Genetic and Environmental Influences on Learning Chinese Language and Literacy Skills

Wing Yin Chow

St. John’s College, University of Oxford

Thesis submitted for the degree of Doctor of Philosophy

Trinity Term 2010

Acknowledgements

I would like to express my sincere gratitude to all those who have given me the possibility to complete this thesis.

My deepest gratitude goes to my supervisor Professor Dorothy Bishop, who has always provided me with unswerving support and care over these years, and has always been there whenever I need help and advice. Her insightful comments, patient guidance, and ongoing encouragement at all stages of my research have brought this thesis to completion. She is always enthusiastic about research, patient, and kind to others, who is an excellent role model for me to grow as a better researcher. I hope in the future I could nurture students in the same way as hers.

My sincere gratitude also goes to Professor Connie Ho and Professor Mary Waye who have shown me great assistance and guidance throughout the course of study. Their tremendous help especially with research grant application, research material preparation, data collection, and zygosity determination has made this thesis possible.

I own my special thanks to Professor Kate Nation for her invaluable comments on my thesis, and taking great care of not only my academic work but also my well-being in these years. I am grateful to Dr. Silvia Paracchini for her helpful feedback at the early stage of my thesis, and co-organizing the genetics journal club which has broaden my horizons on genetics research. I am also thankful for Professor Richard Olson and Professor Brian Byrne for their insightful advice on and discussion of the work in this thesis.

My heartfelt gratitude goes to Professor Catherine McBride-Chang and Professor Ming Ming Chiu who have always given me a lot of invaluable advice and have shared with me the joys and frustrations on paving the way for research and teaching.

I am grateful to all principals, teachers, parents and children who generously contributed their valuable time and efforts to the work in this thesis, and also the research assistants for their company in working some very long days and nights in the data collection period. My special thanks go to all OSCCI teammates for making the department a supportive and pleasant learning environment, and my dearest friends, especially Ben, Loriane, Ornella, Richard, Top and Yi-Chuan, for their generosity and friendship making my life easier and more cheerful in Oxford.

My heartfelt thanks go to Simpson, who has always provided me with endless support and encouragement. Last but not least, I owe my deepest gratitude to my family. They have showed me great understanding, especially in the period which I was staying in Oxford. Their love is the treasure of my life.

Table of Contents

Short Abstract

Genetic and Environmental Influences on Learning Chinese Language and Literacy Skills

Wing Yin Chow

St. John's College, University of Oxford

Thesis submitted for the degree of Doctor of Philosophy, Trinity Term, 2010

This thesis investigated the etiology of individual differences in Chinese language and literacy skills with a two-wave longitudinal design using a sample of 312 Chinese twin pairs aged 3 to 11 in Hong Kong. Children were individually given tasks of Chinese word reading, receptive vocabulary, phonological memory, tone awareness, syllable and rhyme awareness, rapid automatized naming, morphological awareness, and orthographic skills, as well as nonverbal reasoning and audiometric screening tests. They were tested again on the same tasks, except nonverbal reasoning, one year after the initial testing. Children's saliva was collected to perform SNP testing for zygosity determination. Also, their demographic information, home literacy environment profile, and motivation for learning text, were obtained from parent-rated questionnaires.

Overall, there were four major findings on Chinese language and literacy abilities with the effects of age and nonverbal reasoning controlled for. First, genes and environments had differential influences on various skills, and there was a possibility of different etiology in language and reading development. Second, socioeconomic status and home literacy environment were plausible mediators but not moderators of general language and reading abilities. Third, the stability of various skills across a one-year time period was mainly mediated by genetic influences, but shared environmental factors also influenced syllable and rhyme awareness. Also, new genetic and environmental factors came into play at Time 2 for word reading, and results suggested new genetic influences and new shared environmental influences emerged at Time 2 for tone awareness and morphological awareness respectively. Fourth, both genes and environments contributed to parent-rated communicative ability and motivation for learning text. Furthermore, the link between word reading and parent-rated motivation was mediated by genetic processes. The universality of the genetic and environmental origins across languages and their specificity to Chinese, as well as the implications of these findings, were discussed.

Long Abstract

Genetic and Environmental Influences on Learning Chinese Language and Literacy Skills

Wing Yin Chow

St. John's College, University of Oxford

Thesis submitted for the degree of Doctor of Philosophy, Trinity Term, 2010

Individual differences in language and literacy skills tend to emerge early in development and continue over time (Cunningham & Stanovich, 1997). Both heredity and environment play a role in the origins of these individual differences, but their relative contributions can differ in various language and literacy skills as demonstrated in twin studies (e.g., Byrne et al., 2005). In addition, genes and environments contribute to the overlap between language and literacy skills as well as their stability over time (e.g. Harlaar, Dale, & Plomin, 2007a; Petrill, Thompson, Deater-Deckard, DeThorne, & Schatschneider, 2006). However, these studies focused on alphabetic languages, mainly English, and thus whether these roles of heredity and environment are universal or differ across languages remain unknown. Also, these past studies examined a few cognitive skills only and did not investigate literacy-related beliefs, such as motivation. Therefore, the search of genetic and environmental contributions to language and literacy skills has been rather limited.

To fill in this gap of research, this thesis investigated the etiology of individual differences in Chinese language and literacy skills with a two-wave longitudinal design using a sample of 312 Chinese twin pairs aged 3 to 11 in Hong Kong. Chinese has very different linguistic characteristics compared with alphabetic languages, and thus is an appealing

language for study to address the issue of universality. Also, to provide a more comprehensive picture on the roles of genes and environments in language and literacy development, this thesis included cognitive skills which had not been studied in past twin studies (e.g., tone awareness), investigated children's motivation for learning text, and employed both psychometric tests and parental reports. In this thesis, children were individually given tasks of Chinese word reading, receptive vocabulary, phonological memory, tone awareness, syllable and rhyme awareness, rapid automatized naming, morphological awareness, and orthographic skills, as well as nonverbal reasoning and audiometric screening tests. They were tested again on the same tasks, except nonverbal reasoning, one year after the initial testing. Children's saliva was collected to perform SNP testing for zygosity determination. Also, their demographic information, home literacy environment profile, and motivation for learning text, were obtained from parent-rated questionnaires.

Overall, four domains of genetic and environmental influences were examined in this thesis. These influences on language and literacy skills were examined with the effects of age and nonverbal reasoning controlled for. First, **this thesis investigated the genetic and environmental origins and overlap of individual differences in Chinese language and reading skills**. Findings indicated differential roles of genetic and environmental influences in various Chinese language and reading skills, and a possibility of different etiology in language and reading development. Specifically, results of genetic analyses indicated moderate to strong genetic influences on Chinese word reading, tone awareness, phonological memory, and rapid automatized naming, while shared environment played a more important role in Chinese receptive vocabulary, syllable and rhyme awareness, and orthographic skills. Morphological awareness showed significant influences of both genetic and shared environmental factors. Factor analyses on these diverse Chinese tasks indicated two factors: General Language and General Reading, and further genetic analyses were conducted on these two factors. Moderate genetic influence was found for both general language and reading abilities, but bivariate genetic analyses showed nonsignificant genetic or shared environmental links between these abilities.

Second, **it tested whether socioeconomic status (SES) and home literacy environment (HLE) as identified environmental factors mediated environmental influences and/or moderated the genetic effects on Chinese language and reading skills**. Twin pairs with SES and/or HLE information provided by their parents were included (296 and 215 twin pairs for SES and HLE respectively). The roles of SES and HLE on general language and reading abilities were examined in two ways, including a) mediation of shared environmental influences and b) moderation of genetic influences. Results showed that SES and HLE (Reading resources and opportunities at home) mediated the shared environmental contributions to general language and reading abilities, but did not moderate the genetic influences. Also, this thesis investigated whether SES and HLE were plausible environmental mediators of more specific skills which were influenced by shared environments, including receptive vocabulary, syllable and rhyme awareness and orthographic skills. Results showed that SES and HLE (Reading resources and opportunities at home) mediated the shared environmental contributions to receptive vocabulary and syllable and rhyme awareness, but not orthographic skills.

Third**, it investigated genetic and environmental contributions to the developmental etiology of Chinese language and reading skills** across a one-year time period in 292 Chinese twin pairs. Results indicated that various language and reading skills were moderately to largely stable across time. Genetic analyses indicated their stability was mainly mediated by genetic influences, but some evidence showed shared environment exerted stronger influences on the stability of syllable and rhyme awareness. Also, findings showed new genetic and environmental factors came into play at Time 2 for Chinese word reading, and suggested new genetic influences and new shared environmental influences emerged at Time 2 for tone awareness and morphological awareness respectively.

Fourth, **parent-rated children's communicative ability and motivation for learning text, and their links with reading abilities were examined.** Parents' ratings on twenty positive items of Children's Communication Checklist – Second Edition (CCC-2; Bishop, 2003) and four items on motivation for learning text were obtained for 307 and 306 twin pairs respectively. Results indicated significant genetic and shared environmental influences for both communicative ability and motivation for learning text. Specifically, there were stronger shared environmental influences than genetic effects for communicative ability and comparable genetic and environmental contributions to motivation. Also, genetic factors contributed to the link between motivation and word reading.

Overall, by comparing the findings of this thesis and those of past research, three aspects of universal influences have been suggested. First of all, genetic factors exerted greater impacts on word reading, phonological memory, and rapid automatized naming, whereas shared environments had stronger effects on vocabulary knowledge. In addition, home environments, including SES and HLE, mediated the shared environmental influences on language and reading skills. Furthermore, genetic factors contributed to the stability of word reading and phonological awareness, and shared environments did not significantly influence the continuity of rapid automatized naming across time.

This thesis has also demonstrated some findings which were more language- or country- specific. Results showed that individual differences in Chinese phonological awareness at syllable and rhyme levels and orthographic skills were more affected by shared environments than by genes, but past studies highlighted the genetic influences on these skills in alphabetic languages. Also, there was no genetic link between general language and general reading abilities in Chinese, and this was contrastive to past findings on alphabetic languages. Furthermore, there were new sources of genetic and environmental influences on Chinese word reading, and they were demonstrated in the United States samples in past research. Therefore, these new sources of influences could be country-specific.

In sum, this thesis has put genetic, cognitive, behavioural and belief, and developmental perspectives together in the pursuit of the etiology of language and literacy development. It has extended past twin research in the search of genetic and environmental contributions in several ways. To begin with, it investigated the nature of Chinese acquisition and thus, by comparing its findings with past research on alphabetic languages, has suggested the universality of the genetic and environmental influences on children's language and literacy skills. Also, it studied the moderating and the mediating roles of identified home environments in a single study, and so the influences of identified home environments could be better understood. Furthermore, it included cognitive skills which had not been studied in past twin studies (e.g., tone awareness), and investigated motivation for learning text which reflected children's beliefs. Thus, the search of genetic and environmental contributions to language and literacy development has covered not only wider domains of cognitive skills, but also outside these cognitive aspects. Lastly, it examined parent-rated communicative skills in typically developing children, which has only been tested in a twin study focusing on language impaired children. Findings of this thesis help paving the road for fully exploiting children’s potentials and for supporting at risk children to surpass their genetic propensities.

Chapter 1

General Introduction

Overview

Mastery of literacy skills is fundamental in children’s development as it is key to learning in various spheres of life (Snow, Burns, & Griffin, 1998). Being able to read efficiently and fluently is a major milestone and an ideal outcome in literacy acquisition. However, learning to read is effortful and requires the support of a variety of cognitive skills, and thus could be challenging. While some children learn to read with ease, others find it particularly difficult, and thus huge variations in reading skills have been observed (Vellutino & Fletcher, 2005).

Reading requires understanding visual symbols which represent speech (Ziegler & Goswami, 2005). Reading a word involves retrieval of the word’s sound and meaning from its printed form (Snowling & Hulme, 2005). Therefore, one of the major building blocks of reading skills is language abilities, such as vocabulary knowledge. Language variations could be evident at an early age before children have started formally learning how to read. For instance, the top one-tenth of 16-month-olds could produce at least 154 words, but the bottom one-tenth of them could hardly produce any words (Bates, Dale, & Thal, 1995). These early language variations may indicate children’s later potential reading achievement. Longitudinal studies have shown early language abilities predict subsequent language and reading proficiency, and children with poor language skills are more likely to have low language and literacy proficiency when they grow up (e.g., Larney, 2002). Although most of them scored within normal limits, when compared with controls, late-talking toddlers had poorer language-related skills, such as vocabulary, grammar, verbal memory, and reading comprehension, when they became adolescents (Rescorla, 2005). Therefore, these variations on language and reading skills tend to emerge early in development and continue over time (Cunningham & Stanovich, 1997), and so create great practical challenges to parents, educators and therapists.

Mounting research has been conducted to understand the causes of these individual differences on language and literacy skills. A core strand of this area of research is on how genetic factors and environments contribute to the emergence and the stability of these individual differences. The twin study method, as a genetically sensitive design, has been utilized in the pursuit of their genetic and environmental origins. To date, twin studies have demonstrated the roles of genes and environments on learning alphabetic languages, mainly on English (e.g. Byrne et al., 2005; Harlaar et al., 2007a; Petrill et al., 2007; Samuelsson et al., 2008). However, whether these roles of heredity and environment are universal or differ across languages remains unknown. This issue of universality across languages could be best addressed by studying a language that has very different characteristics than that of English. If results generalize across characteristically distant languages, the notion of universality could be better substantiated. For this reason, Chinese is an appealing language to be examined.

To fill in this gap of research, this thesis investigates the genetic and environmental contributions to Chinese language and literacy skills in a two-wave longitudinal design with 312 Chinese twin pairs in Hong Kong. In this thesis, two facets of language and literacy skills are the central foci, and they are vocabulary knowledge and word reading. Broader aspects of language and literacy skills are examined by also considering cognitive skills including phonological memory, tone awareness, syllable and rhyme awareness, rapid automatized naming, morphological awareness and orthographic skills.

In this thesis, four domains of genetic and environmental influences are examined, each of which will be described in detail in the following chapters. First, this thesis investigates the genetic and environmental origins and overlap of Chinese language and reading skills (Chapter 3). Second, it tests whether socioeconomic status (SES) and home literacy environment (HLE) as identified environmental factors mediate environmental influences and/or moderate the genetic effects on Chinese language and reading skills (Chapter 4). Third, it investigates genetic and environmental contributions to the developmental etiology of Chinese language and reading skills across a one-year time period (Chapter 5). Fourth, parent-rated variables, including children's communicative ability and motivation for learning text, and their links with reading abilities are examined (Chapter 6). These four domains examined in a single study provide a more comprehensive picture on the roles of genes and environments in language and reading development.

This thesis as a whole extends past twin research in the search of genetic and environmental contributions in several ways. To begin with, it investigates the nature of Chinese acquisition and thus, by comparing its findings with past research on alphabetic languages, suggests the universality of the genetic and environmental roles in children's language and reading skills. Also, it studies the moderating and the mediating roles of identified home environments in a single study, and so the influences of identified home environments could be better understood. Furthermore, it includes cognitive skills which have not been studied in past twin studies (e.g., tone awareness), and investigates motivation for learning text which reflects children's beliefs. Thus, the search of genetic and environmental contributions to reading covers not only wider domains of cognitive skills, but also outside these cognitive aspects. Lastly, it examines parent-rated communicative skills in typically developing children, which has only been tested in a twin study focusing on language impaired children.

Before going into the details of each domain, this chapter provides a general introduction on the twin study design as a background for the following chapters. This chapter describes how the twin study method has been employed to study nature and nurture influences, and the logics, models and assumptions of the twin study design. It also discusses Chinese character reading in the dual-route approach context and briefly describes Chinese literacy instruction in Hong Kong. Further research backgrounds which are particularly relevant to individual chapters, such as links between specific variables, will be presented later on.

Twin study method as a useful means to understand the nature and nurture influences

Evidence emerges that both nature and nurture take part in determining individual variations in language and literacy skills. While the identification of genes linked to language and reading skills, such as KIAA0319, suggests that language and reading are influenced by innate biological processes (e.g., Paracchini et al., 2008), the effects of various environmental factors, such as schooling, provide strong evidence of the role of nurture (Rutter & Maughan, 2002). However, what remains unclear is how they contribute to these variations, and more specifically which of them plays a more important role in language and reading development.

From a century ago when reading difficulties were found to aggregate in families (Thomas, 1905), resemblance on language and reading abilities has been consistently observed in families, suggesting that heredity has a prominent influence on language and reading development. However, empirical testing of this notion on humans is complicated, because it is not possible to experimentally select and manipulate genotypes and environmental factors in humans. Rather, this investigation has to rely on natural genetic and environmental variations. Family studies made a promising start for this natural investigation by examining the degree of similarity in a trait among members of the same family, including siblings and parents. Francis Galton's study on the links between family members' mental ability in the mid-nineteenth century was a precursor of the field. However, members of a biological family who live together not only have similar genetic makeup, but also experience similar environments, and thus this kind of familial data does not tease genetic and environmental influences apart. Although the absence of familial resemblance suggests the role of heritability is negligible, a finding of correlation of family members' performance could be due to both heredity and shared environment, and does not indicate their relative prominence. To disentangle genetic from environmental sources of resemblance in human traits and abilities, more genetically sensitive designs, such as adoption and twin studies, are needed. In recent decades, these designs have successfully shed light on the differential influence of heredity and environmental factors on language and literacy skills. Owing to the aim of this thesis, twin studies are the focus of the following review.

A twin design for investigating genetic and environmental influences was introduced by Francis Galton in the late-nineteenth century. By comparing two groups of twins distinguished by their physical likeness at birth, he found that resemblance in traits persisted across time in originally similar twins, but not in their originally physically dissimilar counterparts who also experienced similar environments, suggesting a genetic source of resemblance. However, twins' physical likeness at birth is only a rough indicator of zygosity. Later on, twin studies which distinguished monozygotic (MZ) and dizygotic (DZ) twins were conducted, and the earliest twin research on cognitive skills investigated the resemblance of IQ in the 1920s. This twin research showed that MZ twins were more similar than DZ twins in IQ, suggesting the importance of heredity (Merriman, 1924). Twin design has then been employed to study different cognitive constructs, including language and literacy skills (e.g., Mather & Black, 1984; Munsinger & Douglass, 1976).

Logics of twin design

Twin design can disentangle genetic from environmental sources of resemblance based on the difference of the genetic relatedness of MZ and DZ twins. MZ twins come from the split of a single zygote after conception, and thus they share all their genes. On the other hand, DZ twins come from separate fertilizations of two ova, and thus they only share half of their segregating genes, and one-forth of dominance genetic variation, on average (chances of sharing both alleles, one allele, and no alleles are 25%, 50% and 25% respectively).

Under the equal environments assumption, which supposes that twin pairs share equally similar prenatal and postnatal environments regardless of their zygosity, any extra similarities of MZ twins over DZ twins can be attributed to genetic origins. Thus, if cotwins of MZ twin pairs are more similar than those of DZ twin pairs, we may infer that genetic factors are involved in determining the individual differences. The contribution of genetic factors can be quantified by heritability (a^2) which refers to the estimate of the genetic effect size. Specifically, heritability is the proportion of phenotypic variance of a trait that is due to genetic differences of individuals. It ranges from 0, which implies no effects of genes, to 1, which means genetic differences completely account for the phenotypic variance. There are two types of heritability: broad-sense heritability and narrow-sense heritability. Broad-sense heritability refers to all sources of genetic variance, including additive and non-additive ones, while narrow-sense heritability refers to additive genetic variance only. In general, twin studies focus on narrow-sense heritability for simplicity by assuming only additive genetic effects influence a trait. Thus, heritability could be estimated in twin studies by doubling the difference between correlation of MZ twins and that of DZ twins [($a^2 = 2\ (r_{MZ} – r_{DZ})$)].

Phenotypic variance is considered as under the influence of genetic factors plus that of environmental factors (including measurement errors), assuming the gene-environment correlation and interaction are negligible. The effects of shared environment (c^2) which make twins resemble to each other (e.g., parenting practice and family SES) and those of nonshared environment (e^2) which distinguish twins from one another (e.g., illnesses that only one cotwin experienced and cotwin's own peer groups), can be shown in the twin design.

If cotwins of MZ twin pairs are no more similar than those of DZ twin pairs, we may infer that heritability is negligible and shared environmental factors play a role in the individual differences. Strong shared environmental influences are indicated when the similarity between cotwins of MZ twin pairs and that of DZ twin pairs is high and equivalent. Shared environment influence (c^2) can be estimated by subtracting heritability (a^2) from MZ twins' correlation ($c^2 = r_{MZ} – a^2$) because effects of heredity and shared environment contribute to MZ twins' resemblance. On the contrary, a strong genetic influence and a negligible shared environmental effect on cotwins' resemblance is shown when the correlation of MZ twins doubles that of the DZ twins. If cotwins of MZ twin pairs do not show a perfect correlation, we may infer that nonshared environment (including measurement errors) plays a role in the individual variations. Nonshared environment influence (e^2) can be estimated by subtracting MZ twins' correlation from 1 ($e^2 = 1 - r_{MZ}$). Heritability (a^2), shared environmental

(c^2) and nonshared environmental (e^2) influences are proportions of variance and their sum equals 1.

Though the aforementioned logics and computation could be applied to both same-sex and opposite DZ twins, same-sex DZ twins are often a better comparison group for their MZ counterparts who are always same-sex pairs, unless gender difference is a main focus of the study (see Viding et al., 2004). Where there are gender influences on a trait, opposite-sex DZ twins tend to be less similar than same-sex DZ twins, and so including opposite-sex DZ twins in the comparison may inflate heritability.

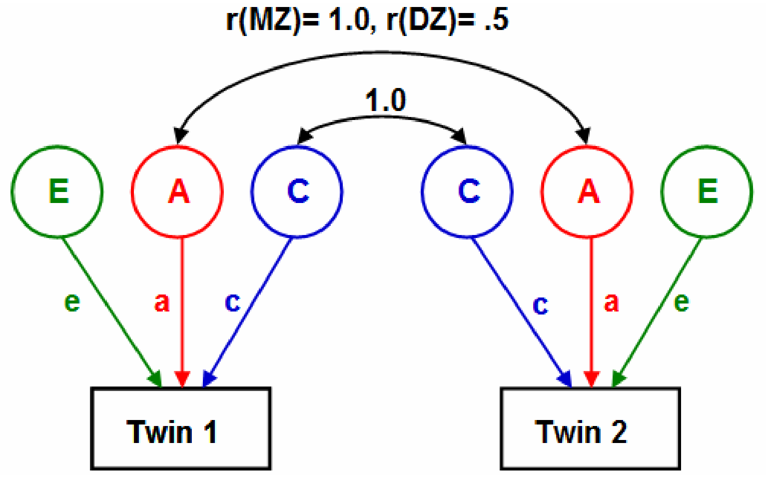
Statistical modeling in twin design

The quantification of genetic and environmental contributions can be achieved by statistical modeling.

