	29	15
A34	m	12	2	8	7	29	10	3	2	4	19	5
A36	f	8	3	1	4	16	10	3	3	10	26	7
A39	m	8	0	7	7	22	9	1	3	15	28	20

A314	m	9	8	7	8	32	13	3	2	16	34	10
A315	f	17	2	7	7	33	13	3	4	18	38	22.5
SUBJ	GEND	PITCH	TUNE	CHORD	RHY	MUSIC	SOUND	CHIN	NUM	GRAM	LANG	RAVEN
A318	m	14	4	10	8	36	14	1	4	12	31	17.5
A319	f	12	8	6	9	35	15	3	4	8	30	12.5
A322	m	13	5	4	5	27	13	4	7	12	36	22.5
A325	f	16	9	11	7	43	15	2	7	15	39	22.5
A330	f	10	5	8	7	30	11	2	4	9	26	20
D8	m	13	8	9	8	38	10	3	4	9	26	25
D9	m	15	6	6	2	29	9	4	2	15	30	10
D13	m	19	6	2	8	35	9	6	4	11	30	3
D18	m	11	5	5	3	24	8	4	2	4	18	2
D10	m	11	5	9	6	31	9	1	1	7	18	2
D14	f	16	3	1	8	28	12	2	2	9	25	0
D4	f	10	3	10	3	26	10	5	1	5	21	20
D2	m	8	4	6	5	23	10	2	2	9	23	12.5
D7	f	13	9	11	6	39	11	4	1	10	26	22.5
D12	m	7	6	5	3	21	2	1	2	6	11	10
D15	m	15	5	6	6	32	10	1	0	5	16	2
B3	m	16	6	8	6	36	9	2	0	4	15	22.5
B4	m	13	3	6	2	24	7	2	0	3	12	10
B5	m	16	6	11	7	40	12	3	0	12	27	0
B8	m	10	5	3	8	26	13	1	2	10	26	12.5
B11	m	10	5	9	10	34	11	2	2	8	23	15
B16	m	12	4	7	8	31	11	1	7	19	38	25
C1	m	8	6	8	5	27	12	1	1	9	23	17.5
C2	m	11	4	7	3	25	12	2	6	4	24	22.5
C3	m	16	3	10	3	32	12	3	3	9	27	8
C8	m	14	8	7	5	34	13	7	2	8	30	22.5
E2	f	9	5	13	5	32	13	4	3	11	31	22.5
E5	f	14	8	9	8	39	15	3	7	9	34	20
E8	f	13	10	18	9	50	15	7	6	14	42	15
E14	f	19	8	9	10	46	10	5	4	13	32	25
E15	f	12	8	12	7	39	11	6	4	14	35	15
E18	f	14	9	11	8	42	13	3	8	14	38	22.5
E20	f	11	7	8	10	36	13	5	4	11	33	20
E21	f	14	6	7	10	37	10	6	6	13	35	17.5

Appendix 9

COD: Subject Code
GEN: Gender
KO: Korean
SPPRO: Spanish Pronunciation
SPRHY: Spanish Rhythm
TOTSP: Total Spanish
LANPR: Total Language Productive
INVOC: Intervocalic interval
VOC: Vocalic interval
MUSPRO: Music Productive
MUSR: Music Receptive
LANR: Total Language Receptive
RAV: Raven

COD	GEN	KO	SPPRO	SPRHY	TOTSP	LANPR	INVOC	VOC	MUSPRO	MUSR	LANR	RAV
con11	m	45	32	51	83	132	52	64	1450	43	31	92
rap13	f	43	31	51	82	128	59	68.1	1502.083	37	40	56
br312	m	43	23	49	72	117	49	55.1	1227.083	24	26	83
rap4	f	44	32	44.5	76.5	123.5	65	58.1	1172.917	49	32	62
br111	f	44	19	43	62	109	76	69.3	1200	31	30	62.5
a15	f	44	29	42	71	120	67	45.7	1420.833	46	40	65
br113	m	36	21.5	42	63.5	101.5	67	65.1	1131.25	28	26	60
con6	m	45	28	41.5	69.5	116.5	48	57.2	1304.167	26	22	41
a118	f	40	22	41	63	106	60	63.3	1033.333	31	36	31
br130	m	40	21	41	62	104	61	55.5	1164.583	30	24	60
rap14	f	40	24	37.5	61.5	104.5	58	45.7	1197.917	46	32	25
a116	m	40	20.5	36	56.5	100.5	73	56.4	1160.417	25	20	50
rap6	f	40	30	36	66	108	46	42.3	1272.917	38	30	40
rap15	f	39	24	34	58	100	70	56.3	1495.833	39	35	15
rap19	f	37	31	34	65	106	60	52.3	1362.5	35	37	40
br16	m	37	19	33	52	91	64	63.9	1377.083	43	33	22.5
rap21	f	45	25	32	57	105	65	49.2	845.8333	37	26	25
rap8	f	36	30	32	62	99	53	53.9	1547.917	50	42	15
rap5	f	47	22	31.5	53.5	102.5	63	61	1335.417	39	34	20
con15	m	41	21	31	52	96	57	50.3	1097.917	33	28	95
br34	m	47	23	29	52	101	52	58	922.9167	29	19	5
rap10	f	31	23	29	52	85	70	57.5	952.0833	44	40	45
br19	m	44	23.5	28	51.5	97.5	62	66.8	1258.333	26	22	22.5
a117	f	34	25	28	53	97	69	45.8	1354.167	42	43	50
br129	f	43	26	28	54	100	69	55.6	1358.333	37	27	43
con8	m	47	19	27	46	96	63	58.9	1039.583	34	30	22.5
rap18	f	36	27	26.5	53.5	90.5	74	60.5	1327.083	42	38	22.5
br32	m	39	27	24.5	51.5	93.5	59	58.9	675	23	25	15

COD	GEN	KO	SPPRO	SPRHY	TOTSP	LANPR	INVOC	VOC	MUSPRO	MUSR	LANR	RAV
a14	m	41	21	24	45	88	57	56.6	1085.417	40	23	10
br322	m	44	20	24	44	91	76	61.7	989.5833	27	36	22.5
rap20	f	43	19	24	43	88	60	52.8	1460.417	36	33	20
br119	f	34	21	24	45	82	75	58.5	810.4167	37	33	87
br12	m	40	18	23	41	84	64	65.2	670.8333	21	17	7
rap12	f	43	23	22	45	90	73	59.9	1545.833	46	40	87
br17	f	42	21	22	43	87	70	49.5	1272.917	37	33	50
br39	m	41	21	19	40	84	84	63	658.3333	22	28	20
a110	f	41	18	19	37	82	63	51.7	1141.667	38	33	75
br313	f	33	18	19	37	73	65	53.8	735.4167	33	43	85
br18	f	38	17	18	35	75			866.6667	20	34	25
rap2	f	37	16	18	34	73	55	63.8	733.3333	32	31	22