Univariate ACE model

A univariate ACE model estimates the proportion of phenotypic variance due to genetic (A), shared environmental (C) and nonshared environmental (E) factors. This model is depicted in Figure 1.1.

Figure 1.1. Univariate ACE model



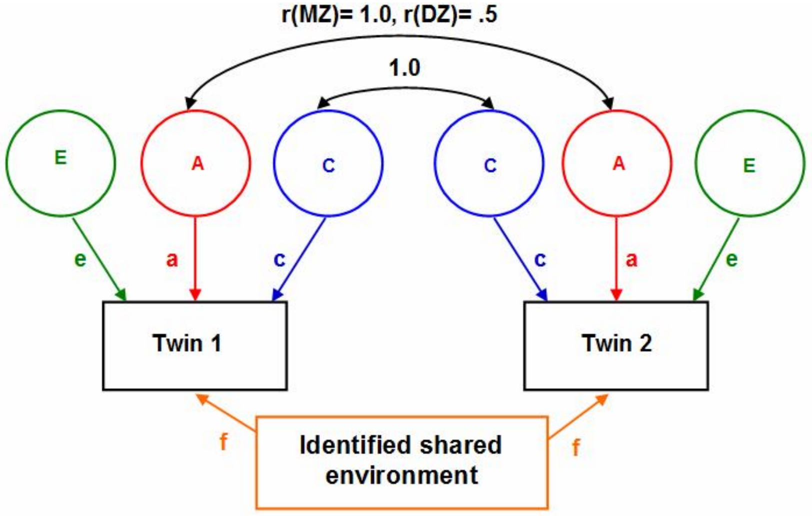
This model includes two measured variables representing the observed performance of the two cotwins in a pair (i.e., Twin 1 and Twin 2), and three latent variables influencing their observed performance, including additive genetic (A), shared environmental (C) and nonshared environmental (E) factors. Because the genetic and environmental factors contribute to the individual differences in the measured variables, single-headed arrows link the latent factors A, C, and E to the measured variables. As MZ twins share all their genes, and DZ twins share half of their segregating genes on average, the genetic terms (A) are linked by a double-headed arrow with a value of 1 for MZ and 0.5 for DZ. By definition, shared environment (C) makes twins resemble to each other whereas nonshared environment (E) is specific to individuals. So, the shared environmental terms (C) are perfectly correlated, but nonshared environmental terms (E) do not correlate. Paths ‘a’, ‘c’ and ‘e’ indicate the independent contributions of the latent factors A, C, and E respectively, when the effects of the other latent factors are accounted for. For example, the path ‘a’ shows the additive genetic influences after the environmental effects are considered. Therefore, the proportion of phenotypic variance explained by genetic, shared environmental and nonshared environmental factors are quantified by a^2, c^2 and e^2 respectively. The sum of these products is the total phenotypic variance ($V = a^2 + c^2 + e^2$).

In this thesis, the univariate ACE model is employed to examine the extent to which genetic, shared environmental and nonshared environmental factors contribute to Chinese language and literacy skills, including word reading, receptive vocabulary, phonological memory, tone awareness, syllable and rhyme awareness, rapid automatized naming, morphological awareness, and orthographic skills (Chapter 3), and parent-rated communicative ability and motivation for learning text (Chapter 4). Apart from these, the univariate ACE model provides a foundation for more sophisticated models for investigating identified environments' mediation and moderation effects (i.e., univariate ACE mediation model and split-group analysis with univariate ACE model), as well as etiological overlap between skills and longitudinal stability (i.e., Cholesky decomposition model).

Univariate ACE mediation model

One of the aims of this thesis is to understand how identified home environments influence individual differences in language and reading skills (Chapter 4). So, a model which builds on the univariate ACE model and includes effects of an environmental mediator is used to pursue this research question. This model concerns identified environments in common between the cotwins in a twin pair, such as family socio-economic status, and the shared environmental term is decomposed into variances explained by identified and unidentified shared environmental influences (see Figure 1.2).

Figure 1.2. Univariate ACE mediation model



As shown in Figure 1.2, on top of a univariate ACE model, there is an extra measured variable labeled as ‘identified shared environment’ linking to the two measured variables of cotwins’ performance by single-headed arrows. Path ‘f’ indicates the independent contributions of the measured variable identified shared environments, after the effects of genes, unidentified shared environments and nonshared environments are accounted for. The proportion of phenotypic variance explained by genetic, identified and unidentified shared environmental, and nonshared environmental factors are quantified by a^2, f^2, c^2 and e^2 respectively. The sum of these products is the total phenotypic variance ($V= a^2 + c^2 + e^2 + f^2$). This model has been successfully employed in past twin studies (e.g., Petrill, Pike, Price, & Plomin, 2004).

Cholesky decomposition model

Multivariate genetic analysis provides a useful means to understand the extent to which genetic or environmental factors contribute to the phenotypic correlation and specificity across variables. A multivariate ACE model basically builds on univariate ACE models each concerns one of the variables, and includes additional parameters linking the variables of interest. Different forms of multivariate genetic models have different assumptions about the underlying nature of development, and thus vary in how the additional parameters representing the links are specified. Among these models, the Cholesky decomposition model has been widely used in reading research to understand the genetic and environmental overlap between skills and longitudinal stability (e.g., Byrne et al., 2005; Harlaar et al., 2007a; Harlaar, Dale, & Plomin, 2007b; Hayiou-Thomas, Harlaar, Dale, & Plomin, 2006; Oliver, Dale, & Plomin, 2005; Petrill et al., 2007; Samuelsson et al., 2008). In general, this Cholesky decomposition model estimates genetic and environmental effects in common between two or more measured variables as well as those unique to a variable(s) which are independent to the previously entered variable(s). Its logic is similar to hierarchical regression in phenotypic analysis. The Cholesky decomposition model concerning two measured variables is depicted in Figure 1.3.

Figure 1.3.a. Cholesky decomposition model (full view).

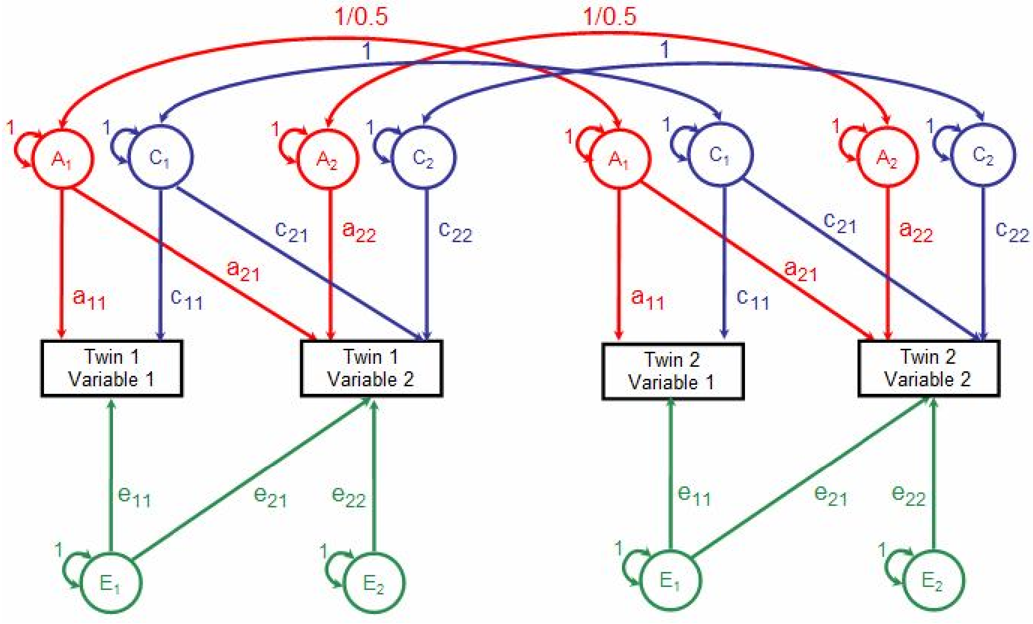
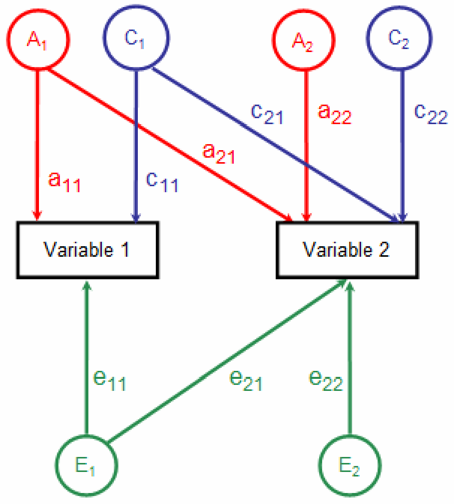


Figure 1.3.b. Cholesky decomposition model (simplified view).



As shown in the full view figure, two sets of ACE terms are specified in this model. The first set (A_1, C_1 and E_1) links to both variables 1 and 2 within each cotwin. It represents the genetic (A), shared environmental (C) and nonshared environmental (E) effects that contribute to the total variance in variable 1 (i.e., paths a_{11}, c_{11}, and e_{11}), as well as their effects that contribute to the covariance between the two variables (i.e., paths a_{21}, c_{21}, and e_{21}). Hence, the shared paths linking A_1, C_1 and E_1 to variable 2 (i.e., paths a_{21}, c_{21}, and e_{21}) represent the genetic and environmental contributions to the phenotypic correlation. The second set (A_2, C_2, and E_2) connects to variable 2 only (i.e., paths a_{22}, c_{22}, and e_{22}). These specific paths linking A_2, C_2, and E_2 to variable 2 performance indicate the unique A, C, and E contributions to variable 2 independent of those to variable 1. Apart from the within-twin paths, there are links connecting the A and C terms across cotwins to represent familial resemblance. Specifically, as MZ twins are 100% similar in genetic make-up whereas DZ twins are 50% similar on average, the cross-twin A_1 to A_1 and A_2 to A_2 links are 1 for monozygotic twins and 0.5 for dizygotic twins. However, the cross-twin links of shared environmental terms, C_1 to C_1 and C_2 to C_2, are 1 for both MZ and DZ twins. In this way, the model could estimate genetic and environmental contributions from the cross-twin cross-variable covariance and its difference between the MZ and DZ twins. In fact, a higher cross-twin cross-variable correlation in MZ twins than DZ twins implies genetic contributions to the link. Also, the phenotypic correlation as well as independence of the two variables is dissected into genetic, shared environmental and nonshared environmental components. It should be noted that interpretation of the nonshared environmental paths should be cautious, because they include measurement errors.

Apart from the Cholesky decomposition model path estimates, genetic correlation (r_A), shared environmental correlation (r_C), and nonshared environmental correlation (r_E) reflect the extent to which genetic and environmental factors overlap or are unique to the two variables. How these correlations are specified in a multivariate model is shown in Figure 1.4.

Figure 1.4.a. A model with genetic correlation (r_A), shared environmental correlation (r_C), and nonshared environmental correlation (r_E) specified (full view)

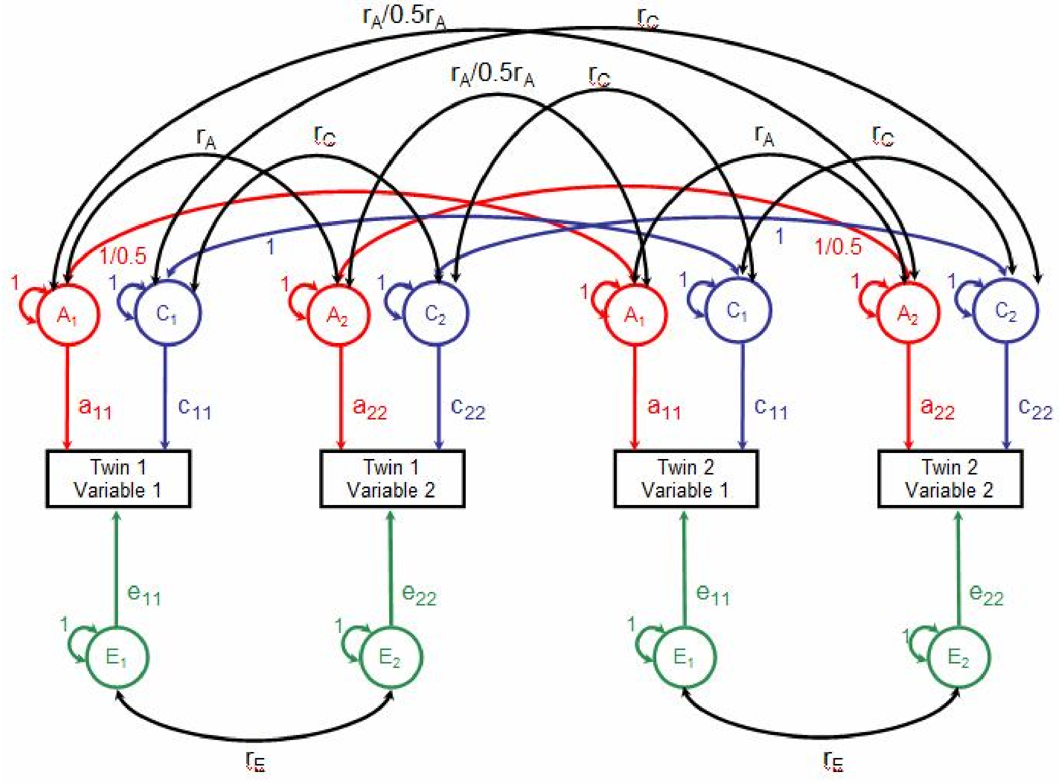
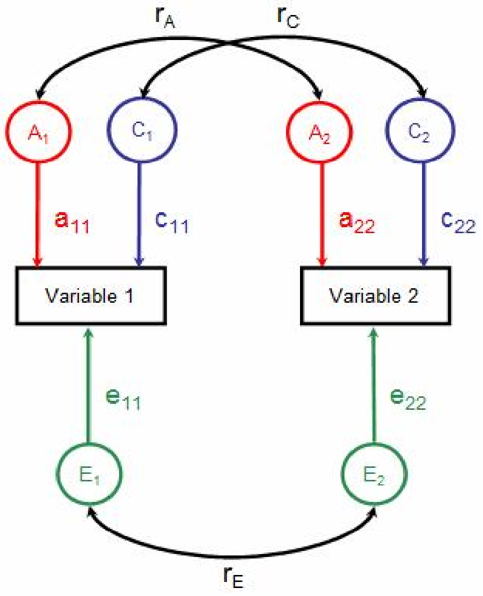


Figure 1.4.b. A model with genetic correlation (r_A), shared environmental correlation (r_C), and nonshared environmental correlation (r_E) specified (simplified view)



As indicated in the full view figure, genetic correlation (r_A), shared environmental correlation (r_C), and nonshared environmental correlation (r_E) link the two sets of A, C or E terms. Regarding genetic connections, there are four genetic correlation links in the model: within-twin performance across variables for each of the cotwins (2 links) and cross-twin performance across variables (2 links, i.e., twin 1 variable 1 to twin 2 variable 2, and twin 2 variable 1 to twin 1 variable 2). These also apply to the shared environmental correlation. However, nonshared environmental correlation only links within-twin performance, but not cross-twin performance. The cross-twin genetic links are multiplied by .5 for dizygotic twins, because their genetic make-up is 50% similar on average. A genetic correlation (r_A) of 1 indicates that all genetic factors which exerted impacts on variable 1 also influence variable 2, and a genetic correlation of 0 implies that genetic factors contributed to variable 1 are independent from those to variable 2. Similarly, a shared environmental correlation (r_C) of 1 shows a complete overlap of environmental factors which make cotwins more similar in the performance of the two variables, and that of 0 implies environmental factors which make cotwins more similar in variable 1 differ from those in variable 2. A significant nonshared environmental correlation (r_E) suggests the environmental factors which are nonshared between cotwins influence performance of the two variables. To reiterate, cautious interpretation of nonshared environmental correlation should be made, because it includes measurement errors.

These correlations represent the proportion of genetic or environmental factors in common across two variables, and thus show the extent of genetic or environmental overlap between the two variables clearly. However, a high genetic or environmental correlation does not necessarily means a strong phenotypic link. A high genetic correlation and a weak phenotypic link could happen, when both variables are influenced by the same sets of genetic factors which have small contributions to each variable. To indicate the impacts of genes with both genetic overlap and phenotypic link considered, bivariate heritability should be used. Bivariate heritability is the genetic contributions to the phenotypic correlation. For instance, a bivariate heritability of .3 of a phenotypic correlation of .6 means genetic factors contribute to 50% of the phenotypic correlation. It could be derived from the genetic correlation and the univariate genetic contributions, by the product of a_{11}, a_{22} and r_A (see Figure 1.4.b). The shared environmental and the nonshared environmental contributions to the phenotypic correlation could be computed in the same way (the product of c_{11}, c_{22} and r_C and the product of e_{11}, e_{22} and r_E for shared environment and nonshared environment respectively). Though not explicitly shown in the Cholesky decomposition model, the bivariate heritability can be converted from its path estimates. In the Cholesky decomposition model, the product of path estimates a_{11} and a_{21} (see Figure 1.3.b) represents bivariate heritability. In the same vein, environmental correlations and contributions to phenotypic correlations can be computed from the path estimates.

The Cholesky decomposition model, genetic correlation (r_A), shared environmental correlation (r_C), and nonshared environmental correlation (r_E) are employed to investigate the etiological overlap between general language and general reading abilities (Chapter 3), and between parent-rated communicative skills and motivation, and word reading and general reading abilities (Chapter 6). Also, they are used to examine the longitudinal stability and instability in various Chinese language and reading skills, including word reading, receptive vocabulary, phonological memory, tone awareness, syllable and rhyme awareness, rapid automatized naming, morphological awareness, and orthographic skills (Chapter 5).

Major assumptions and criticisms of twin design

Like all other statistical methods, the twin design has its assumptions and criticisms. In general, there are five major assumptions.

Equal environments assumption

The equal environments assumption supposes that the twin pairs share equally similar prenatal and postnatal environments regardless of their zygosity. In order words, it assumes that the shared environmental effects on MZ and DZ twins are equally similar. With this assumption, the difference between MZ and DZ twins' resemblance can be attributed to the contrast between MZ and DZ twins' genetic origins, and heredity can thus be indicated. However, violation of this assumption will lead to inaccurate heredity estimation.

Cotwins of MZ twin pairs are more likely to experience greater prenatal competition as they tend to share the same chorion. So, they tend to have greater weight and body size differences, compared to their DZ counterparts (Rutter, 2006). Because these experiences are assumed to be shared between twin pairs, these different experiences MZ twin pairs have will lead to an underestimation of heredity. Perhaps greater concerns are raised on the possibility that MZ twins may share more similar experiences than DZ twins. If MZ twins' shared experiences are more alike than DZ twins' ones, for example MZ twins are treated more similarly, heredity will be inflated because twins' resemblance influenced by these experiences will be mistakenly involved as a part of genetic effects. Note that this assumption focuses on the features of environments which are relevant to the traits studied. For instance, in a study of reading ability, if MZ twins have more similar home literacy environments than

DZ twin pairs, this might be an important violation of this assumption, but if MZ twins are dressed more similarly than DZ twins, this is not likely to be relevant to reading development.

Studies examined the equal environments assumption by comparing twins who were correctly and incorrectly classified as MZ or DZ by themselves or their parents (e.g., Kendler, Neale, Kessler, Heath, & Eaves, 1993; Munsinger & Douglass, 1976). They found comparable resemblance on psychological traits of misclassified twins (MZ twins who were mistakenly labeled as DZ twins or vice versa) with that of the correctly classified twins. It suggests MZ and DZ twin pairs experience equally similar shared environments, or differences in environmental similarity have negligible effects on biasing twin study results if they exist.

Additive genetic effects

The logic of the twin study design outlined so far assumes that genetic effects are additive. In other words, the effects of genes combine in a linear manner. With this assumption, heritability can be estimated by doubling the difference of the correlation of MZ twins and that of DZ twins. However, if there are nonadditive genetic effects, such as dominance genetic influences, heritability will be overestimated. Though twin studies generally focus on narrow-sense heritability, non-additive genetic effects can be estimated if the evidence suggests they are influential. For instance, nonadditive genetic effects are likely when the estimated heritability exceeds the MZ twin correlation. In model fitting, non-additive genetic effects can be estimated by utilizing the fact that MZ twins share all dominance genetic variations while DZ twins share one-forth of them on average, but only when shared environmental effects are nonsignificant and are thus dropped in the model.

No assortative mating

The twin study method assumes that there is no genetically relevant assortative mating. That is, mating is random and couples are not more genetically similar than two independent individuals. If assortative mating happens, the actual genetic relatedness of MZ twins will be unaffected, while that of DZ twins will be greater than the expected .5. This will lead one to underestimate heritability and overestimate shared environmental effects. Assortative mating can be reflected by spouses' resemblance. Studies have indicated assortative mating effects on cognitive ability, such as IQ scores (Bouchard & McGue, 1981). To account for the effects of assortative mating, the expected genetic correlation between DZ twins can be adjusted based on the resemblance between their parents.

Negligible genotype-environment correlation and interaction

The twin study method aims to tease genetic and environmental influences apart by assuming negligible genotype-environment correlation and interaction. Genotype-environment correlation occurs when the exposure to different environments is at least partially determined by an individual's genotype. For instance, a child who is genetically at risk of language difficulty may also have impoverished language learning environment at home, and these jointly contribute to the poor language ability of the child. Scarr and McCartney (1983) proposed three types of genotype-environment correlation. A child can passively receive both genes and environments form his or her parents (passive), evoke experiences from others' reactions to his or her behaviours according to his or her genotypes (evocative), or actively search for environments which link to his or her genotypes (active). Studies have shown family environment variables, such as parent-child interaction, could be partly influenced by genetic factors (Plomin, Reiss, Hetherington, & Howe, 1994). The heritability of family environment was found to be 18 to 30% in a review paper (Kendler & Baker, 2007). Genotype-environment correlation could lead to spurious findings in twin design, as environmental effects indicated could be genetic ones in origin. Note that the traits which are highly influenced by genetic factors are more prone to this concern. Nevertheless, researchers have pointed out that this genotype-environment correlation was negligible and results of twin research could reflect the true shared environmental influences (Petrill et al., 2004). In general, with the potential genotype-environmental correlation borne in mind, the twin study method helps identify the sources of environmental influences and demonstrate their roles in individual differences with genetic impacts considered.

Genotype-environment interaction suggests environments modify the expression of genotypes. If genotype-environment interaction is influential, environments will contribute to the development of different phenotypes from a genotype, and genes will exert greater influences in certain environments. Models of genotype-environment interaction include Diathesis-stress and Bioecological models, with the former suggesting higher heritability in impoverished environments, while the latter predicts greater heritability in enriched environments. Studies on language skills have found tentative evidence for genotype-environment interaction between 6p22 and 15q21 loci, and different environmental measures, with the effects of home language and literacy environment supporting the Bioecological model, while the effects of number of ear infections substantiating the Diathesis-stress model (McGrath et al., 2007). Thus, genotype-environment interaction will lead to inflated or underestimated heritability depending on the particular environments involved.

Generalizability of twin data

Twin study supposes that the trait variances of twins and singletons are similar and research findings from twin data can generalize to the singleton population. However, concerns about the generalizability of twin data have been raised because twinship may have specific effects on development. Regarding physical condition at birth, twins are born around 3 weeks prematurely on average and weigh less at birth compared to singletons (Bleker, Breur, & Huidekoper, 1979; Luke & Keith, 1992). Apart from this, twins often show lower language ability, such as using shorter sentences and producing fewer utterances, than singletons (Conway, Lytton, & Pysh, 1980). Rutter and colleagues (2003) found that the language ability of twins was 1.7 months and 3.1 months lagging behind that of singletons at age of 20 months and 36 months respectively. Prenatal factors (e.g., poorer physical condition at birth) might contribute to the slower language development in twins, but recent studies found that they have less or even no effect on language development (Rutter, Thorpe, Greenwood, Northstone, & Golding, 2003). Postnatal factors (e.g., reduced parents'oral communication and interaction with twins) tend to be more important in explaining the contrast between twins' and singletons' language skills (Lytton, Conway, Sauvé, 1977; Thorpe, Rutter, & Greenwood, 2003). Note that this twin-singleton difference is within the normal-range variations and is not great enough to prevent the use of the twin design to study the genetic origins of language development, though its potential influences should be borne in mind (Rutter, 2006)

Twinning in different cultures

Twinning rates differ across ethnic groups, and are highest for blacks (1.32%) and lowest for Asians (0.72%), and fall between these two ends for whites (1.01%) (Pollard, 1995). The twinning rate is 7.18 per 1000 births for Chinese (Pollard, 1995). The number of multiple births has increased over time across countries, partly owing to the use of in vitro fertilization and hormonal induction of ovulation (Imaizumi, 2003). This rising trend was observed in both DZ and MZ twins in different countries, though MZ rates remained stable in some countries (Imaizumi, 2003). The DZ and MZ twinning ratio tended to be lower in Asian populations. For instance, the number of DZ twins was double that of MZ twins in England and Wales, whereas they were about the same in Japan, in 1998 (Imaizumi, 2003). In Hong Kong, the DZ to MZ twinning ratio was found to be 0.86 in 1994 and 1995 (Tong, Caddy, & Short, 1997).

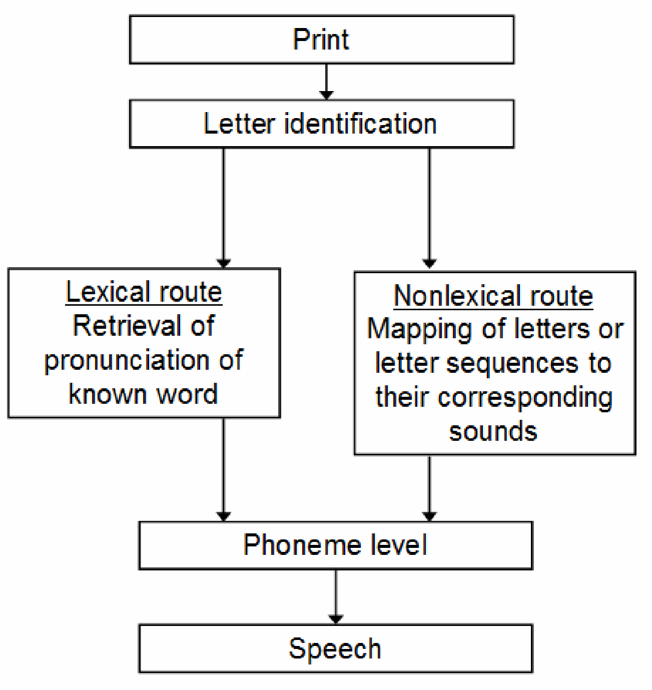
Chinese character reading in the dual route model context

This section aims to discuss important characteristics of Chinese character reading in the framework of the dual route approach. The discussion of Chinese character reading in the dual route model context is mainly motivated by the possible ways in which novel Chinese characters could be read and learnt, and the differential findings on whether both surface and phonological dyslexia appeared in Chinese reading difficulties (e.g., Butterworth & Wengang, 1991; Ho, Chan, Chung, Lee, & Tsang, 2007). There are alternative connectionist approaches (e.g., the “triangle” connectionist approach, Seidenberg & McClelland, 1989), but they also have an underlying architecture that distinguishes between orthography-phonology and orthography-lexicon mappings.

Dual-route approach to reading

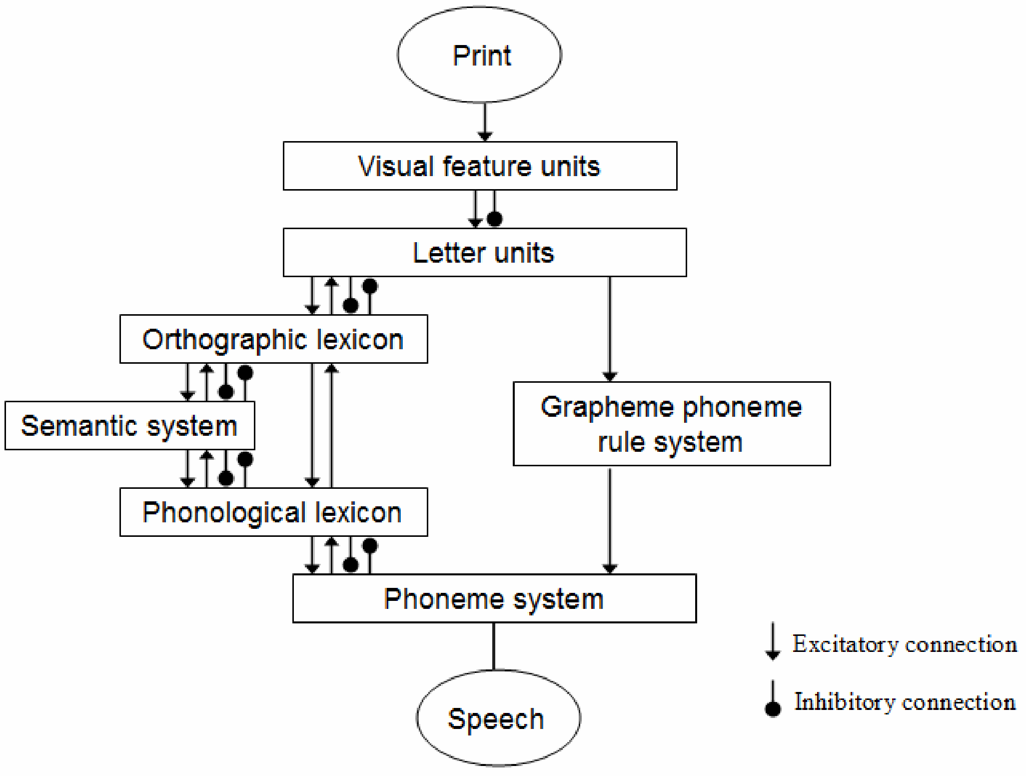
The dual-route model was motivated by the two distinctive processes involved in reading words which differ on their regularity in grapheme-phoneme correspondence (Coltheart, 2005). Specifically, a pseudoword (e.g., sare), which does not exist in the mental lexicon, is read by mapping letters or letter strings to phonemes with grapheme-phoneme correspondence rules, and this strategy refers to the nonlexical route in the dual-route model. This also applies to reading a novel word that one has not encountered before. On the contrary, an exception word is irregular in terms of grapheme-phoneme mappings (e.g., yacht), and thus relying on the grapheme-phoneme correspondence rules will produce a wrong pronunciation. So, it can only be pronounced correctly by searching its representation in the mental lexicon, and this strategy is considered as the lexical route in the dual route model. Reading of a regular word (e.g., cat) can involve both the nonlexical and the lexical strategies. Figure 1.5 depicts a generic dual route approach to reading aloud.

Figure 1.5. A generic dual-route approach to reading aloud. (Diagram adapted from Jackson & Coltheart, 2001, p.41)



A more specific dual-route approach, Dual-Route Cascaded model (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001), is shown in Figure 1.6.

Figure 1.6. Dual-Route Cascaded model. (Diagram adapted from Coltheart, 2005, p.12)



The Dual-Route Cascaded model is a computational model which captures two procedures of processing from print to speech, and they are a lexical route and a nonlexical route. When a print is encountered, different modules will be activated depending on the print's nature. Activations pass from one module to another continuously instead of relying on thresholds to be met in this model. The process begins with visual feature analysis of components in the letter string, and these analyzed components then activate the corresponding letter representations in the letter units module. The process then diverts to one of the routes depending on the output of the letter units module. When a familiar real word (regular or irregular) is viewed, the output of the letter units module activates the word's representation in the orthographic lexicon, which possesses a distinct unit for each word. The orthographic lexicon then contributes to the activation of its corresponding unit in the phonological lexicon directly or through the semantic system. The phonological lexicon units further contribute to the activation of the phoneme units which they are connected to in the phoneme system, and this leads to the word's pronunciation. Note that both excitatory and inhibitory connections are made, and the outputs are fed both forward and backward. For instance, the word unit activated in the orthographic lexicon provides feedback to the letter units module by activating the letter units which it is connected to and inhibiting the letter units which it does not contain. Through these continuous forward and backward flows of excitatory and inhibitory connections, the selection of phoneme units is fine-tuned, and at last the activations of correct phonemes reach a critical level and a reading response is given. This procedure is considered as the lexical route.

When a novel word, a pseudoword, or a nonword is encountered, the output of the letter units module fails to connect with any units in the orthographic lexicon. So, the reading process relies on the nonlexical route. It starts with decoding the first letter of the letter string into its phoneme according to the grapheme-phoneme rules, and then activating the corresponding phoneme unit in the phoneme system. This process operates serially from left to right across the letter string. Note that the lexical and nonlexical routes operate simultaneously. In the condition of which the two routes provide conflicting outputs of the phoneme system, such as irregular words, the reading response depends on the relative strengths of the lexical and the nonlexical routes.

The dual route model provides satisfactory accounts to reading in English, but its generalizabiltiy to other languages has been challenged. Given English as an alphabetic language with relatively high irregularity in grapheme-phoneme correspondence, the regular-irregular dualism that inspired the dual route model may only be a special observation in English (Share, 2008). So, two separate routes for regular and irregular words may be inappropriate for reading other scripts, especially the highly regular or highly irregular ones. For highly regular scripts, such as German, exception words are rare and the majority of words can be pronounced by grapheme-phoneme mappings, and so the nonlexical route can be adequate for reading and a separate lexical route is not necessary (Ziegler & Goswami, 2005). For scripts which are even more irregular than English, such as nonalphabetic scripts (e.g., Chinese) in which graphemes do not directly map onto phonemes, the lexical route may play the main role in reading and the nonlexical route may be redundant. In both cases, it is speculated that a single route can be sufficient for reading.

Whether the dual route model can generalize beyond English and be successful in demonstrating Chinese reading depends on the following three issues, 1) how does the nonlexical route operate in Chinese reading; 2) are the lexical and the nonlexical strategies independent in Chinese reading; and 3) does Chinese reading involve strategies which are beyond those in English reading.

How does the nonlexical route operate in Chinese reading?

In English, a letter is considered as the basic orthographic unit, and a letter or a letter string maps onto a phoneme. The nonlexical route represents the grapheme-phoneme correspondence. However, there is no equivalent in Chinese, which is a logographic script. In Chinese, the basic graphic unit is a character which simultaneously maps onto a syllable and a morpheme. A Chinese character and its components (strokes or radicals) do not represent phonemes. Thus, the grapheme-phoneme mappings are irrelevant in reading Chinese, and orthography represents syllables in Chinese instead. The different orthography-phonology mappings across the two scripts correspond to the research findings that phoneme awareness and syllable awareness are relatively more important in English reading and Chinese reading respectively (McBride-Chang, Bialystok, Chong & Li, 2004).

Among the Chinese characters, only the semantic-phonetic compound characters (also called ideophonetic compound characters), which constitute around 80% of the Chinese characters (Hoosain, 1991), contain cues to pronunciation. A semantic-phonetic compound character consists of a semantic radical and a phonetic radical, which provide cues to its meaning and pronunciation respectively. For instance, the semantic-phonetic compound 洋 (ocean /joeng4/[1]) has a phonetic radical on its right 羊 /joeng4/ and a semantic radical 氵 meaning water on its left. Thus, the nonlexical strategy in Chinese refers to pronouncing characters with the cues provided by their phonetic radicals, and is relevant to most, but not all, of the characters. It should be noted that phonetic radials provide cues to the character as a whole rather than a part of it. Phonetic radical can be a stand alone character itself (e.g., 羊 /joeng4/) or a bound phonetic which itself does not appear as an independent character (e.g.,畺). Among characters used in elementary school, 92% of semantic-phonetic compounds have a ‘stand alone’ phonetic radical, and 8% of them have a bound phonetic radical (Chen, Anderson, Wu, & Xuan, 2003). Reading a character is achieved by either deriving the pronunciation directly from the phonetic radical (e.g., 洋 /joeng4/ with a phonetic radical 羊 /joeng4/), or through analogy of characters in the same phonetic family (i.e., characters which share the same phonetic radical; e.g., 僵, 疆, 韁 which share the phonetic radical 畺 are

---

[1] Cantonese is transcribed in Jyutping, or Cantonese romanization, standardized by the Linguistic Society of Hong Kong in 1993.

pronounced as /goeng1/). While both strategies can be applied to the ‘stand alone’ phonetic radicals, only analogy is relevant for the bound phonetic radicals. Therefore, in order to obtain the cues from the phonological radical, one should have already mastered the character which appears as the phonological radical, or different characters which share the same phonetic radical with the character encountered. As the phonetic radical takes the main role in the nonlexical route, and they can be character themselves, it is believed that the nonlexical route should be labelled as the sublexical route. The term ‘nonlexical route’ is used in this section so as to make the discussion on Chinese reading easier to be referred to the original dual route model.

Only around 40% of the Chinese characters can be successfully decoded by the orthography-phonology correspondence rules, because some phonetic radicals do not provide exact or useful pronunciation cues (Shu et al., 2003). Though these rules are not very reliable, they are the only conceivable strategies that the readers can rely on when a novel character is encountered. Their effects become more significant when the huge number of characters that a proficient reader needs to master is taken into account (around 4500 characters are in regular use; Liu, Chuang, & Wang, 1975). So, the nonlexical route is still relevant in reading Chinese. This nonlexical route may be more significant for an intermediate Chinese learner than a proficient or a beginning Chinese readers. It is because compared to the proficient reader, he or she tends to encounter relatively more novel characters than known characters, and so the nonlexical route is kept utilized, while compared to the beginning reader, he or she has a greater character base for obtaining phonological cues from phonetic radicals.

To conclude, the nonlexical strategy in Chinese refers to pronouncing characters with the cues provided by their phonetic radicals, and proficiency in Chinese may be linked to the significance of the nonlexical route.

Are the lexical and the nonlexical strategies independent to each other in reading Chinese?

The lexical and the nonlexical routes operate independently in English reading (Coltheart, 2005). The independency of the two routes has been supported by different subtypes of reading disability (Castles & Coltheart, 1993). Surface dyslexics suffer from an impaired lexical route, but their nonlexical route remains intact. They are able to read pseudowords and nonwords, but fail to read irregular words. On the contrary, deep dyslexics and phonological dyslexics suffer from an impaired nonlexial route, but their lexical route is relatively intact (the semantic system mediating the lexical route is damaged in deep dyslexics). They are able to read irregular words (though meaning-related errors present among deep dyslexics), but not pseudowords and nonwords. This dissociation is less clear in Chinese dyslexics compared to their English counterparts. While some studies showed the presence of two distinct groups of dyslexics (e.g., Butterworth & Wengang, 1991), recent studies identified only surface dyslexia, but not phonological dyslexia among dyslexics (e.g., Ho et al., 2007). The failure to find dyslexics having problems with the nonlexical route only leads to two speculations, a) only the lexical route is involved in reading, or b) the operation of the nonlexical route highly depends on the lexical route.

As discussed before, the nonlexical strategy in Chinese refers to pronouncing characters with the cues provided by their phonetic radicals. Obtaining phonological cues from a phonetic radical requires the mastery of either the character which appears as the phonological radical, or different characters which share the same phonetic radical with the target character. In other words, the phonetic radical is only useful when relevant referents are available in the mental lexicon. Thus, while the nonlexical route works through the orthography-phonology correspondence rules, the lexical route is involved, as a reference for the phonological cues. The role of the lexical route on the nonlexical route is underscored when tone is considered. Chinese is a tonal language, in which a syllable in different tones represents different meanings and each of them is denoted by a different character (there are six lexical tones in Cantonese). For instance, the syllable /ji/ in different tones refers to the following meanings: /ji1/ (clothing) /ji2/ (chair), /ji3/ (opinion), /ji4/ (son), /ji5/ (ear), and /ji6/ (two). However, tone is not explicitly represented in a character as there is no symbol in a character signifies tone. Therefore, simply decoding a character with orthography-phonology rules will only generate several possible candidates, but fail to produce the exact pronunciation because the orthography does not explicitly label tone. Therefore, at least some information of the character, such as tone, should be retrieved from the lexicon during reading.

Assuming the nonlexical route which is independent from the lexical route is possible, the nonlexical route should have its own pronunciation representations for each radical, such as ‘吉’ /gat1/ is mapped onto syllables /gat1/ (as the character 桔 pronounced) and /git3/ (as the character 結 pronounced), just like the English letter ‘c’ is mapped on /k/ or /s/. There are around 800 phonetic radicals (McBride-Chang, 2004), and thus the quantity of representations is greater in Chinese than in English. In an English pseudoword, letters can provide cues on whether ‘c’ is pronounced as /k/ or /s/ (‘c’ followed by ‘e’, ‘i’, or ‘y’ is pronounced as /s/). However, in a Chinese pseudocharacter (pseudocharacters described in this section refers to those consist of a semantic radical and a phonetic radical in their legal position, but this combination does not exist in Chinese; e.g., 垍), the components provide no cues on whether the phonetic radical '吉' refers to /gat1/ or /git3/. Therefore, the selection is either by choosing the most possible one (such as the one with the highest frequency) or by chance, whenever this phonetic radical is encountered. Nevertheless, this approach is very unreliable.

To conclude, the lexical and the nonlexical routes are linked. Specifically, the nonlexical route relies on the referents in the lexicon in the process of phonological decoding according to the orthography-phonology correspondence rules.

Does reading Chinese require strategies which are beyond those in reading English?

Studies have indicated remarkable similarity in the cognitive predictors of reading in English and Chinese, which differ from each other in characteristics substantially (McBride-Chang & Kail, 2002). For instance, phonological awareness at the syllable level was a strong predictor of Chinese and English reading abilities of Chinese children learning English as a foreign language, and also of English ability in English-speaking children, but general visual processing skills did not uniquely contribute to reading in these children (McBride-Chang & Kail, 2002). While some skills are fundamental across the two scripts, different strategies can emerge given the unique characteristics of the scripts. This notion is in line with the psycholinguistic grain size theory, which proposed English readers are encouraged to develop various strategies involving large and small grain size in reading, because of the irregularity of the grapheme-phoneme correspondence in English, while the small grain size strategy is sufficient for readers of highly regular scripts, such as German (Zeigler & Goswami, 2005). Different strategies may also be involved in Chinese reading. For instance, Chinese reading is much more analytical than English reading, and so analytical processing strategies may emerge in Chinese reading. To apply orthography-phonology correspondence rules in Chinese, analytical processing on at least three aspects is needed.

Analytical processing on the orthographic structure is required. The semantic and the phonetic radicals have their legal positions in a square-shaped character, mostly left-right or top-bottom structured. In a left-right structured character, the semantic radical always takes the left position, while the phonetic radical always takes the right position. In a top-bottom structured character, the semantic radical is always placed on the top, while the phonetic radical is always placed at the bottom. Therefore, the orthographic structure should be understood and analysed in order to locate the phonetic radical which contains pronunciation cues for the character. An orthography analysis should be done before the execution of orthography-phonology mappings in the nonlexical route.

Also, how phonetic radicals are utilized to read characters is ambiguous. Phonetic radicals convey phonological information in three ways. First, they provide full phonological information, that is, they possess the same pronunciation as that of the character (i.e., regular; e.g., 桔 /gat1/ has the same pronunciation as the phonetic radical 吉 /gat1/]). Second, they provide partial phonological information. Specifically, their pronunciation is similar to but not exactly the same as that of the character. Their tone, onset, or rime, is different from that of the character (i.e., semi-regular; e.g., 結 /git3/ has the same onset but different rime and tone as the phonetic radical 吉 /gat1/]). There are about 26% and 39% of the semantic-phonetic compounds taught in elementary school which are fully regular and semiregular respectively (Shu et al., 2003). Derivation and analogy strategies are useful to decode these characters. Nevertheless, in about 15% of the semantic-phonetic compounds used in elementary school, the phonetic radicals provide no direct or even no useful information in pronouncing the character (i.e., irregular; e.g., 海 /hoi2/ has different tone, onset and rime from the phonetic radical 每/mui5/) (Shu et al., 2003). Some of them consist of bound phonetic radicals which do not provide direct cues to pronunciation (e.g.,畺). However, characters sharing the same bound phonetic radicals often possess the same pronunciation, and thus making analogy with characters in the same phonetic family is helpful to reading (e.g., 僵, 疆, 韁 which share the phonetic radical 畺 are pronounced as /goeng1/). Therefore, Chinese readers have to be analytical in figuring out the amount of phonetic information that the phonetic radicals provide and the strategy (derivation or analogy) that is appropriate for a particular situation.

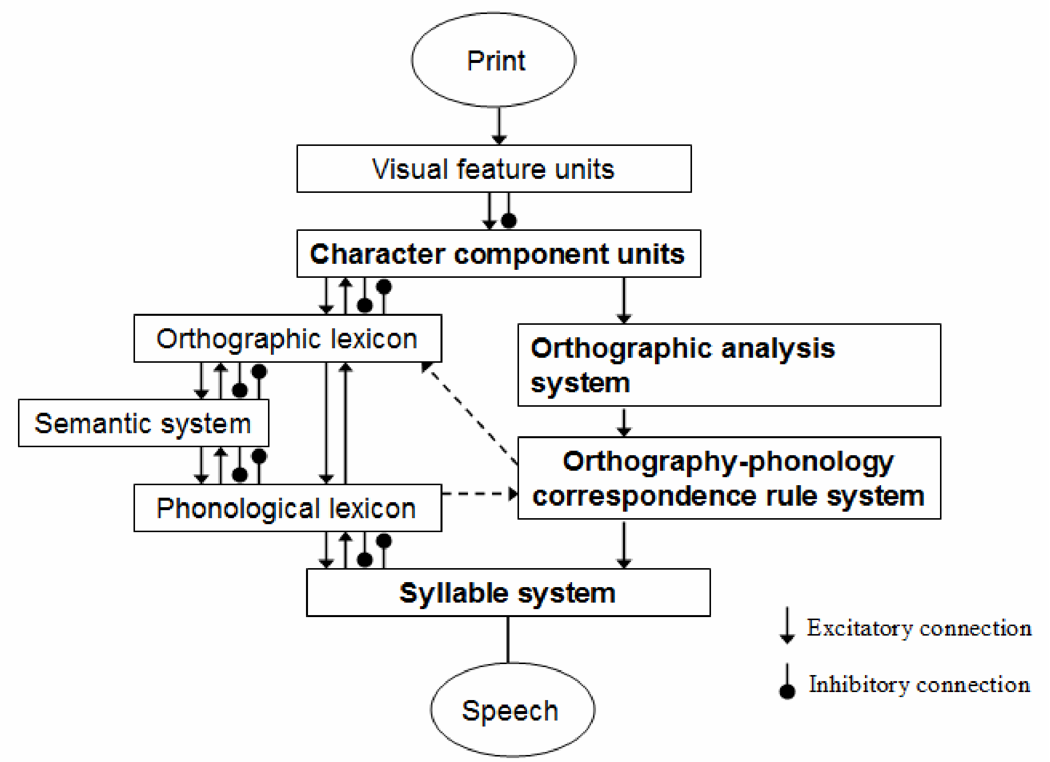
So, reading Chinese involves relatively more extensive analytical processing, including the processing of the orthographic structure of the character, how much phonetic information is provided, and which strategies should be applied. These analytical processing skills develop with reading experience, and explicit instruction could be useful. The dual route model in its current form is insufficient in explaining this analytical processing aspect in Chinese reading.

A hypothetical model for reading Chinese: A modified dual route model

Although the Chinese script is irregular and ambiguous in terms of the orthography-phonology correspondence, the nonlexical route plays a role in Chinese reading. However, the dual route model with two separate routes cannot fully explain Chinese reading. The operation of the nonlexical route highly depends on the lexical route in Chinese reading, and may involve an additional analytical system. Therefore, it is believed that the dual route model should be modified in order to provide a better account to the Chinese reading process. Figure 1.7 shows the modified Dual-route Cascaded model for Chinese character reading.

In the modified model, the letter units, the phoneme system and the grapheme-phoneme rule system of the original model are replaced by character component units, syllable system and orthography-phonology correspondence rule system respectively. The character components refer to the spatially separated constituting parts of a character. They are usually a part of the radical. The lexical route is the same with that in the original model. However, an orthographic analysis system is added before the orthography-phonology correspondence rule system in the nonlexical route. Within this system, the orthographic structure of the character is analysed, to locate the phonetic radical so that phonological cues can be extracted. The information is then passed to the orthography-phonology correspondence rule system. The phonological cues are obtained from the phonetic radical with reference to the mental lexicon. Therefore, additional connections flowing from the orthography-phonology correspondence system to the mental lexicon, and going back to the system form the mental lexicon are added. Based on the information obtained form the mental lexicon, the orthography-phonology system generates a syllable with either derivation or analogy strategies. Modifications of the Dual-Route Cascaded model involve redefining some basic units, adding an analytical processing unit to the nonlexical route, and adding the links between the lexical and the nonlexical routes. These modifications help model to provide a better account to the Chinese reading process.

Figure 1.7. A hypothetical model for Chinese character reading: A modified Dual-Route Cascaded model. (Modified or new components are in bold and newly added connections are in dash arrows.)



A pseudocharacter ⿰土吉 is used to illustrate how the nonlexical route may operate (only excitatory connections are described for simplicity). The visual features of ⿰土吉 activates the character component units, such as 土, 士 and 口. These character component units then enter the orthographic analysis system and the phonetic radical 吉 is located. The orthography-phonology correspondence rule system starts to obtain the phonological cue of 吉 by matching it with the relevant representations containing this phonetic radical in the mental lexicon (e.g., 吉, 結). The information is then sent back to the system and generates a syllable with either derivation or analogy strategies. As in this case the phonetic radical is a stand alone character, both strategies are relevant. The pronunciation of /gat1/ and /git3/ will be given through derivation and analogy respectively. Depending on the relative strengths of these two strategies on the particular phonetic radical, the stronger one will be chosen as the output of the syllable system.

This section aims to highlight important characteristics of Chinese character reading and summarize the differences between Chinese character reading and English word reading in the dual route approach framework.

Chinese literacy instruction in Hong Kong

Most Hong Kong children start learning to read and write Chinese characters at a very young age. The majority of Hong Kong children started learning to read and write at age 4 or below (98.4% and 91.9% for learning to read and write respectively) (Li & Rao, 2005). On entering kindergarten at around age three, they are taught Chinese characters using Cantonese pronunciation, mainly by look-and-say method (Fu, 1987; Li & Rao, 2000). As there is a lack of systematic phonological transcription approach to learning Cantonese pronunciation of characters, children are encouraged to learn characters as holistic units and rely heavily on rote memorization (Cheung & Ng, 2003). To learn a new character, they are required to copy it for several times to ensure that they can recognize and reproduce it (Chan & Wang, 2003). Some teachers show their students how to decompose a Chinese character to phonetic and semantic radicals, which provide pronunciation or meaning hints of the characters, but there is a lack of systematic guidelines for this practice (Cheung & Ng, 2003). In fact, the Hong Kong Education Bureau has included understanding basic phonological units, orthographic structure, and homophones as suggested learning objectives for primary school students (Hong Kong

Education Bureau, 2008), but there is no concrete teaching plans given on these more analytic approaches in Chinese character instruction. Thus, the teaching methods employed could vary greatly, especially across kindergartens. In general, the number of Chinese lessons in primary school is around 8 to 11 per week, with each lesson lasts for approximately 35 minutes (Cheung & Ng, 2003).

The demand of learning to read is mounting across age, with children in the first grade are expected to have learnt approximately 460 characters, and they are expected to know about 2080 characters by grade four (Hong Kong Department of Education, 1990). Note that characters combine together to form words, and so it is even more taxing in terms of word reading. Parents in Hong Kong generally have positive attitudes towards early literacy instruction and home literacy activities. The majority of parents taught their children to read Chinese characters at home (73%), and believed parent-child reading was useful to develop their child’s Chinese literacy skills (82.5%) (Li & Rao, 2000). These positive attitudes are helpful to children’s Chinese acquisition.

Chapter 2

Method

Participants

This thesis included typically developing Chinese twins who were a) aged from 3 to 11 (born between Nov 1997 and Nov 2004), b) of the same sex, and c) using Cantonese as their mother tongue. Twins were identified through kindergartens and primary schools in all of the 18 geographic districts in Hong Kong, to enhance the representativeness of this sample to the general population in Hong Kong. A list of kindergartens and primary schools for typically developing children, excluding international schools in which English was used as the main language of communication, was produced with reference to the school information available on the Hong Kong Education Bureau website. There were 912 kindergartens and 585 primary schools on the list. An invitation letter containing information of the research was sent to each of these schools by fax. Principals of these schools were then contacted individually on phone to provide them with further information of the study, and to invite them for participation. Due to the nature of this thesis, only schools which had twin students were included. Informed consent was obtained from the participating schools, and number of twin pairs in these schools was gathered. There were 120 kindergartens and 94 primary schools which had twin students participated.

Invitation letters and consent forms for parents describing the aims, procedure and implications of and participants’ rights in the research were sent to these schools. These documents were then distributed to twins’ parents through the schools. Parents were asked for their permission for their twins to take part in the research of this thesis. After receiving the parents’ informed consent, the names of participating twins and their contact information were obtained. The majority of twins were recruited through schools, but the research was also advertised in community centers and discussion forums for parents, in order to maximize the number of twins participating in the research. Parents recruited through community centers and discussion forums provided us with the names of the participating twins and their contact information, and were then given invitation letters and consent forms by post. There were 338 twin pairs participated in the first wave in 2007. All children were given an audiometric screening test to ensure they had normal-range hearing for speech frequencies. Zygosity was determined by SNP testing.

Among the 338 twin pairs, there were 182 male twin pairs, and 156 female twin pairs. Also, 181 twin pairs were kindergartners, and 157 twin pairs were primary school students. Five twin pairs with unreliable results in audiometric screening test were excluded. Among this sample, children who could not hear 35dB or above with the better ear were excluded, thereby yielding a sample of 312 twin pairs in the first wave. There were 228 pairs of monozygotic twins (116 male pairs and 112 female pairs; 110 kindergartner pairs and 118 primary school pairs) and 84 pairs of same-sex dizygotic twins (50 male pairs and 34 female pairs; 52 kindergartner pairs and 32 primary school pairs). See table 2.1.

Table 2.1. Number of MZ and DZ Twin Pairs by Gender and Educational Level in the First Wave.

	MZ	**DZ**	**Total**
Boy	116	50	166
Girl	112	34	146
Total	**228**	**84**	**312**
Kindergarten-level	110	52	162
Primary-level	118	32	150
Total	**228**	**84**	**312**

In the second wave conducted in 2008, 11 twin pairs dropped out, and children who could not hear 35dB or above with the better ear were excluded, thereby yielding a sample of 292 twins pairs in the second wave. There were 215 pairs of monozygotic twins (108 male pairs and 107 female pairs; 71 kindergartner pairs and 144 primary school pairs) and 77 pairs of same-sex dizygotic twins (45 male pairs and 32 female pairs; 30 kindergartner pairs and 47 primary school pairs). See table 2.2.

Table 2.2. Number of MZ and DZ Twin Pairs by Gender and Educational Level in the Second Wave.

	MZ	**DZ**	**Total**
Boy	108	45	153
Girl	107	32	139
Total	**215**	**77**	**292**
Kindergarten-level	71	30	101
Primary-level	144	47	191
Total	**215**	**77**	**292**

The same-sex dizygotic to monozygotic twin ratio in our sample was 0.37 in wave 1 and 0.36 in wave 2. With opposite-sex twin pairs included, the dizygotic to monozygotic twin ratios were around 0.65 for twins born to Chinese fathers or mothers in Singapore from 1986 to 2001 (Chia et al., 2004), and 0.86 for twins born in Hong Kong from 1994 to 1995 (Tong et al., 1997). Assuming there are equal number of same-sex and opposite-sex dizygotic twin pairs, the same-sex dizygotic to monozygotic twin ratios were around 0.33 and 0.43 respectively in these studies. So, the proportion of twin types in our study was comparable to that of the population prevalence.

Measures

Tasks tapping various aspects of Chinese language and literacy, and related cognitive skills, were administered. They included tasks of word reading, receptive vocabulary, phonological memory, rapid naming, syllable and rhyme awareness, tone awareness, morphological awareness, orthographic skills, and nonverbal reasoning ability. In the following tasks, unless specified otherwise, a correct response of a test trial was awarded one point, and failing a test trial referred to obtaining a zero mark for that trial.

Pilot testing

Pilot testing was conducted on 90 kindergartners and junior primary school students (15 in each grade level, six grades levels in total: all three grades in kindergarten, and the first to the third grades in primary school). There were several aims of the pilot testing. First, the pilot testing data was used for evaluating and refining tasks, to ensure that the measures were suitable for tapping the variability among the children. It is especially important for tasks which had not been administered to Chinese children in past research. Second, the participating children in this thesis included 3- to 11-year-olds, an age range that covered a broad range of language and literacy abilities. To prepare a battery of tasks which was able to tap individual differences across these ages, the pilot testing data was important for ranking items in increasing difficulty, and setting up entry points, basal rules, and ceiling rules for the measures. Third, it provided an estimation of time needed for administering these tasks.

However, some tasks were not included in the pilot testing. Items adapted from the Hong Kong Test of Specific Learning Difficulties in Reading and Writing (HKT-SpLD) (Ho, Chan, Tsang & Lee, 2000) were not administered because they had been developed and normed for Hong Kong Chinese children. Also, the Raven's Colored Progressive Matrices

(RCPM; Raven, Court, & Raven, 1995), and the speeded number naming task, had been administered successfully on Hong Kong Chinese children in past studies (e.g. Lin et al., in press; McBride-Chang, Cheung, Chow, Chow, & Choi, 2006, for RCPM; Chan, Ho, Tsang, Lee, & Chung, 2007; Chow, McBride-Chang, & Burgess, 2005; McBride-Chang & Ho, 2000, for speeded number naming), and thus were not included in pilot testing. Furthermore, the character recognition task was not included because its items were adapted from a longer list developed by Ho, Leung and Cheung (submitted) and permission to use their raw data for item selection was obtained.

Word reading

A 48-item character reading list and 150 items adapted from the reading subtest of the Hong Kong Test of Specific Learning Difficulties in Reading and Writing (HKT-SpLD) (Ho et al., 2000) were combined. The character recognition list, consisted of 31 single characters and 17 two-character words, was adapted from a longer list developed from characters and words in Chinese textbooks for Hong Kong kindergartners by Ho et al., (submitted). This thesis included the 48 items with the best discriminative indices, ranked in increasing difficulty, based on the data by Ho et al., (submitted). The HKT-SpLD was a standardized test developed for Hong Kong primary school children. Its reading subtest consisted of 150 two-character words in an order of progressive difficulty was used in this thesis.

The child was required to read each word aloud. Testing stopped when the child failed to read 15 consecutive items. Kindergartners started from the character reading list, and were given the items adapted from the HKT-SpLD if they progressed beyond this list. However, the first item of the HKT-SpLD reading subtest was regarded as the entry point for all primary school children. They were given the character reading list only if they failed to read the easiest 15 consecutive items. The maximum score of the combined task was 198, and its Cronbach's alpha was .996.

Receptive vocabulary

The receptive vocabulary test consisted of 2 practice trials and 80 test trials translated and adapted form the Peabody Picture Vocabulary Test – Fourth Edition (PPVT-IV**;** Dunn & Dunn, 2007). The 228 PPVT-IV items were translated into Chinese, and were administered in the pilot testing. Eighty items with the best discriminative indices were included in the main testing. For each trial, the experimenter read out the target item and the child was required to select a picture from the four options to match it. Every item had one target vocabulary accompanied by three distracters (a semantic and two unrelated distracters). An entry point for each grade level and a basal rule were set according to pilot testing data. The basal rule was fulfilled if correct responses were given in nine or all trials in the first 10 consecutive trials from the corresponding entry point. Testing stopped when the child failed 11 or all trials in 12 consecutive trials. The maximum score of this task was 80, and its Cronbach's alpha was .96.

Phonological memory

The nonword repetition task was used to assess phonological memory. This task consisted of a series of nonword strings ranging from two syllables to seven syllables. A nonword string was constituted by Cantonese syllables and had no lexical meaning as a whole (e.g., 芳令/*fong1 ling1*/). There were two practice trials and 14 test trials, ranked by increasing complexity based on the number of syllables involved, and the pilot testing data for items with the same number of syllables. For each trial, the child was presented a nonword string, in which the inter-syllable interval was 0.5 second, by a mp3 player. The child was then requested to repeat the nonword string in the exact order of syllables presented, and the response was recorded. The child's responses were scored according to the number and the order of correct syllables given. Specifically, for a nonword string, a point was given for each correct syllable, and also for each correct pair of consecutive syllables, but a point was deducted for each excessive syllable. So, the maximum score of each item was twice the total number of syllables minus one. Testing stopped when the child failed four consecutive items. The maximum score of this task was 124, and its Cronbach's α was .90.

Syllable and rhyme awareness

This test was composed of the syllable deletion and the rime detection tasks. The syllable deletion task consisted of three blocks of trials in an increasing difficulty order: real word, nonwords and nonsense words. A real word was combined by Cantonese syllables and had lexical meaning (e.g., 望遠鏡 /*mong6 jyun5 geng3*/ [binoculars]). A nonword was constituted by Cantonese syllables and had no lexical meaning as a whole (e.g., 女任綠 /*neoi5 jam6 luk6*/). A nonsense word was created from nonsense syllables that conformed to the phonological constraints of Cantonese, and neither the nonsense syllables nor the compound as a whole had lexical meaning (e.g., /*fou2 moi1 peng5*/). There were two practice trials for the real word block, and one practice trial for the other two blocks. Each block contained five test trials, in which two of them were two-syllable items and three of them were three-syllable items. The items were orally presented by the experimenter and the child was required to produce an answer orally with one syllable taken away from the compound words. In each block, two trials required deletion of the first syllable, two trials required deletion of the last syllable, and one trial required deletion of the middle syllable. For example, the real word ‘望遠鏡’ /*mong6 jyun5 geng3*/ (binoculars) without ‘遠’/ *jyun5*/ is ‘望鏡’ /*mong6 geng3*/. The target answers of some real words were meaningful words. The maximum score on this task was 15.

The rime detection task consisted of two practice trials and nine test trials. For each item, the experimenter read out a target syllable, and then read out three syllables and simultaneously showed three pictures illustrating each of them. The child was required to select a syllable from the three options which rhymed with the target syllable. For example: ‘人’ /*jan4*/ (human) was read out as the target syllable, and ‘牙’ /*ngaa4*/ (tooth), ‘猴’ /*hau4*/ (monkey) and ‘雲’ /*wan4*/ (cloud) were then presented with their illustrations. The child was asked which of the three syllables rhymed with ‘人’ /*jan4*/ (human), and the answer was ‘雲’ /*wan4*/ (cloud) (see Appendix 1A). The maximum score of this task was nine. The score of the combined task was 24, and its Cronbach’s α was .88.

Rapid automatized naming

The speeded number naming task was employed to assess rapid automatized naming ability. It consisted of six rows of five digits (2, 4, 5, 7, 9). These digits were arranged in different orders for each row (see Appendix 1B). The child was instructed to name all digits at the fastest speed possible, and was timed with a stop watch. Two trials were completed, and the time was averaged. The measure for analyses was (1/ average time), so a higher score indicated a better rapid automatized naming ability, proportional to number of words read per unit time.

Morphological awareness

Three tasks of morphological awareness were combined to tap morphological awareness. They were arranged in an order of increasing difficulty based on the tasks' nature: the receptive morphological construction task, the morphological construction task, and the homophone task. Items of each task were ordered in progressive difficulty according to the pilot testing data. Testing stopped when the child failed four out of five consecutive items.

The receptive morphological awareness task consisted of 2 practice trials and 10 test trials. For each item, five pictures were presented to the child simultaneously and the experimenter read out the target item which represented a novel concept created by combination of morphemes. The child was required to select a picture from the five options which illustrated the target item. Every trial had one target item accompanied by four distracters: a) a real object or action that shared a morpheme with the target; b) a novel concept that shared a morpheme with the target c) two separate real objects or actions each of which represented a morpheme of the target; d) two separate real objects or actions in which one of them represented the target morpheme while the other one represented an irrelevant morpheme used in b). For example, in one trial of this task, the target item was a striped elephant (斑象 /*baan1 zoeng6*/) which represented a new concept of an elephant with stripes on its body. Five pictures showing (a) a zebra (斑馬/*baan1 maa5*/), (b) a striped dog (斑狗 /*baan1 gau2*/), (c) stripes and an elephant (斑+象/*baan1*/ + /*zoeng6*/), (d) a dog and an elephant (狗+象 / *gau2*/ + /*zoeng6*/), and (e) a striped elephant (斑象 /*baan1 zoeng6*/), were presented (See Appendix 1C for this example item). Selection of the distracter (a) indicated difficulty in interpreting a word in a novel morphological context; (b) showed difficulty in interpreting both morphemes, but was able to interpret a word in a novel morphological context; (c) indicated ability to interpret both morphemes, but difficulty in manipulating them in a novel morphological context, and (d) showed difficulty in interpreting both morphemes and manipulating morphemes in a novel morphological context. The maximum score of this task was 10.

The morphological construction task consisted of 1 practice trial and 12 test trials. For each item, a scenario was orally presented by the experimenter, and the child was asked to actively construct words for the newly presented objects or concepts according to the scenarios. The practice trial was aided with illustration. One test trial is indicated below as an example. “An island that is full of yellow chrysanthemums, is called a yellow chrysanthemum island (黃菊島 /*wong4 guk1 dou2*/). What will we call the island which is full of red peach blossom?” The answer is a red peach blossom island (紅桃島 /*hung4 tou4 dou2*/). A correct answer had all core morphemes combined in a right order with no redundant or irrelevant morphemes included, and was awarded two points (e.g., 紅桃島 /*hung4 tou4 dou2*/ [red peach blossom island]). A response with all core morphemes selected but combined in a wrong order (e.g., 桃紅島 / *tou4 hung4 dou2*/) [peach blossom red island] or included any redundant or irrelevant morphemes (e.g., 紅桃菊島 /*hung4 tou4 guk1 dou2*/ [red peach blossom chrysanthemum island]), or that with a core morpheme omitted but was correct in combination order and had no redundant or irrelevant morphemes (e.g., 桃島 / *tou4 dou2*/ [peach blossom island]), was considered as a partially correct answer and was awarded one point. The maximum score of this task was 24.

The homophone task consisted of one practice trial and five test trials. For each trial, a character was orally presented in the context of a word, and the child was required to produce as many words constituting this character as possible in 10 seconds. The child was then asked to produce as many words constituting the homophones of this character as they could in 10 seconds. For example, in one trial of the task, the target character was ‘兒’ /*ji4*/ of the word ‘兒童’ (children /*ji4 tung4*/). The words constituting this character could be ‘兒子’ (son /*ji4 zi2*/), ‘兒歌’ (children’s song /*ji4 go1*/) etc., while the words composed of its homophones could be ‘姨媽’ (aunt /*ji4 maa1*/), ‘懷疑’ (suspect /*waa14 ji4*/) etc. Two points were given for at least one correct word produced in both parts, while one point was awarded for at least one correct word given in one part only. The maximum score of this task was 10. The maximum score of the combined task was 44, and its Cronbach’s α was .90.

Orthographic skills

Two tasks of orthographic skills were combined, ranked in increasing difficulty based on the tasks’ nature: a left-right reversal task and a lexical decision task. Items for these tasks were adapted from the Hong Kong Test of Specific Learning Difficulties in Reading and Writing (HKT-SpLD; Ho et al., 2000).

The left-right reversal task assessed the knowledge of correct orientation of highly frequent orthographic units in Chinese characters. This task consisted of 21 simple Chinese characters and 4 alphabetic numbers, of which 14 of them were left-right reversed. The child was presented all items simultaneously arranged in a five by five matrix, and required to cross out items with an incorrect orientation with a pencil. A point was given for an item identified correctly, i.e., an incorrect oriented item was crossed out or a correct oriented item was left uncrossed. The maximum score of this task was 25.

The lexical decision task assessed the knowledge of Chinese character structure. This task consisted of 30 rare real characters, and 30 noncharacters with radicals placed in illegal positions. All of them are left-right structured composed of two radials. A real left-right structured Chinese character normally had a semantic radical on the left and a phonetic radical on the right. A noncharacter was the combination of two semantic radicals, two phonetic radicals, or a semantic radical and a phonetic radical in their illegal positions, all of which were illegal in Chinese character structure. The child was presented items arranged in 12 rows of 5 items each on 2 separate pages, and was required to cross out the noncharacters. A point was given for an item identified correctly (i.e., a noncharacter was crossed out or a real character was left uncrossed). The maximum score on this task was 60. The maximum score of the combined task was 85, and its Cronbach's α was .93.

Tone awareness

The Cantonese tone task consisted of 3 practice trials and 15 test trials, administered with a computer. There were three blocks of test trials arranged in the following order: three-syllable, two-syllable, and one-syllable blocks, and each had five trials. The three-syllable block was presented first, while the one-syllable block was shown last, because the more syllables given could provide more cues on identifying the correct tones and were thus relatively easier. For each trial, three pictures each illustrated a syllable (in the one-syllable block)/ a group of syllables (in the two-syllable and three-syllable blocks), in which these syllables/ these groups of syllables had different tones, were shown. The child was required to label each of them, and was given the syllables if they were not able to label them correctly. This procedure was to ensure that the child knew the syllables represented by the pictures before proceeding to the actual tone test. Then, a sound of a lexical tone (in the one-syllable block)/ of a group of lexical tones (in the two-syllable and three-syllable blocks) was presented, and the child was asked to select the picture representing the syllable(s) which matched with the sound of lexical tone(s). For instance, in a one-syllable trial, a Cantonese first lexical tone sound (i.e., high-level tone sound) was presented, and three pictures illustrated a letter (信 /*seon3*/), a lock (鎖 /*so2*/) and a pig (豬 /*zyu1*/) respectively, were shown. The child was then asked which of the three options had the same tone as the lexical tone sound. The answer was a pig (豬 /*zyu1*/) which had a Cantonese first tone (see Appendix 1D). The maximum score was 15, and its Cronbach's α was .66.

Nonverbal reasoning

The Raven's Colored Progressive Matrices (RCPM; Raven et al., 1995) was an assessment of nonverbal intelligence in children between five to eleven years old. The test was composed of three subsets A, Ab, and B, each consisting of 12 items. For each item, the child was shown a visual matrix with a missing part and six options with the size of the missing part simultaneously. The child was requested to select from the six options the one that fitted the pattern of the visual matrix. This task had been successfully administered to Hong Kong kindergarteners including those who were under 5 years old in past studies (e.g., Lin et al., in press; McBride-Chang et al., 2006). Thus, it was administered to children as young as four years old in this thesis. As this task had not yet been normed in the Chinese population, raw scores were used. There were two trial items in Set A and they were not included in the total scores. The maximum score was 34, and its Cronbach's α was .93. To adjust for age effects, the raw scores of RCPM were regressed on children's age, and the standardized residuals represented children's nonverbal reasoning ability.

Audiometric screening

An audiometric screening test was administered using Interacoustics Manual Audiometer AS208 or DSP Pure Tone Audiometer to ensure subjects had normal-range hearing speech abilities. The child was put on earphones on both ears, and was asked to raise his/her hand when a pure tone sound was heard. The experimenter administered the test behind the child so that no visual cues were given on whether a pure tone sound was played. The test started with the right ear at 40dB to familiarize the child with the tone. The intensity was reduced to 25dB when the child could respond successfully. Frequencies in the order of 1000 Hz, 2000 Hz, 4000 Hz, and then 500 Hz were tested at 25dB. For each frequency, the intensity was raised in 5dB steps if there was a failure of response, and was reduced in 10 dB steps if a correct response was given. The same procedures were used for the left ear. The lowest intensity of each frequency was recorded.

Zygosity testing

An Oragene DNA self-collection kit (model no. OG-100 vial format or OG-250 disc format) was used to collect saliva from each child. Zygosity of twin pairs was determined by SNP testing conducted by professionals at the Department of Biochemistry in the Chinese University of Hong Kong and the Genome Quebec Innovation Centre in the University of McGill.

Questionnaire

Parents were requested to fill in a questionnaire to provide information on participating twins’ demographic background, home literacy environment, communicative ability, and motivation for learning Chinese text. Items on twins’ demographic background included twins’ sex, date of birth, maternal and paternal education levels and family income. There were 18 items on their home literacy environment and 4 items on their motivation for learning Chinese text. The questionnaires also included 20 items adapted from Children’s Communication Checklist - Second Edition (Bishop, 2003) translated into Chinese by Wong (2007) to gather information on the child’s communicative ability. This questionnaire was printed in Chinese. The items of twins’ demographic background, home literacy environment and motivation for learning Chinese text, and their English translation, are presented in Appendix 1E.

Procedure

The sample was shared with another Doctor of Philosophy candidate of the Department of Experimental Psychology at the University of Oxford who was interested in studying Chinese twins learning English as a second language. The children were tested in two 1-hour sessions by trained psychology major undergraduates and graduates in their school, their home, or University of Hong Kong according to parents’ preference. Chinese measures were administered in one session (Session A), and English, nonverbal reasoning and hearing screening test were employed in another session (Session B). Children were given a ten-minute break in between sessions, unless otherwise was requested by the children or the parents. Data on Chinese and nonverbal reasoning measures were included in this thesis.

The measures and their order of administration included in this thesis are listed in Table 2.3. This order considered the task nature and children's attention. Tasks tapping similar skills and with similar testing procedures were scattered apart.

Table 2.3. Order of Task Administration.

Session A
Rapid automatized naming
Receptive vocabulary
Phonological memory
Orthographic skills
Syllable and rhyme awareness
Word reading
Morphological awareness
Tone awareness
Session B
Audiometric screening test
Colour Raven's Progressive Matrices

Two experimenters visited a pair of twins together. Each experimenter administered tasks of a particular session (i.e., either session A or session B) to both twins during each visit, thereby reducing the assessment errors given by two different persons assessing particular tasks on a pair of twins. The twins were seen in separate rooms, or in two corners of a big room if the former was not possible. Children were tested again on the same tasks, except Colour Raven's Progressive Matrices, around one year after the initial testing. Saliva was collected from cotwins with DNA kits at the end of the testing for zygosity assessment, and questionnaires were sent to and obtained from parents by post, in the first year of subjects' participation. The author tested 108 children in the first wave and 110 children in the second wave of testing.

Chapter 3

Chinese Language and Reading Skills: Genetic and environmental influences on their individual differences and overlap

Overview

This part of the thesis investigates the etiology of individual differences in Chinese language and reading skills in 312 typically developing Chinese twins aged from 3 to 11 years in Hong Kong. Children were given tasks of Chinese word reading, receptive vocabulary, phonological memory, tone awareness, syllable and rhyme awareness, rapid automatized naming, morphological awareness, and orthographic skills individually. All analyses controlled the effects of age and nonverbal reasoning. Results of genetic analyses indicated moderate to strong genetic influences on Chinese word reading, tone awareness, phonological memory, and rapid automatized naming, while shared environment played a more important role in Chinese receptive vocabulary, syllable and rhyme awareness, and orthographic skills. Morphological awareness showed significant influences of both genetic and shared environmental factors. Factor analyses on these diverse Chinese tasks indicated two factors: General Language and General Reading, and further genetic analyses were conducted on these two factors. Moderate genetic influences were found for both general language and reading abilities, but bivariate genetic analyses showed nonsignificant genetic or shared environmental links between these abilities. Findings of this thesis suggest the differential roles of genetic and environmental influences in various Chinese language and reading skills. When more general Chinese language and reading abilities are considered, heredity plays a more central role in determining their variations than shared environment, and there is a possibility that they have different etiology in development.

Research background

Children show great individual variations on language and reading skills (e.g., Bates et al., 1995). Individual variations in language and reading skills tend to inter-correlate, suggesting possible common etiology in the development of language and reading skills (e.g., Harlaar, Hayiou-Thomas, Dale, & Plomin, 2008). Both heredity and environment play a role in the origins of these individual differences, but their relative contributions can differ in various language and reading skills as demonstrated in twin studies (e.g., Byrne et al., 2005; Petrill et al., 2006; Samuelsson et al., 2005). To date, twin studies have been conducted on alphabetic languages, mainly on English. Thus, whether these roles of heredity and environment are universal or differ across languages remains unknown. Chinese has very different linguistic and orthographic characteristics compared with alphabetic scripts. So, the investigation of the genetic influences on Chinese learning will lead to enhanced understanding of, not only the nature of Chinese acquisition, but also the universal or specific factors of language and reading acquisition across languages. This part of the thesis investigates the genetic and environmental origins and overlap of Chinese language and reading skills in 312 Chinese twin pairs in Hong Kong.

Cognitive skills important to reading abilities

Language and reading development of children is interwoven (Snow et al., 1998). The relationships between language and reading skills vary across children’s reading stages.

Specifically, language skills link to early-stage reading abilities indirectly through cognitive skills, such as phonological awareness (Storch & Whitehurst, 2002), but they contribute to reading abilities beyond phonological skills in elementary grades (Nation, 2004). Strong positive links have been demonstrated between English reading abilities and phonological processing skills, including phonological awareness (e.g., Bradley & Bryant, 1983; Mann & Liberman, 1984; McBride-Chang & Kail, 2002; Torgesen, Morgan, & Davis, 1992), and rapid naming ability (e.g., Manis, Seidenberg, & Doi, 1999; Torgesen, Wagner, Rashotte, Burgess, & Hecht, 1997; Wagner et al.,1997). Among these phonological skills, phonemic awareness is a consistent and robust predictor of English reading (e.g., Hulme et al., 2002; Nation & Hulme, 1997). Phonological memory is correlated with English reading (e.g., Gathercole, Willis, & Baddeley, 1991; Mann & Liberman, 1984), but often does not contribute uniquely over phonological awareness and rapid naming in predicting English reading (e.g., Wagner et al., 1997).

While the roles of phonological skills have been clearly demonstrated, the importance of other cognitive skills has drawn attention in reading research. It has been underscored by Tunmer and Nesdale (1985), who noted that phonological awareness was necessary but not sufficient for reading acquisition. Converging research evidence has shown the prominence of morphological awareness and orthographic skills in English reading, making a unique contribution after phonological skills have been taken into account (e.g., Deacon & Kirby, 2004; Mahony, Singson, & Mann, 2000; Siegel, 2008 for morphological awareness; e.g., Badian, 2001; Barker, Torgesen, & Wagner, 1992; Cunningham, Perry, & Stanovich, 2001 for orthographic skills).

In contrast to the many studies on the cognitive features of English acquisition, relatively little is known about Chinese language and reading development. Because Chinese has very different characteristics compared with English, many of the research findings about English may not apply in Chinese. Chinese is a morphosyllabic language in which each character, the primary unit of writing, represents both a syllable and a lexical morpheme. A character and its components (strokes or radicals) do not represent phonemes. These one-to-one-to-one relations among Chinese characters, syllables, and morphemes promote a high degree of metalinguistic prominence pertaining to syllables (e.g., Chow et al., 2005) and morphemes (e.g., Chen, Hao, Geva, Zhu, & Shu, 2009), with some arguing that their role in Chinese acquisition is analogous to that of phonemic awareness in English acquisition (Nagy & Anderson, 1999).

Apart from the morphosyllabic nature of Chinese, the extensive lexical compounding and large number of homophonic morphemes further underscore the importance of morphological awareness in learning Chinese (e.g., McBride-Chang, Shu, Zhou, Wat, & Wagner, 2003). First, Chinese involves extensive lexical compounding and polymorphemic words are common. For example, the word 電視 (*/din6 si6/; television*) is composed of the morphemes 電 (*/din6/; electric*) and 視 (*/si6/; vision*). Hence, meanings of Chinese words can often be derived through morphological analysis (Ku & Anderson, 2003). Second, there are a large number of homophonic morphemes in Chinese. For example, in the Cantonese dialect, on average nearly three regularly used characters, each representing a distinct meaning, share one tonal syllable (Chow, McBride-Chang, Cheung, & Chow, 2008). So, the awareness on morphological construction or compounding (e.g., Chen et al., 2009) and homophone (e.g., McBride-Chang et al., 2003) plays a prominent role in learning Chinese.

Furthermore, learning of syllables is always associated with learning of tones in Chinese, because a Chinese syllable in different tones (six tones in Cantonese) represents different meanings and each of them is denoted by a different character. Although there is no symbol in a character that signifies tone, this awareness of tone is essential for distinguishing the meaning of a particular syllable which allows the correct identification of a character (Shu, Peng, & McBride-Chang, 2008). So, phonological awareness at the syllable and tone levels supports Chinese language and reading development.

However, the orthography-phonology mappings are ambiguous in Chinese, and this highlights the contribution of rapid automatized naming ability and phonological memory to Chinese reading skills. In Chinese, orthography-phonology correspondence refers to pronouncing characters with the cues provided by their phonetic radicals. This is relevant only to the semantic-phonetic compound characters, each of which consists of a phonetic radical and a semantic radical that provide cues to its pronunciation and meaning respectively. Although semantic-phonetic compound characters constitute around 80% of the Chinese characters, only around 40% of the Chinese characters can be successfully decoded by the orthography-phonology correspondence rules, because some phonetic radicals provide no useful information in reading the characters (Shu et al., 2003). Therefore, mastering the orthography-phonology correspondence rules is not sufficient for gaining proficiency in Chinese reading. Thus, learning of Chinese relies heavily on forming relatively arbitrary associations between print and language, especially when the need to master a huge number of characters (4500 regularly used character, Liu et al., 1975) is considered. These print-sound associations are also central elements of rapid automatized naming which reflects the speed of access to the lexicon. So, rapid automatized naming skills are linked to Chinese reading ability (e.g., McBride-Chang & Ho, 2000). Also, a great number of phonological units have to be stored in and retrieved from the lexicon during the process of learning Chinese, and thus phonological memory contributes to reading acquisition in Chinese (e.g., Hu & Catts, 1998). Regarding Chinese orthography, each Chinese character occupies a fixed square space in print, and its components and radicals have habitual positions. Knowledge of correct orientation of orthographic units aids the formation of character representations in the lexicon, and locating cues of pronunciation and meaning of characters. These orthographic skills are prominent in reading Chinese (e.g., Cheung, Chan, & Chong, 2007).

In general, past research has indicated that morphological awareness (e.g., Ku & Anderson, 2003), phonological awareness at the syllable and onset-rime levels (e.g., Chow et al., 2005; Siok & Fletcher, 2001), tone awareness (e.g., Shu et al., 2008), rapid automatized naming (e.g., McBride-Chang & Ho, 2000), phonological memory (e.g., Hu & Catts, 1998), and orthographic skills (e.g., Cheung et al., 2007) contribute to Chinese language and reading skills. Owing to the contrasting characteristics of Chinese and English, metalinguistic skills which are crucial in English reading, such as phonemic awareness, might not play such an important role in Chinese reading (e.g., Siok & Fletcher, 2001). Conversely, morphological awareness tends to be more prominent in Chinese reading than English reading (McBride-Chang et al., 2005). The causes of individual variations of these cognitive skills in Chinese are examined in this chapter.

Twin studies on language and reading skills

The twin design has been employed to study the acquisition of speaking and reading alphabetic languages, mainly English. Robust research evidence has indicated strong genetic and negligible to small environmental influences on reading skills (e.g., Byrne et al., 2009; Hart, Pertrill, Thompson, & Plomin, 2009), and rapid naming ability (e.g., Davis, Knopik, Olson, Wadsworth, & DeFries, 2001; Petrill et al., 2006). Moreover, converging research evidence has supported genetic factors, compared to environmental factors, exert relatively greater influences on phonological memory (e.g., Kovas et al., 2005; Wadsworth et al., 1995), phonological awareness (e.g., Byrne et al., 2005; Hohnen & Stevenson, 1999), and orthographic skills (Gayán & Olson, 2003). However, shared environmental influences are relatively greater on vocabulary knowledge (e.g., Byrne et al, 2002; Dionne, Dale, Boivin, & Plomin, 2003), and grammar/morphological skills (Samuelsson et al, 2005; Samuelsson et al, 2007). Results are mixed on general language abilities (stronger genetic influences: Haworth et al., 2009; stronger environmental influences: Harlaar et al., 2008; comparable influences: Colledge et al, 2002; Hayiou-Thomas et al., 2006). Key results on the aforementioned skills in normal variations are summarized in Table 3.1.

Table 3.1. Twin Studies Examining Word Reading and Vocabulary, and Related Cognitive Skills in Typically-developing Samples.

Authors	Country	Age in years (range/waves)	N twin pairs: MZ/same-sex DZ /opposite-sex DZ	Measures	Heritability estimate (95% CI)	Shared Env estimates (95% CI)
Word reading						
Byrne et al., 2005 [a]	Australia, USA	preschool	172/153	TOWRE	.70(.52,.93)	.22(.00,.40)
Byrne et al., 2007 [a]	Australia, USA	7.0 – 7.4 in Grade1	167/152	TOWRE	Grade1 .72(.49,.82) Kinder .49(.32,.69)	Grade1 .05(.00,.28) Kinder .28(.09,.44)
Byrne et al., 2009 [a]	Australia, USA, Scandinavia	7.9 – 8.7	303/312	TOWRE	.82(.67,.88)	.03(.00,.19)
Samuelsson et al., 2007 [a]	Australia, USA, Scandinavia	4.8 – 5.1	401/408	TOWRE Word recognition subtest	USA .61(.45,.81) Australia .91(.67,.94)	USA .33(.12,.49) Australia .00(.00,.24)
Samuelsson et al., 2008 [a]	Australia, USA, Scandinavia	6.0 – 6.7	402/410	TOWRE word and nonword subtests	Kinder .33 to .84 (all sig) Grade1 .79 to .83 (all sig)	Kinder .09 to .52 (sig in USA and Scandinavia) Grade1 .02 to .07 (all nonsig)
Brooks et al., 1990 [b]	USA	12;6 (7;8-20;6)	86/60	PIAT Reading recognition subtest	.45 [CA]	.07 [CA]
Gayàn & Olson, 2003[b]	USA	10.6 (8 - 18)	257/183	PIAT Word recognition subtest; Time-limited word recognition test	.85(.69,.92) [CA]	.04(.00,.19) [CA]
Keenan et al., 2006 [b]	USA	11 (8-17)	70/61/60	PIAT word recognition subtest; Timed Oral Reading of Single Words	.66(.42,.82) .65(.42,.81) [CA]	.11(.00,.32) .11(.00,.32) [CA]
Hart, Petrill, Thompson, & Plomin, 2009 [c]	USA	6 in wave1; (4 waves)	128/175	WRMT-R Word identification subtest	.45 to .94 (all sig)	.00 to .28 (all nonsig)
Petrill et al., 2006 [c]	USA	6.1	102/140	WRMT-R Word identification subtest	.68(.48,.91) .67(.44,.88) [CA]	.22(.00,.42) .18(.00,.40) [CA]
Harlaar, Dale, & Plomin, 2005 [d]	UK	7.06	1019/948	TOWRE (Same / Different teachers)	.63(.53,.74)/ .74(.59,.86)	.22(.11,.32)/ .09(.00,.24)
Vocabulary[#]						
Byrne et al., 2002 [a]	Australia, USA, Norway	4.9 – 5.1	125/125	WPPSI-R Vocabulary subtest; Hundred Pictures Naming Test	.18	.49(sig)
Byrne et al., 2009 [a]	Australia, USA, Scandinavia	7.9 – 8.7	303/312	Boston Naming Test	.44(.31,.59)	.36(.22,.49)
Samuelsson et al., 2005 [a]	Australia, USA, Scandinavia	4.8 - 5.1	312/315	WPPSI-R Vocabulary subtest; Hundred Pictures Naming Test	.32(.06,.56)	.60(.38,.81)

Hart, Petrill, DeThorne, et al., 2009[c]	USA	6.1 in wave1; (3 waves)	128/175	Boston Naming Test	.29 to .49 (all sig)	.37 to .46 (all sig)
Mather & Black, 1984[e]	USA	4.5	50/29	Peabody Picture Vocabulary Test	.68	--
Van Hulle et al., 2004	USA	(1;7-3.2)	116/125/112	MCDI–Short Form II	Girls .08(.05,.14) Boys .20(.12,.28)	Girls .77(.63,.95) Boys .53(.41,.69)
Dionne et al., 2003[d]	UK	2 and 3 (2 waves, 2 cohorts)	771/734 in 1994 cohort 502/547 in 1995 cohort	MCDI: United Kingdom Short Form Vocabulary scale (1994/1995))	.10 to .21(all sig) .14-.17 (all sig)	.75 to.84(all sig) .80-.82 (all sig)
Phonological memory						
Byrne et al., 2002[a]	Australia, USA, Norway	4.9 – 5.1	125/125	Children's Test of Nonword Repetition-Revised	.19(nonsig)	.22.(nonsig)
Samuelsson et al., 2005[a]	Australia, USA, Scandinavia	4.8 - 5.1	312/315	Nonword repetition; WPPSI-R Sentence memory subtest; WRAML Sound Symbol subtest	.57(.35,.79)	.29(.08,.48)
Wadsworth et al., 1995[b]	USA	12.3 (8-20)	258/188	Detroit Test of Learning Aptitude subtests: Auditory attention span for related syllables, Auditory attention span for unrelated words; WISC-R or WAIS-R Digit span subtest	.95	.02
Kovas, Hayiou-Thomas, et al., 2005[d]	UK	4;6	281/275/231	Children's Test of Nonword Repetition	.41(.18,.57)	.09(.00,.27)
Phonological awareness						
Bus & Out, 2009	Netherlands	4.6	27/39	Rhyme production	Nonsig	.73(.58,.82)
Byrne et al., 2002[a]	Australia, USA, Norway	4.9 – 5.1	125/125	CTOPP; Lonigan's task	.52(sig)	.16
Byrne et al., 2005[a]	Australia USA	preschool	173/153	CTOPP Elision, Blending, Sound matching subtests	.63(.36,.92)	.28(.00,.53)
Samuelsson et al., 2005[a]	Australia, USA, Scandinavia	4.8 - 5.1	312/315	Word blending; Syllable and phoneme blending; Sound matching; Word elision; Syllable and phoneme elision; Rhyme and final sound	.60(.37,.85)	.32(.08,.52)
Gayàn & Olson, 2003[b]	USA	10.6 (8-18)	257/183	Phoneme transposition; Phoneme deletion; Lindamood auditory conceptualization test	.83(.62,.94)[CA]	.08(.00,.27)[CA]
Petrill et al., 2006[c]	USA	6.1	102/140	6 subtests from Robertson	.48(.33,.68) .30(.09,.52)[CA]	.43(.23,.58) .50(.29,.66)[CA]
Petrill et al., 2007[c]	USA	6.1 in wave1 7.2 in wave2	119/164 in w1 88/123 in w2	Robertson and Salter's (1997) Phonological Awareness Test	Wave1 .59(.34,.78) Wave2 .14(.04,.39)	Wave1 .16..02,.40) Wave2 .47(.23,.65)

Hohnen & Stevenson, 1999	UK	5 – 7	66/60	Phoneme blending; Sound categorization; Age7only: Phoneme deletion	Age6 .52 Age7 .62	Age6 .35 Age7 .28
Kovas, Hayiou-Thomas, et al., 2005 [d]	UK	4;6	281/275/231	Authors devised phoneme detection task based on Bird, Bishop, & Freeman, 1995	.38(.13,.53)	.06(.00,.24)
Rapid automatized naming						
Byrne et al., 2002 [a]	Australia, USA, Norway	4.9 – 5.1	125/125	CTOPP subtest		
				Rapid naming objects	.00	.30(sig)
				Rapid naming colours	.66(sig)	.00
Byrne et al., 2005 [a]	Australia, USA	preschool	172/153	CTOPP Color, digit and letter Naming subtests	.60(.33,.82)	.17(.00,.41)
Samuelsson et al., 2005 [a]	Australia, USA, Scandinavia	4.8 - 5.1	312/315	CTOPP Object and colour naming subtests	.64(.40,.81)	.11(.00,.57)
Davis et al., 2001	USA	11.7 (7.9-20.4)	221/139	Rapid naming paradigm Numbers, colours , pictures and letters subtests (Denckla, & Rudel, 1976)	.62	.03
Hart, Petrill, Thompson, & Plomin, 2009 [c]	USA	6 in wave1; (4 waves)	128/175	CTOPP Letter and number naming subtests	.42 to .79(sig in first 3 waves)	.00 to .17(all nonsig)
Petrill et al., 2006 [c]	USA	6.1	102/140	CTOPP Letter and number naming subtests	.77(.49,.84) .72(.43,.83)^{CA}	.01(.00,.26) .04(.00,.29) ^{CA}
Morphology/Grammar						
Byrne et al., 2002 [a]	Australia, USA, Norway	4.9 – 5.1	125/125	Productive morphology task based on Berko (1958)	.31(nonsig)	.27(nonsig)
Samuelsson et al., 2005 [a]	Australia, USA, Scandinavia	4.8 - 5.1	312/315	Illinois Test of Psycholinguistic Abilities Grammatic Closure subtest; Productive morphology test based on Berko's(1958)	.29(.07,.53)	.59(.38,.77)
Mather & Black, 1984 [e]	USA	4.5	50/29	Berko Test of English Morphology	No sig differences between MZ and DZ correlations	
Orthographic coding						
Gayàn & Olson, 2003 [b]	USA	10.6 (8-18)	257/183	Orthographic choice; Homonym choice; PIAT Spelling subcomponent	.87(.75,.94) ^{CA}	.01(.00,.11) ^{CA}

Note. Studies based on the same sample are denoted by the same superscript in the Authors column. In case of studies examined the same variable based on the same sample, the study with the largest sample in the analyses is included. Studies with samples in the age range of 3 to 11 years are reviewed in this table. Superscript CA denotes estimates with general cognitive ability accounted for.

[#]Past children twin studies focused on expressive vocabulary. Expressive vocabulary and receptive vocabulary skills are highly correlated. The corrected correlations between the Expressive One-Word Picture Vocabulary Test and the Receptive One-Word Picture Vocabulary Test and the Peabody Picture vocabulary Test-Third Edition are .72 and .76 respectively (Brownell, 2000).

CTOPP - Comprehensive Test of Phonological Processing; MCDI - MacArthur Communicative Development Inventories; PIAT - Peabody Individual Achievement Test; TOWRE - Test of Word Reading Efficiency; WAIS-R – Wechsler Adult Reasoning Scale – Revised; WISC-R – Wechsler Reasoning Scale for Children – Revised; WPPSI-R - Wechsler Preschool and Primary Scale of Reasoning-Revised; WRMT-R - Woodcock Reading Mastery Tests-Revised

These genetic influences could overlap and constitute common genetic origins of language and reading skills. For instance, reading abilities have been shown to share common genetic influences with rapid naming ability (Davis et al., 2001), phonological memory (Wadsworth et al., 1995), phonological awareness (Petrill et al., 2006), and oral language skills (Haworth et al., 2009). Strong genetic overlap has also been found between language skills at 4.5 years old and reading abilities at 7 years old (Hayiou-Thomas et al., 2006), and reading performance across ages 7, 9 and 10 (Harlaar et al., 2007b). This line of research evidence has shed light on the underlying mechanisms of the links between language and reading skills.

However, it is important to know whether or not nonverbal ability has been taken into account when interpreting such results. As noted by Brooks, Fulker and DeFries (1990), most cognitive tests show a 'positive manifold' of inter-correlations. Plomin and Kovas (2005) more recently have stressed the extent to which different cognitive skills have common genetic influences from 'generalist genes'. The interest of this thesis, however, was in the extent to which different skills predict language and reading skills when nonspecific effects of general ability have been taken into account. Accordingly, the analyses this thesis shall present are based on scores corrected for nonverbal reasoning as well as age.

The growing twin research in alphabetic languages has contributed to better understanding on the roles of heredity and environment on the development of these languages, mainly English. However, there is a lack of twin study on languages which have very different linguistic characteristics compared with alphabetic ones, such as Chinese. This thesis investigates how heredity and environment determine the development of Chinese language and reading skills, and related cognitive skills, in 312 Chinese twin pairs.

Research questions

There are three research questions in this chapter. First, what are the phenotypic relationships between diverse facets of Chinese language and reading skills, including word reading, receptive vocabulary, phonological memory, tone awareness, syllable and rhyme awareness, rapid automatized naming, morphological awareness, and orthographic skills? These relationships are tested in two ways, a) do various language and reading related skills contribute to word reading and receptive vocabulary; and b) how do all these skills cluster to form more general factors, after the effects of age and nonverbal reasoning are accounted for. It is predicted that these various skills will each link to word reading and receptive vocabulary, and the way of which all these skills cluster will be explored.

Second, how does heredity and environment contribute to individual differences in these skills? It is hypothesized that strong genetic influences will be found for word reading and rapid automatized naming, because of the robust evidence from past twin research and also the less diverse environments for the development of these skills (i.e., reading skills are mainly acquired through formal instruction and explicit teaching on naming speed is not common). In addition, it is predicted that phonological memory and orthographic skills will be influenced more by genetic factors than environmental ones, based on past research findings. However, learning orthographic skills in Chinese could be relatively demanding than in English, as Chinese orthographic units are visually more complex and many of these units are unpronounceable. This challenge might yield a wider range of learning experiences, and so the genetic effects will be lower than those indicated in English twin studies. Furthermore, it is expected that environment will have stronger effects on receptive vocabulary and syllable and rhyme awareness. It is because receptive vocabulary is largely influenced by home environment and language interaction with others, and these experiences are diverse, highlighting the importance of the environmental effects. As phonological cues in Chinese are much more unreliable in reading than are those in English, learning experiences may exert greater influences on learning to read Chinese and on developing syllable and rhyme awareness, thus diminishing the effects of heredity on them in Chinese. Furthermore, past research suggested greater effects of environmental factors than genetic ones on morphological awareness. These environmental effects could be even stronger in Chinese, because morphological units play a particularly important role in learning to Chinese (e.g., morphosyllabic nature and extensive lexical compounding) and so the various environments where children experience and use the language could enhance their awareness of morphological units. Lastly, the genetic and environmental contributions to tone awareness are explored in this thesis, given the lack of existing relevant research findings.

Third, is there etiological overlap between general Chinese language and reading abilities? It is expected that their etiological overlap will be determined by genetic influences, given the common genetic origins of language and reading skills demonstrated.

Method

This thesis tested 312 Chinese twin pairs aged from 3 to 11 on tasks of Chinese word reading, receptive vocabulary, phonological memory, tone awareness, syllable and rhyme awareness, rapid automatized naming, morphological awareness, and orthographic skills, and Raven’s Colored Progressive Matrices (see Chapter 2 for participant and task details).

Results

Table 3.2 shows descriptive statistics of age and test scores by zygosity.

Table 3.2. Descriptive Statistics of Age and Test Scores by Zygosity.

	MZ		DZ	
	M	*S.D*	*M*	*S.D*
Age (years)	6.78	1.92	6.52	1.81
Raven's Colored Progressive Matrices (raw scores)	21.83	8.10	20.28	7.80
Nonverbal reasoning (age-adjusted Raven's Colored Progressive Matrices)	0.04	0.99	-0.10	1.02
Word recognition	85.63	63.57	74.39	61.18
Receptive vocabulary	53.13	16.85	51.67	16.87
Phonological memory	74.04	28.03	70.88	28.87
Tone awareness	7.31	3.19	6.94	2.69
Syllable and rhyme awareness	15.30	5.92	14.95	5.74
Rapid automatized naming	.06	.03	.06	.03
Morphological awareness	22.91	12.26	20.92	11.67
Orthographic skills	63.04	15.57	61.10	15.71

Note. *N*= 455 for MZ; *N*= 168 for DZ.

Standardized scores adjusted for age and nonverbal reasoning were employed in all analyses. A child had refused to complete the Raven's Colored Progressive Matrices, and so this twin pair was not included in genetic analyses, but the performance of this child's cotwin was included in phenotypic analyses. Results of independent sample T-tests on the scores of a cotwin randomly selected in each twin pair showed MZ twins and DZ twins did not significantly differ on their performance across tasks (*p*s>.05).

Phenotypic relationships

Bivariate correlation and hierarchical regression analyses were conducted on the scores of a cotwin randomly selected in each twin pair.

Correlations among language and reading measures. The phenotypic bivariate correlations of word reading with the other tasks controlling for age and nonverbal reasoning were all significant (r ranged from .13 to .33, ps<.05), except phonological memory (p>.05). Correlations of receptive vocabulary with the other tasks controlling for age and nonverbal reasoning were all significant (r ranged from .13 to .43, ps<.05), except tone awareness and rapid automatized naming (ps>.05). The phenotypic bivariate correlations across tasks with age and nonverbal reasoning accounted for are indicated in Table 3.3.

Table 3.3. Correlations Controlling for Age and Nonverbal Reasoning across Tasks

	Word reading	Receptive vocabulary	Phonological memory	Tone awareness	Syllable and rhyme awareness	Rapid automatized naming	Morphological awareness
Receptive vocabulary	.13*						
Phonological memory	.08	.29**					
Tone awareness	.18**	.01	.10				
Syllable and rhyme awareness	.23**	.32**	.34**	.20**			
Rapid automatized naming	.31**	.04	.05	.04	.26**		
Morphological awareness	.28**	.43**	.33**	.23**	.35**	.17**	
Orthographic skills	.33**	.22**	.09	.05	.28**	.23**	.27**

Note. **p*<.05; ***p* < .01. *N* ranged from 302 to 311.

Unique contributions to word reading and receptive vocabulary. To understand the unique contributions of these skills to word reading, a hierarchical regression explaining variance in word reading with age and nonverbal reasoning controlled was conducted. Phonological memory was not entered as a predictor as it was not significantly correlated with word reading. The variables were entered in the order of receptive vocabulary, tone awareness, syllable and rhyme awareness, rapid automatized naming, morphological awareness and finally orthographic skills. Receptive vocabulary was placed as the first variable in the hierarchical regression as it represents more general language ability. Tone awareness was entered next to understand its relation with word reading over general language ability. Syllable and rhyme awareness and rapid automatized naming was entered as the third and the fourth variables because they are more specific language skills which have strong links with word reading ability (e.g., Hu & Catts, 1998; McBride-Chang & Ho, 2000). Morphological awareness was then placed fifth as its unique contribution to word reading over and above vocabulary knowledge, phonological awareness and rapid automatized naming was demonstrated (e.g., McBride-Chang et al., 2003). Orthographic skills were placed in the final step to understand its relationship with word reading controlling for verbal and rapid naming tasks administered. These seven variables explained 22% of variance in word reading, after the effects of age and nonverbal reasoning were controlled. The entry of each of them yielded a significant R^2 change, demonstrating the unique contributions of these skills to word reading ability. When all independent variables were entered in one step, tone awareness, rapid automatized naming, morphological awareness and orthographic skills significantly predicted word reading (see Table 3.4).

Table 3.4. Hierarchical Regressions Explaining Variance in Word Reading Controlling for Age and Nonverbal Reasoning.

Steps	Variables	*B*	*t*	R^2	R^2 change
1	Receptive vocabulary	.01	0.13	.02	.02*
2	Tone awareness	.13	2.25*	.05	.03**
3	Syllable and rhyme awareness	.03	0.44	.08	.03**
4	Rapid automatized naming	.23	4.11**	.15	.07**
5	Morphological awareness	.13	2.11*	.17	.02**
6	Orthographic skills	.26	4.19**	.22	.05**

Note. *p<.05; **p < .01. N= 299.

A comparable hierarchical regression explaining variance in receptive vocabulary with age and nonverbal reasoning controlled was also conducted. Variables with significant correlations with receptive vocabulary were entered. These included word reading, phonological memory, syllable and rhyme awareness, morphological awareness, and orthographic skills. The order of variable entry corresponded to that of word reading, except word reading replaced receptive vocabulary as an independent variable in the first step and phonological memory was entered after that. These five variables explained 23% of variance in receptive vocabulary, after the effects of age and nonverbal reasoning were controlled. The entry of each of them yielded a significant R^2 change, except orthographic skills, demonstrating the unique contributions of these skills to receptive vocabulary knowledge. When all independent variables were entered in one step, syllable and rhyme awareness and morphological awareness significantly predicted receptive vocabulary (see table 3.5).

Table 3.5. Hierarchical Regressions Explaining Variance in Receptive Vocabulary Controlling for Age and Nonverbal Reasoning.

Steps	Variables	*B*	*t*	R^2	R^2 change
1	Word reading	-.03	-.56	.02	.02*
2	Phonological memory	.10	1.80	.08	.06**
3	Syllable and rhyme awareness	.16	2.74**	.13	.05**
4	Morphological awareness	.32	5.53**	.22	.09**
5	Orthographic skills	.09	1.46	.23	.01

Note. *p<.05; **p < .01. N= 306.

Factor structures of language and reading measures. An exploratory factor analysis was employed to explore the dimensionality of the eight variables. Factor scores of any factors indicated would be applied to both twins in subsequent genetic analyses, and thus this factor analysis included all twins. Two factors with eigenvalues over one were indicated. Direct Oblimin rotation showed items clustered on two factors. The factor loadings are shown in Table 3.6.

Table 3.6. Rotated Factor Loadings of Exploratory Factor Analysis with Direct Oblimin Rotation on Scores Controlling for Age and Nonverbal Reasoning.

Variables	General Language	General Reading
Word reading	.04	**.74**
Receptive vocabulary	**.69**	-.04
Phonological memory	**.67**	-.09
Tone awareness	**.50**	-.06
Syllable and rhyme awareness	**.56**	.34
Rapid automatized naming	-.15	**.74**
Morphological awareness	**.70**	.18
Orthographic skills	.12	**.61**

Note. Factor loadings over .40 are bolded. N= 594.

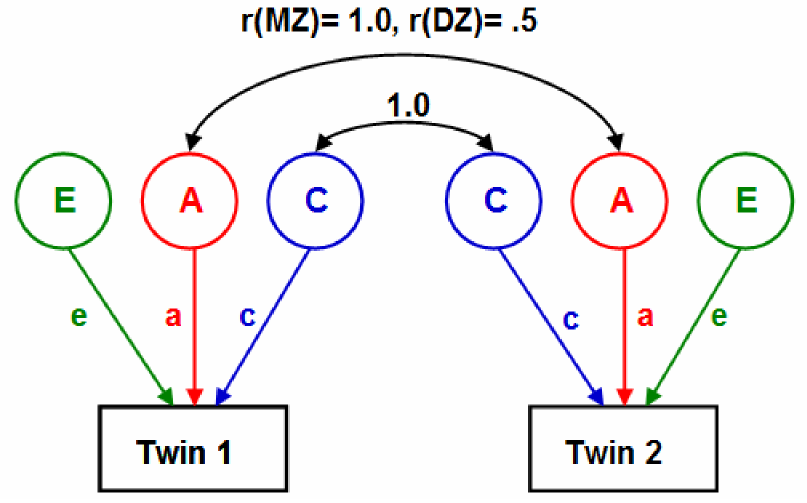
Receptive vocabulary, phonological memory, tone awareness, syllable and rhyme awareness, and morphological awareness were moderately to highly loaded on the first factor, which was referred as General Language. Word reading, rapid automatized naming and orthographic skills were moderately to highly loaded on another factor, which represented General Reading. The correlation between these two factors was .24 which was significant ($p<.01$).

Univariate genetic analyses

To reduce the effects of extreme scores on model fitting, standardized scores that exceeded absolute score of 3 were censored, by replacing them with 3 or -3 as appropriate. Also, to minimize the possible effects of different performance between MZ and DZ twins on model fitting, all scores were then standardized by zygosity. These adjusted scores were first fitted to the full ACE model, and then nested AE or CE models, for each variable by the Mx statistical modeling package (Neale, Boker, Xie, & Maes, 2006). The correlations of cotwins' scores on different tasks controlled for age and nonverbal reasoning by zygosity are shown in Table 3.7.

ACE models. The univariate ACE model was employed to examine the genetic and environmental contributions to each variable. A, C and E represent additive genetic, shared environmental and nonshared environmental factors, and their contributions are indicated by the paths a, c and e respectively (see Figure 3.1).

Figure 3.1. Univariate ACE model



Results indicated that all models provided a satisfactory goodness-of-fit (*p*> .05). Model parameter estimates are indicated in Table 3.7.

Table 3.7. Twin Correlations by Zygosity and Genetic Model Parameter Estimates of Each Variable Controlling for Age and Nonverbal Reasoning. (95% Confidence Intervals in Parentheses)

	Twin Correlations		ACE Models						
Variables	MZ	DZ	a^2	c^2	e^2	χ^2 (df=3)	*p*	AIC	RMSEA
Word reading	.85	.52	.68	.17	.15	0.72	.88	-5.29	.00
			(.43, .88)	(.00, .42)	(.12, .18)				
Receptive vocabulary	.54	.50	.09	.45	.46	1.25	.74	-4.75	.00
			(.00, .46)	(.10, .60)	(.38, .55)				
Phonological memory	.71	.46	.50	.21	.29	3.35	.34	-2.66	.03
			(.21, .76)	(.00, .49)	(.23, .36)				
Tone awareness	.51	.22	.50	.00	.50	1.35	.72	-4.65	.00
			(.16, .59)	(.00, .31)	(.41, .60)				
Syllable and rhyme awareness	.57	.53	.08	.49	.43	0.86	.84	-5.15	.00
			(.00, .44)	(.15, .63)	(.35, .52)				
Rapid automatized naming	.58	.37	.44	.14	.42	0.08	.99	-5.92	.00
			(.08, .66)	(.00, .47)	(.34, .51)				
Morphological awareness	.48	.36	.25	.23	.52	1.35	.72	-4.65	.00
			(.00, .56)	(.00, .51)	(.43, .62)				
Orthographic skills	.45	.43	.06	.39	.55	.34	.95	-5.66	.00
			(.00, .47)	(.01, .53)	(.45, .65)				
General Language	.65	.49	.33	.32	.35	.16	.98	-5.84	.00
			(.04, .69)	(.00, .59)	(.28, .43)				
General Reading	.70	.48	.45	.25	.30	.50	.92	-5.50	.00
			(.16, .75)	(.00, .52)	(.24, .37)				

Note. Number of twin pairs ranged from 210 to 227 for MZ and from 80 to 84 for DZ. a^2 = additive genetic variance; c^2 = shared environment variance; e^2 = nonshared environment variance; AIC = Akaike's information criterion; RMSEA = root mean square error of approximation.

Strong and significant genetic influences were found on word reading, phonological memory, and tone awareness, with two-thirds of the variance in word reading and half of the variances in phonological memory and tone awareness explained by genetic influences. Genetic factors also exerted moderate and significant influences on rapid automatized naming. In contrast, relatively strong and significant influences of shared environment were indicated in receptive vocabulary, syllable and rhyme awareness and orthographic skills, in which near half of the variance in syllable and rhyme awareness and around two-fifths of variances in receptive vocabulary and orthographic skills were explained by shared environmental factors. Nonsignificant genetic and shared environmental contribution was indicated on morphological awareness.

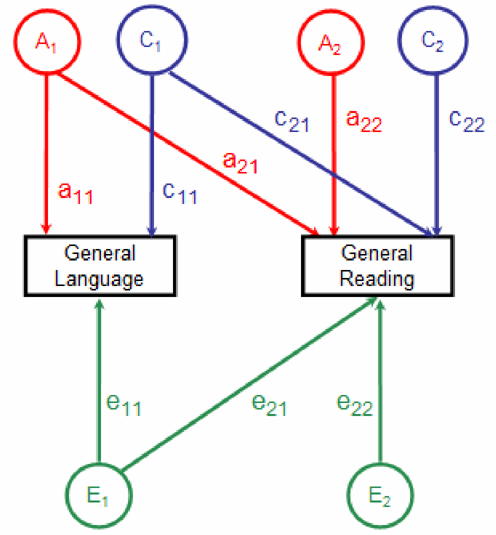
Genetic influences were significant and relatively stronger than those of shared environment on General Language and General Reading skills. Over one-thirds of their variances were contributed by genetic effects. Effects of nonshared environment, which included measurement errors, were moderate across tasks, except word reading. Table 3.7 shows the estimates of genetic and environmental influences on scores controlling for age and nonverbal reasoning.

Nested models. The nonsignificant parameter (either A or C) of the ACE models was dropped and the model fit of the nested models (AC or AE models) was compared to that of the full ACE models to further clarify the roles of genetics and environment in the aforementioned skills. Nested models with the C parameter dropped resulted in a nonsignificant decrease in model fit for word reading, phonological memory, tone awareness, rapid automatized naming, General Language, and General Reading. Also, nested models with the A parameter removed did not reduce model fit significantly for receptive vocabulary, syllable and rhyme awareness, and orthographic skills. These results further confirm the full ACE model results that heredity plays a more important role in word reading and rapid automatized naming, and shared environment has stronger effects than genetic influences on receptive vocabulary and syllable and rhyme awareness. For morphological awareness, the AE and the CE models did not have poorer model fit than the full ACE model, but the model fit decreased substantially when the A parameter or the C parameter was dropped (the E model) in the AE and the CE models. These results suggested that both genetic and shared environmental influences could contribute comparably to morphological awareness. The model fitting statistics are shown in Appendix 2A.

Bivariate genetic analyses

Genetic Cholesky decomposition model and genetic correlations. The Cholesky decomposition model was employed to investigate the genetic and environmental links between General Language and General Reading abilities (see Figure 3.2).

Figure 3.2. Cholesky decomposition model on General Language and General Reading.



Two sets of ACE terms were specified in the model. The first set (A1, C1 and E1) linked to both General Language and General Reading, and the second set (A2, C2, and E2) contributed to General Reading only. This model yielded a good fit ($\chi^2_{(11)}$ = 4.00, p=.97). The genetic and shared environmental influences on General Language (A1 and C1) were significant, but they did not significantly link to General Reading. However, General Reading was affected by specific genetic and shared environmental factors (A2 and C2) which were independent from those on General Language. All coefficients of nonshared environmental paths, which included measurement errors, were significant. Consistent with these results, nonshared environmental correlation was significant, but not the genetic and shared environmental correlations [r_A = .23 (-.48, .81); r_C = .37 (-1.00, .1.00); r_E = .15 (.01, .27)]. To further examine whether controlling for nonverbal reasoning effects yielded these nonsignificant links, the same model was tested with scores of General Language and General Reading accounting for age only. Results showed nonsignificant genetic and shared environmental links. The path coefficients and model fitting statistics from the Cholesky decomposition models are shown in Table 3.8.

Table 3.8. Standardized Path Coefficients from a Cholesky Decomposition Model of Genetic (A); Shared-Environment (C); and Nonshared-Environment (E) Influences on General Language and General Reading Factors Controlling for Age and nonverbal ability, and Age only. (95% Confidence Intervals in Parentheses)

	Age and IQ controlled		Age controlled	
	Paths		Paths	
Factors	A1	A2	A1	A2
1. General Language	.57		.56	
	(.29, .86)		(.35, .76)	
2. General Reading	.15	.65	.31	.62
	(-.25, .55)	(.41, .89)	(-.03, .65)	(.35, .89)
	C1	C2	C1	C2
1. General Language	.57		.70	
	(.29, .85)		(52, .88)	
2. General Reading	.19	.47	.22	.47
	(-.20, .57)	(.15, .79)	(-.05, .50)	(.15, .78)
	E1	E2	E1	E2
1. General Language	.59		.45	
	(.53, .65)		(.41, .49)	
2. General Reading	.08	.54	-.07	.51
	(.01, .15)	(.49, .59)	(-.14, -.002)	(.46, .55)

Note. Model fitting statistics of the model controlling for age and nonverbal reasoning: $\chi^2_{(11)} = 4.00$, p=.97, AIC= -18.00, RMSEA=.00; the model controlling for age: $\chi^2_{(11)} = 6.82$, p=.81, AIC= -15.18, RMSEA=.00; Confidence intervals were computed from standard errors generated by Mx.

Discussion

This part of the thesis investigated the etiology of individual differences in Chinese language and reading skills in 312 Chinese twin pairs. There are three notable findings in this part of the thesis. First, various facets of language and reading related skills contributed uniquely to Chinese word reading, except phonological memory. Morphological awareness played an important role in both Chinese word reading and receptive vocabulary. Also, the eight tasks tapping different Chinese skills yielded two factors: General Language and General Reading. Second, differential roles of genetic and environmental influences on various Chinese language and reading skills were indicated, which were partly consistent with findings on alphabetic languages. Lastly, bivariate genetic analyses showed nonsignificant genetic or shared environmental links between general Chinese language and reading abilities, which were inconsistent with findings on alphabetic languages.

Phenotypic relationships

Phenotypic analyses showed all tasks had unique contributions to the variance of word reading, except phonological memory. Consistent with past studies, morphological awareness contributed uniquely to word reading over and above age, nonverbal ability, vocabulary knowledge and phonological processing skills in Chinese (e.g. Chen et al., 2009; McBride-Chang et al., 2003). Our results have extended these findings by showing even after controlling for tone awareness in addition to these skills, morphological awareness still significantly contributed to the variance in Chinese word reading. It has strengthened the evidence on the strong link between morphological awareness and word reading, which is unique from that of phonological skills. When all skills were considered in one step, tone awareness, rapid automatized naming, morphological awareness and orthographic skills had relatively stronger links with word reading than the other skills. Therefore, while phonological awareness is a robust predictor of English word reading (Wagner & Torgesen, 1987), its role in Chinese word reading is reduced (except the suprasegmental tone units). The prominent role of morphological awareness was also found in our analyses on Chinese vocabulary knowledge. Specifically, morphological awareness contributed uniquely to receptive vocabulary knowledge over and above age, nonverbal ability, word reading and phonological processing skills, which is in line with past studies (e.g. McBride-Chang et al., 2006).

In addition, how these eight diverse Chinese tasks represent more general Chinese abilities was indicated. Results showed that these tasks were represented by two dimensions, namely General Language and General Reading. General Language reflected receptive and expressive verbal skills, while General Reading reflected skills of print processing and speeded naming.

Relative genetic and environmental contributions

This thesis indicated differential roles of genetic and environmental influences on various Chinese language and reading skills. Specifically, genetic factors exerted moderate to substantial influences on Chinese word reading, tone awareness, phonological memory, and rapid automatized naming, while shared environment played a more important role in Chinese receptive vocabulary, syllable and rhyme awareness, and orthographic skills, when effects of age and nonverbal reasoning were controlled. Heredity played a more central role in determining variation in both general Chinese language and reading abilities. These findings were consistent to the hypotheses, except environmental effects on orthographic skills were greater than expected.

The stronger influences of genetic factors on Chinese word reading, phonological memory, and rapid automatized naming, and the greater effects of shared environmental factors on Chinese receptive vocabulary, agree with past research evidence on English-speaking children (e.g., Byrne et al, 2009; Gayàn & Olson, 2003; Kovas et al., 2005). They suggest these genetic or environmental contributions are universal across languages, even for those with very different characteristics.

Findings of this thesis also suggest more language-specific influences. Results indicated genetic factors exerted greater influences on Chinese tone awareness. This is a novel and language-specific finding because tones (six tones in Cantonese) represent lexical meanings in Chinese, and it is a unique characteristic which English does not possess. In addition, Chinese phonological awareness at syllable and rhyme levels and orthographic skills were more affected by shared environments than by genes. Past studies have highlighted the prominence of genetic influences on English phonological awareness in kindergarteners (e.g., Byrne et al., 2005), junior elementary school children (Hohnen & Stevenson, 1999), and older children (e.g., Gayán & Olson, 2003). These converging links did not extend to Chinese phonological awareness. Chinese has ambiguous print-sound correspondence, and so the importance of phonological awareness is diminished. While phonics is widely used as an effective method in teaching English, which provides a comparatively more standard learning environment on phonological skills, its application in Chinese is not as effective and direct as that in English, owing to the ambiguous print-sound correspondence in Chinese. So, phonological awareness development may have heavier reliance on how far this skill is specifically taught. Phonological scripts can be employed in learning Chinese, but there is geographical diversity in use of the different methods: Pinyin in mainland China, Zhu Yin Fu Hao in Taiwan, and lack of systematic teaching on Cantonese phonetics in Hong Kong. These may explain why shared environmental effects on phonological awareness are stronger in Chinese than those in English. Regarding orthographic skills, the basic orthographic components in Chinese are unpronounceable strokes which combine to form orthographic units, such as radicals. These orthographic units are visually more complex, and more numerous than letters in English, and a number of them are unpronounceable. Also, they orient in various positions within a square frame, unlike letters which always arrange from left to right in English letter strings, and thus are spatially more complicated. To acquire the knowledge on legal stroke patterns and orthographic units as well as their orientations, Chinese children rely on various methods, such as copying and writing (Siok & Fletcher, 2001), which may contribute to relatively stronger environmental influences on orthographic skills.

Furthermore, results of this thesis suggest both genetic and shared environmental influences may be important in Chinese morphological awareness. However, more than one model gave satisfactory fit and so these results need further research. General language and general reading abilities in Chinese were both influenced more by genetic factors than by environments, which coincided with findings from English twin studies (Haworth et al., 2009). Therefore, in a broader perspective, individual differences in general Chinese learning abilities were more determined by genetic factors than shared environments.

Genetic and environmental links between language and reading skills

Bivariate genetic analyses showed nonsignificant genetic and shared environmental links between these general Chinese language and reading abilities. They are in marked contrast to findings of past English twin studies which indicated substantial genetic correlations between English language and reading skills, ranging from 0.67 to as high as 1.00 (Plomin & Kovas, 2005). These findings of past research fit with the generalist genes hypothesis (Plomin & Kovas, 2005) and is further substantiated by genetic linkage studies which identified language-associated genes also played a role in reading phenotypes, such as KIAA0319 (Rice, Smith & Gayán, 2009). One possible reason for the discrepancy with results of this thesis is that most of these studies did not control for nonverbal reasoning, which is known to have shared etiology with both language and reading ability. An exception is the study of Hohnen and Stevenson (1999), but they still found genetic overlap between language and literacy skills after removing the effect of nonverbal ability. Hayiou-Thomas, Harlaar, Dale and Plomin (2010) also reported that findings of significant genetic overlap between language and later literacy skills were unchanged when nonverbal ability was controlled for. It therefore seems that the explanation for the differences between results of this thesis and those previously reported might have to do with different facets of skills tested in this thesis and in past research on English. Measures which are important to one language might be less relevant in another. For instance, tone awareness was tested in our study as Chinese is a tonal language, while studies on English skills did not include it. So, the inclusion of different facets could be essential in understanding the development of various languages.

The nonsignificant shared environmental links reflect the shared environments associated to Chinese language abilities might not influence Chinese reading skills. This echoes with past research findings on the specific impacts of home and classroom literacy experiences on either language or literacy skills in Chinese. For instance, a recent training study has found shared book reading enhanced Chinese language skills, whereas parents' explicit metalinguistic training better prepared children for learning to read Chinese, in Hong Kong children (Chow et al., 2008). Also, only formal literacy experiences, such as instruction on reading and writing, were linked to Chinese literacy abilities, but informal literacy experiences, such as storybook reading, did not (Li, Corrie, & Wong, 2008). These research findings are also consistent with studies on children learning English as a mother tongue (e.g., Sénéchal, 2006), and have further supported the Home Literacy Model (Sénéchal & LeFevre, 2002), which proposed various home literacy experiences linked to language and literacy abilities differently. They inform parents and educators that various activities could have differential and specific effects on language or reading growth, and the design of activities should match the child's own needs and abilities.

Significant nonshared environmental link between language and reading abilities was demonstrated. However, this might reflect the impacts of the testing procedures associated with differential performances between cotwins across tasks, such as order of tasks administered. So, further studies are warranted to confirm these results on the nonshared environmental link. It should be noted that these findings on genetic, shared environmental and nonshared environment influences suggest several possibilities of the links between Chinese language and reading skills. First, Chinese language and reading skills have different etiology. Second, they have different genetic and shared environmental origins, but have common nonshared environmental factors. Third, potential overlap was not detected because of the relatively small sample size and large confidence intervals in this thesis. Therefore, further studies are warranted to confirm the relationship between Chinese language and reading skills.

Limitations

Findings in this chapter should be interpreted in the context of two caveats. First, children of a relatively wide age range were included. This age range covered children in different stages of language and reading development. Therefore, it should be noted that this thesis indicated the general patterns of children's individual variations in Chinese language and reading skills, instead of those in a particular developmental stage. Second, the sample size might not be large enough to detect the significant effects of some factors, especially those in the bivariate genetics analyses. So, cautious interpretations of these results should be taken, and further studies are warranted to confirm these results.

Implications and conclusions

To conclude, this thesis has extended past twin research on language and reading abilities in alphabetic languages to Chinese. There are several important implications. First, this thesis has indicated the differential roles of genetic and environmental influences on, and the lack of genetic and shared environmental links between, Chinese language and reading abilities. These findings are useful in constructing a more comprehensive Chinese acquisition model, and facilitating future investigation on the root causes of Chinese language and reading difficulties. For instance, models should take special account of specific factors contributing to the development of different skills. These findings also provide insights to educators and parents on structuring effective Chinese language and literacy learning environment to the children. Second, by comparing findings of this thesis with those of past studies on English abilities, the universal or specific factors of language and reading acquisition across languages could be understood. Specifically, the genetic contributions to word reading, phonological memory, and rapid automatized naming, and the shared environmental influences on receptive vocabulary, are likely to be universal across languages. However, general language and reading abilities in Chinese did not have common genetic and shared environmental etiology, which is unique to Chinese, though these findings need replications. Lastly, genetic factors played a prominent role in certain Chinese skills, including word reading, tone awareness, phonological memory, and rapid automatized naming. These are helpful for future molecular genetics studies, especially in selecting target language or reading skills for investigation in the Chinese populations.

Chapter 4

Home environmental influences on children’s language and reading skills in a genetically sensitive design: Socioeconomic status and home literacy environment are plausible environmental mediators but not moderators

Overview

This part of the thesis examines how family socioeconomic status (SES) and home literacy environment (HLE) contribute to Chinese language and reading skills in a genetically sensitive design. Twin pairs with SES and/or HLE information provided by their parents were included (296 and 215 twin pairs for SES and HLE respectively). The roles of SES and HLE in general language and reading abilities were examined in two ways, including a) mediation of shared environmental influences and b) moderation of genetic influences. Results showed that SES and HLE (Reading resources and opportunities at home) mediated the shared environmental contributions to general language and reading abilities, but did not moderate the genetic influences. Also, this thesis aims to investigate whether SES and HLE are plausible environmental mediators of more specific skills which are influenced by shared environments, including receptive vocabulary, syllable and rhyme awareness and orthographic skills. Results showed that SES and HLE (Reading resources and opportunities at home) mediated the shared environmental contributions to receptive vocabulary and syllable and rhyme awareness, but not orthographic skills.

Findings have indicated that SES and HLE are plausible mediators of shared environmental influences rather than moderators of genetic contributions. They have added to past research on alphabetic languages that SES and HLE mediated the shared environmental influences on language and reading skills, suggesting these links could be universal across languages. Also, the significant contributions of SES and HLE have extended to phonological awareness at the syllable and rhyme levels when genetic influences were accounted for, and their contributions might differ across reading-related cognitive skills.

Research background

Home environment plays an important role in children's language and reading development (Griffin & Morrison, 1997). Ample research evidence has demonstrated the links between children's language and reading skills with various home environmental factors, particularly family socioeconomic status (SES; e.g., Hecht, Burgess, Torgesen, Wagner & Rashotte, 2000) and home literacy environment (HLE; e.g., Payne, Whitehurst, & Angell, 1994). A few twin studies have further examined the mechanisms of these links, and have shown home environment, including SES, home chaos, and parental education, mediated the environmental influences (e.g., Petrill et al., 2004) and moderated the genetic effects (e.g., Friend, DeFries, & Olson, 2008) on language and reading skills. However, studies of this kind have been scarce and mixed results have been demonstrated, and so solid conclusions have not yet been made. This thesis aims to provide a more complete picture of the roles of SES and HLE by examining whether they mediate the environmental influences and/or moderate the genetic effects on Chinese language and reading skills in a single study with 312 Chinese twin pairs aged 3 to 11 years.

Socioeconomic status and home literacy environment

The impacts of family SES and HLE on children's language and literacy development have been well documented in psychological and educational research. Family SES reflects the resources and assets that the household possesses, which is not limited to material resources but also includes nonmaterial ones, such as education opportunities and social networks (Bradley & Corwyn, 2002). Despite its relatively broad nature in conceptualization, family SES has been widely tapped by household income, and parental education level and occupational status in developmental psychology studies (Bradley & Corwyn, 2002). Children from a high SES family often perform better in reading and related cognitive skills, including phonological awareness (e.g., Bowey, 1995; Raz & Bryant, 1990), and significant and persisting effects of SES on reading skills across time have been shown in longitudinal studies (e.g., Molfese, Modglin, & Molfese, 2003). For instance, Hecht et al. (2000) indicated SES differences in word decoding in each year from kindergarten to grade four when prior word decoding skills were controlled. Also, they found family SES in kindergarten explained 7% of variances in fourth-grade word decoding skills after prior decoding skills, general verbal intelligence, phonological awareness, rate of access, and print knowledge were accounted for. These influences of family SES have not been confined to Western cultures. Research has found family SES represented by maternal education was significantly linked to children's vocabulary knowledge in the Chinese culture (e.g., Lau & McBride-Chang, 2005). In general, SES is considered as a marker variable which covaries with other conditions which influence children's language and literacy learning. For example, higher SES is linked to more enriched language and literacy experiences (e.g., Hoff-Ginsberg, 1998). Compared to SES, HLE is a more proximal construct in linking with language and reading growth. However, HLE is not necessarily more predictive of language and reading skills than SES, as their relative contributions depend on how these measures are conceptualized and the samples included (Scarborough & Dobrich, 1994). Therefore, it is important to examine the influences of both SES and HLE under the same context.

HLE has been conceptualized in different ways, but there is a consensus that a multifaceted HLE could more adequately describe its relations with language and literacy development, as different HLE aspects may vary in their impacts on different outcomes (Sénéchal & LeFevre, 2002). Burgess, Hecht and Lonigan (2002) identified six interrelated facets of HLE, including Limiting Environment (parents' ability and disposition to provide literacy opportunities), Literacy Interface (parents' participation in children's literacy activities and their attitudes on literacy), Passive HLE (parents' activities which provide models of literacy usage or leisure activities), Active HLE (parents' direct engagement in children's literacy activities), Shared Reading (parents read with children), and Overall HLE (unit-weighted composite of the five facets above). They found that these HLE facets predicted significant portions of variances in subsequent word decoding and oral language, except Limiting Environment and Passive HLE, in a one-year longitudinal study. Likewise, other concurrent and longitudinal research further underscored the multifaceted nature of HLE and the differential importance of various facets in predicting language and reading skills (e.g., Umek, Podlesek, & Fekonja, 2005; Weigel, Martin, & Bennett, 2006), although they differed in the way of which HLE domains were classified. Similar findings were obtained in Chinese families as well, in which children's reading ability was linked to direct literacy teaching at home, but not the availability of home reading resources (Li et al., 2008). Because of the complicated nature of the Chinese script (i.e., ambiguous print-sound mapping and complicated visual features), researchers have pointed out that parental input in literacy teaching could be more important in learning Chinese than in learning alphabetic languages (Li & Rao, 2000). Across the various HLE domains, the importance of shared book reading has consistently been highlighted, especially for language development (e.g., Payne et al., 1994). Shared book reading explained on average 8% of the variances of various language and reading outcomes (Bus, van Ijzendoorn, & Pelligrini, 1995), and it enhanced children's language and literacy growth as documented in experimental and longitudinal research (e.g., Whitehurst et al., 1988; Wood, 2002). Similarly, studies focused on the Chinese families demonstrated the effectiveness of shared book reading in enhancing children's vocabulary knowledge (e.g., Chow & McBride-Chang, 2003).

Overall, these studies have identified SES and HLE as prominent environmental factors at the family level which contribute to children's language and reading growth. However, they have not considered the possible genetic confounds. Children in higher SES families are likely to be genetically predisposed to better learning abilities, and thus the apparent effects of SES could be confounded by the genetic contributions. Without the consideration of the possible genetic influences, the impacts of SES and HLE considered as home environmental factors might be inflated. Therefore, to better understand the impacts of SES and HLE, a genetically sensitive design, such as the twin study, which allows the estimation of environmental contribution with genetic effects taken into account, is essential.

Mediation and moderation effects

The twin study method provides a useful means to examine two types of SES and HLE effects with genetic effects considered, and they are a) mediation of shared environmental influences and b) moderation of genetic impacts. In general, a mediator is a variable which accounts for the link between an independent and a dependent variables, and thus reflects the process of which this link occurs, while a moderator is a variable which influences the direction and/or strength of the link, and thus specifies the situations where the link occurs (Baron & Kenny, 1986).

First of all, the twin study method estimates the genetic, shared environmental and nonshared environmental contributions to the variance of a phenotype. Within this context, family SES and HLE are environmental factors which are shared between the cotwins of a twin pair, and thus are conceptualized as shared environmental factors. Therefore, if SES or HLE influences are present, they will account for a significant portion of the shared environment contribution, and in other words they will show a mediation effect on the shared environmental influences. A handful of twin studies have examined this mediating role of family SES and HLE on language skills. In one of these studies, Petrill et al. (2004) indicated family SES and chaos each explained 3 to 4% of the total variances in parent-rated expressive vocabulary and grammatical complexity, and 5 to 7% of the shared environmental variances, in children assessed at age 3 and a year later, when SES and chaos were fitted into separate univariate genetic models. They also showed that when both SES and chaos were considered together, they accounted for 6% of the total variances in parent-rated expressive vocabulary and grammatical complexity and 10% of the shared environmental variances. SES was tapped by an index computed from parents' educational level and occupational status, and the maternal age at birth of the eldest child in this study. In another study, Hart, Petrill, DeThorne, et al. (2009) found that HLE, tapped by the parental report of the parents' and child's reading behaviours at home, explained 6 to 10% of the total variance in expressive vocabulary and mediated 13 to 27% of the shared environmental variances, in a group of children assessed longitudinally in ages 6, 7 and 8.

On the other hand, twin study can indicate whether SES or HLE moderate the genetic influences on a phenotype by showing a change in heritability of a phenotype when SES or HLE differs. If heritability is stronger in higher SES or better HLE families, it suggests the genetic potential is better actualized in more supportive home environment, which is in line with the Bioecological model (Bronfenbrenner & Ceci, 1994). However, if heritability is stronger in lower SES or poorer HLE families, it suggests stressors in less favorable environment could lead to the expression of the deleterious genes, and this is consistent to the view of the Diathesis-stress model (Scarr, 1992). These interactions of gene and environment have begun to draw the attention of language and reading researchers, and several twin studies have examined how heritability and environmental influences might vary in different home environments. However, these studies focused on different age ranges and mixed results have been found. In the studies examining general verbal ability, Asbury, Wachs and Plomin (2005) demonstrated higher heritability in verbal ability in families with higher chaos, and less informal and instructive parent-child communication in 4-year-olds, which was in line with the Diathesis-stress model, but Turkheimer, Haley, Waldron, D'Onofrio and Gottesman (2003) indicated SES, which was a composite of parental education, occupational status and income, did not moderate heritability of verbal ability in 7-year-olds. However, when more specific language and reading skills are concerned, Rowe, Jacobson and Van den Oord (1999) found heritability in vocabulary knowledge increased with higher parental education in 16-year-olds, supporting the Bioecological model. In particular, they found the heritability rose from .25 in adolescents whose parents with less than 12 years of education to .74 in those whose parents with above high school education. This parental education moderation effect was demonstrated even in later stage of life, specifically in word recognition ability of middle-aged men, and thus the influences of parental education could be long-lasting (Kremen et al., 2005). Some studies have extended this investigation on gene and environment interaction to the lower and the higher reading ability groups. Friend and colleagues indicated parental education moderated the heritability in reading abilities, but its direction differed in lower and higher reading ability groups (Friend et al., 2008; Friend et al., 2009). Specifically, the genetic influences were stronger in children of better-educated parents in the lower reading ability group (mean age of 11.5 years, and were identified with reading performance 1.5 S.D. or more below the mean of the group with no school history of reading difficulties), but the heritability was stronger in children of poorer-educated parents in the higher reading ability group (kindergarten to the sixth grade, and were identified with reading performance 1 S.D. above the sample mean). Friend et al. (2009) concluded these findings, which sounded contrastive, actually provided a coherent account by suggesting genetic influences were more prominent when reading ability was inconsistent with the environmental expectations, that was, either having poor performance in more supportive environment or good performance in less favorable environment. Researchers have proposed the same sets of genetic influences would function across the entire distribution of reading ability (Plomin & Kovas, 2005), and so it could be the same genes but with different size of effect when environments differ.

These studies have shown the mediating and the moderating roles of SES and HLE, and have dissected variances of language and reading skills nicely into genetic and environmental contributions. However, the presence of genotype-environment correlation may create potential complications in these findings (Kendler & Baker, 2007; Plomin &

Neiderhiser, 1992). Genotype-environment correlation occurs when the exposure to different environments is at least partially determined by an individual’s genotype. For instance, a child who is genetically at risk of language difficulty may also have impoverished language learning environment at home because the parents have the same genetic risk, and these jointly contribute to the poor language ability of the child. There are several mechanisms of this genotype-environment correlation, including passive, evocative and active (Scarr & McCartney, 1983). A child can passively receive both genes and environments form his or her parents (passive), evoke experiences from others’ reactions to his or her behaviours according to his/her genotypes (evocative), or actively search for environments which link to his or her genotypes (active). Studies have shown family environment variables, such as parent-child interaction, could be partly influenced by genetic factors (Plomin et al., 1994). The heritability of family environment was found to be 18 to 30% in a review paper (Kendler & Baker, 2007). In twin study design, genetic and environmental influences are divided clearly and the environmental effects are considered as the portion of variances that is not accounted for by the genetic ones. Therefore, the presence of genotype-environment correlation effects could lead to spurious findings of environmental mediation because the family environment variables are not pure measures of shared environment but with the genetic influences subsumed (Purcell & Koenen, 2005; Turkheimer, D’Onofrio, Maes, & Eaves, 2005). The variables which their variances are highly determined by genetic factors are particularly prone to these masked genetic influences. The potential genotype and environmental correlation effects should be taken into consideration when interpretating results of twin research. There are three ways to deal with this problem, including a) comparing correlations of measured environments with traits between adoptive and nonadoptive families; b) conducting analyses for traits where there are no genetic effects; and c) conducing analyses using measured environments that are not heritable. Nevertheless, researchers have pointed out that this genotype and environment correlation was negligible and results of twin research could reflect the true shared environmental influences (Petrill et al., 2004). In general, with the potential genotype and environmental correlation borne in mind, the twin study method helps identify the sources of environmental influences and demonstrate their roles in individual differences with genetic impacts considered.

Research questions

This thesis investigates how family socioeconomic status (SES) and home literacy environment (HLE) as identified home environmental factors contribute to the genetic and environmental influences on Chinese language and reading skills in 312 Chinese twin pairs. The effects of family SES and HLE are tested independently, to indicate the influences of relatively distal and proximal home environmental factors respectively. These relationships were examined in two ways. First, the extent to which SES/ HLE mediates the shared environmental influences is tested. It is predicted that SES and HLE will each explain a significant portion of variances in language and reading skills. As parental input could be particularly important in learning to read Chinese (Li & Rao, 2000), it is expected that family SES and HLE will explain a higher portion of variances than those found in past twin studies on alphabetic languages. Second, the moderation effects of SES/ HLE on the relative genetic and environmental contributions are examined. Given the mixed findings from past twin studies, no specific predictions are made and so this thesis aims to explore their moderation effects.

The extent to which the SES and HLE contribute to more general language and reading abilities (i.e., General Language and General Reading; see Chapter 3) through mediation of shared environmental influences and/or moderation of heritability is examined, so as to understand how they might affect language and reading development in a broader context. The need to investigate these broader aspects of general performance has been underscored (Hart, Petrill, DeThorne, et al., 2009). In addition, this thesis aims to test the mediation effects of SES and HLE on several specific language and reading skills, including receptive vocabulary, syllable and rhyme awareness, and orthographic skills, which showed significant environmental contributions in the univariate genetic analyses (see Chapter 3).

Method

Family SES and HLE conditions were obtained by parent-reported questionnaire. Among the 312 families participated, family SES and HLE conditions were successfully collected from 296 and 215 families respectively. The scores of receptive vocabulary, syllable and rhyme awareness, orthographic skills, General Language and General Reading were included in the following analyses (see Chapter 2 for participant and task details).

Measures of SES and HLE

Socioeconomic status (SES). Information on paternal and maternal education level and family income was collected from questionnaires filled in by parents. An exploratory factor analysis indicated these three variables yielded a single-factor solution with an eigenvalue of 2.27 and similar factor loadings on each variable (.88, .87, and .85 for paternal and maternal education level and family income respectively). So, a composite index of SES was computed from the average of their standardized scores.

Home literacy environment (HLE). Home literacy environment information was obtained through 18 items in a parent-reported questionnaire. These items are presented in Table 4.1. Each cotwin of a twin pair got a rating on each item, except the item about the duration of parents' own reading per day. An exploratory factor analysis was employed to explore the dimensionality of the 18 home literacy environment items. Five factors with eigenvalues over one were indicated. Direct Oblimin rotation showed items clustered on five factors, namely 1) Reading resources and opportunities at home; 2) Time when parental instruction begun; 3) Literacy activities engaged; 4) Duration of parents' daily instruction; and 5) Parents' literacy teaching methods. However, one of these items (i.e., Duration of parents' own reading per day) did not have a factor loading greater than .40 on any of the five factors. The factor loadings are shown in Table 4.1. Factor scores of each factor were used in further analyses. Cotwins' correlations on each factor were all greater than .99, indicating the ratings of cotwins in a family were almost identical. Therefore, the cotwins' ratings on each factor were averaged to represent the home literacy situation in each family.

Table 4.1. Rotated Factor Loadings of Exploratory Factor Analysis on 18 Home Literacy Environment Items.

Items	HLE1_ Reading resources & opportunities at home	HLE2_ Time when parental instruction began	HLE3_ Literacy activities engaged	HLE4_ Duration of parents' daily instruction	HLE5_ Parents’ literacy teaching methods
No. of books owned by the child	**.77**	-.18	-.02	-.01	-.05
No. of books read by the child per week	**.68**	.12	.22	-.10	.08
Duration of parent-child reading per week	**.44**	.21	.08	.37	.15
Any discussion of book content	**.52**	-.04	-.23	.11	.34
When did parent-child reading start	**-.53**	**.55**	.03	.16	.00
When did word reading instruction start	-.06	**.86**	-.01	-.01	-.17
When did word writing instruction start	.18	**.83**	-.15	-.13	-.07
Library visit per month	.03	-.07	**.61**	-.02	-.05
Duration of Chinese learning extracurricular activities participated by the child	-.04	.03	**.68**	.00	.05
Duration of educational TV/DVD watching for Chinese learning by the child per week	-.02	.03	**.72**	-.03	.23
Duration of educational software usage for Chinese learning per week	.12	-.11	**.70**	.13	-.14
Duration of parents working with child on his/her homework per day	.14	-.25	.05	**.60**	-.15
Duration of word reading instruction per day	-.04	.03	.02	**.92**	.07
Duration of word writing instruction per day	-.10	.05	.00	**.92**	.11
Is learning words through daily life experiences encouraged (e.g., labels)	.04	-.12	-.05	.20	**.68**
Are flash cards or posters used	.02	-.15	.06	.08	**.76**
Are songs or nursery rhymes used	.01	.02	.08	-.13	**.74**
Duration of parents' own reading per day	.37	.10	.11	.19	.03

Note. Factor loadings over .4 are in bold; Signs are reversed for all items in Duration of parents’ daily instruction for easier interpretation.

Results

Age- and nonverbal reasoning-adjusted language and reading skill scores were used in all analyses.

Phenotypic relationships

Descriptive statistics of and correlations between SES and HLE factors were presented in Tables 4.2 and 4.3.

Table 4.2. Descriptive Statistics of Home Literacy Environment Factors.

	MZ		DZ		*t*
	M	*S.D.*	*M*	*S.D.*	
SES	-0.06	0.86	0.18	0.87	-3.01*
HLE1_Reading resources and opportunities at home	-0.04	0.98	0.09	1.06	-0.85
HLE2_Time when parental instruction began	-0.06	0.99	0.14	1.02	-1.33
HLE3_Literacy activities engaged	0.04	0.96	-0.09	1.09	0.86
HLE4_Duration of parents' daily instruction	0.05	0.97	-0.12	1.06	1.15
HLE5_Parents’ literacy teaching methods	-0.05	1.03	0.11	0.92	-1.07

Note. **p*<.05; *N* of SES= 218 MZ pairs and 80 DZ pairs; *N* of HLE= 150 MZ pairs and 65 DZ pairs.

Table 4.3. Correlations between SES, Home Literacy Environment Factors, and Language and Reading Measures.

	SES	HLE1_ Reading resources and opportunities at home	HLE2_ Time when parental instruction began	HLE3_ Literacy activities engaged	HLE4_ Duration of parents' daily instruction	HLE5_ Parents' literacy teaching methods
Home literacy environment factors						
HLE1_Reading resources and opportunities at home	.39**					
HLE2_Time when parental instruction began	-.10	-.15*				
HLE3_Literacy activities engaged	-.12	.19**	-.09			
HLE4_Duration of parents' daily instruction	.01	.23**	-.07	.21**		
HLE5_Parents' literacy teaching methods	.18**	.28**	-.09	.14*	.23**	
Language and reading variables						
General Language	.23**	.28**	-.11	-.10	.12	.11
General Reading	.10	.17*	-.29**	.19**	.15*	.05
Receptive vocabulary	.13*	.31**	.02	.02	.08	.04
Syllable and rhyme awareness	.23**	.20**	-.07	.01	.27**	.17*
Orthographic skills	-.02	.15*	-.12	.15*	.18**	.07

Note. *p<.05; ** p<.01. N ranged from 207 to 298. Language and reading variables included the scores of a twin randomly selected from each pair.

SES significantly correlated with two HLE factors (*Reading resources and opportunities at home,* and *Parents' literacy teaching methods*). Also, all HLE factors were linked with each other, except *Time when parental instruction begun* was correlated with *Reading resources and opportunities at home* only. Though the MZ and the DZ twins did not differ significantly on all HLE factors, the DZ twins had higher SES than the MZ twins. However, the SES difference had little consequence for the genetic analyses because these analyses were based on the cotwins' covariances.

Correlations between SES and HLE, and language and reading variables are presented in Table 4.3. These correlations were computed based on the language and reading scores of a twin randomly selected from each twin pair. SES significantly correlated with General Language, receptive vocabulary, and syllable and rhyme awareness (*r* ranged from .13 to .23, *p*s<.05). The five HLE factors had differential relationships with the language and reading variables. Among the HLE factors, only *Reading resources and opportunities at home* was significantly correlated with all of these variables (*r* ranged from .15 to .31, *p*s<.05), showing better reading resources and opportunities at home were linked to higher language and reading abilities. Other HLE factors were linked to some of the language and reading variables. Specifically, *Time when parental instruction begun* was negatively correlated with General Reading only, indicating parental instruction started at an earlier age was associated with better General Reading ability. *Literacy activity engaged* and *Duration of parents' daily instruction* were positively correlated with General reading and orthographic skills, and *Duration of parents' daily instruction* was linked to syllable and rhyme awareness as well. Lastly, *Parents' literacy teaching methods* was significantly linked to syllable and rhyme awareness only. Given the importance of *Reading resources and opportunities at home* on all language and reading variables of interest, this factor was included in further analyses. To reiterate, five items loaded on this factor including, a) number of books owned by the child, b) number of books read by the child, c) duration of parent-child reading, d) discussion of book content, and e) time when parent-child reading started. Also, to test if the correlations of SES and *Reading resources and opportunities at home* with language and reading variables differ between the MZ and the DZ twins, each pair of contrasting correlation coefficients were transformed with the Fisher Z-transform. These correlations did not differ significantly across zygosity (*p*s>.05; see Table 4.4).

Table 4.4. Correlations between SES, Reading Resources and Opportunities at Home and Language and Reading Measures by Zygosity.

	SES			Reading resources and opportunities at home		
	MZ	DZ	*z* score	MZ	DZ	*z* score
General Language	.22**	.25*	-0.81	.26**	.32*	-0.42
General Reading	.09	.12	-0.22	.15	.22	-0.47
Receptive vocabulary	.10	.19	-0.69	.31**	.31*	0.00
Syllable and rhyme awareness	.22**	.23*	-0.08	.17*	.26*	-0.62
Orthographic skills	.00	-.08	0.60	.24**	-.03	1.81

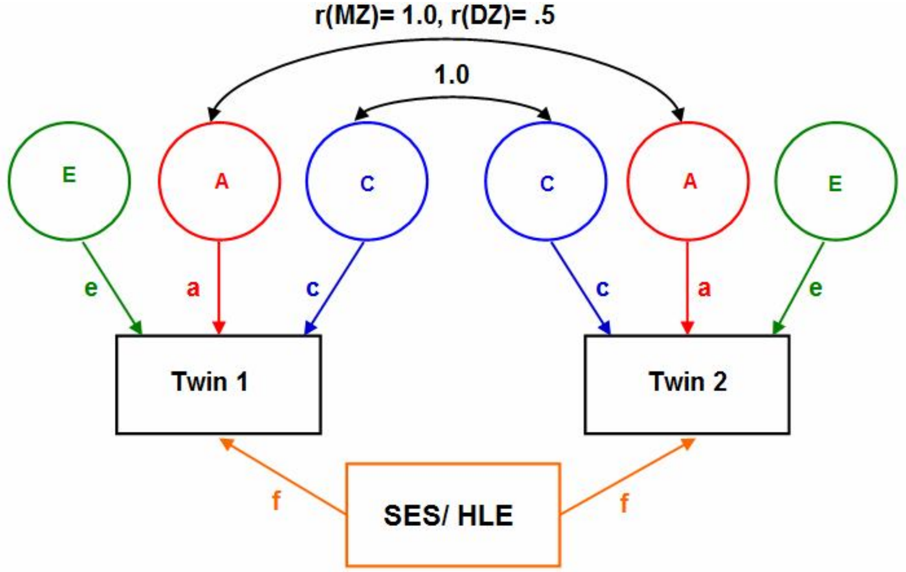
Note. **p*<.05; ***p*<.01. *N* ranged from 125 to 159 for MZ and from 61 to 80 for DZ. *z* score with an absolute value exceeding 1.96 indicates a significant difference.

Mediation of identified home environmental factors on shared environmental influences

Scores of language and reading tasks were normalized to reduce the influences of skewness and extreme cases in model fitting. Specifically, the age- and nonverbal reasoning-adjusted scores were transformed into new standardized scores based on their cumulative frequencies for each variable. These scores and either SES or HLE were fitted to the ACE mediation models by OpenMx in R statistical modeling package (Braun & Murdoch, 2007).

Genetic analyses. A model adapted from Petrill et al. (2004) was employed to understand the contributions of the identified home environmental factors to the individual differences (see Figure 4.1). Specifically, this model estimates the proportion of variance of the observed variable explained by the identified home environmental factors together with the genetic (A), shared environmental (C) and nonshared environmental (E) factors. The home environmental factors of interest, in particular SES and *Reading resources and opportunities at home*, were fitted into the model separately. SES and *Reading resources and opportunities at home* are conceptualized as an aspect of shared environmental influences because they are in common between the cotwins of a twin pair. Thus, shared environment is decomposed into variances explained by SES or *Reading resources and opportunities at home* and unidentified shared environmental influences in each of the models.

Figure 4.1. Univariate ACE mediation model estimating genetic, shared environmental, nonshared environmental, and SES/HLE contributions to variances in language and reading variables.



So, for each variable, three ACE models were tested; a) a model with SES; b) a model with *Reading resources and opportunities at home*; and c) a model with neither SES nor *Reading resources and opportunities at home*. These mediation models had satisfactory goodness-of-fit (see Appendix 2B for model fitting statistics). Path estimates and proportion of contribution to variances are shown in Table 4.5. Model fitting statistics comparing the models with and without a mediator are also shown in Table 4.5.

Table 4.5. Model Fitting Statistics and Parameter Estimates of Genetic Models with SES or Reading Resources and Opportunities at Home on Language and Reading Measures. (95% Confidence Intervals in Parentheses)

	Model fit		Path estimates (unstandardized)				Proportion of variance explained (squared standardized paths)			
	$\Delta\chi^2$ (Δdf =1)	*p*	a	c	e	SES/ HLE	a^2	c^2	e^2	SES^2/ HLE^2
SES										
General Language	17.59	.00	0.59 (0.31,0.87)	0.30 (-0.25,0.85)	0.59 (0.53,0.64)	0.47 (0.36, 0.57)	.35	.09	.34	.22
General Reading	10.66	.00	0.68 (0.43,0.92)	0.29 (-0.29,0.86)	0.54 (0.49,0.59)	0.41 (0.29,0.54)	.45	.08	.29	.17
Receptive vocabulary	7.38	.01	0.31 (-0.23,0.85)	0.55 (0.26,0.85)	0.68 (0.62,0.74)	0.37 (0.24,0.50)	.10	.31	.46	.14
Syllable and rhyme awareness	24.93	.00	0.29 (-0.26,0.83)	0.49 (0.18,0.80)	0.65 (0.60,0.71)	0.50 (0.40,0.60)	.08	.24	.43	.25
Orthographic skills	0.40	.53	0.24 (-0.54,1.02)	0.63 (0.35,0.91)	0.74 (0.67,0.81)	--	.06	.40	.55	--
HLE (Reading resources and opportunities at home)										
General Language	13.63	.00	0.58 (0.29,0.86)	0.32 (-0.19,0.83)	0.59 (0.53,0.65)	0.47 (0.35,0.59)	.33	.10	.35	.22
General Reading	6.42	.01	0.67 (0.43,0.92)	0.31 (-0.22,0.85)	0.54 (0.49,0.59)	0.40 (0.25,0.55)	.45	.10	.29	.16
Receptive vocabulary	13.95	.00	0.27 (-0.33,0.87)	0.49 (0.18,0.81)	0.68 (0.62,0.74)	0.47 (0. 35,0.59)	.07	.24	.46	.22
Syllable and rhyme awareness	7.73	.01	0.29 (-0.25,0.83)	0.57 (0.29,0.84)	0.65 (0.60,0.71)	0.41 (0.27,0.55)	.08	.32	.43	.17
Orthographic skills	2.99	.08	0.24 (-0.54,1.02)	0.63 (0.35,0.91)	0.74 (0.67,0.81)	--	.06	.40	.55	--

Note. Model fit shows model fitting statistics of the ACE model with SES/HLE as a mediator compared to the model without SES/HLE. Number of twin pairs ranged from 142 to 214 for MZ and from 59 to 77 for DZ. a^2 = additive genetic variance; c^2 = shared environment variance; e^2 = nonshared environment variance; SES^2 = socioeconomic status variance; HLE^2 = home literacy environment variance.

SES. Results indicated that the model fit decreased significantly when SES was removed from the model for General Language, General Reading, receptive vocabulary and syllable and rhyme awareness (*p*s<.05), and the SES paths were significant in these models, indicating SES played a prominent role in explaining the variances of these skills. The proportion of variances explained by SES was obtained from standardizing and then squaring the SES path estimate. It was found that SES accounted for 22%, 17%, 14% and 25% of the total variances of General Language, General Reading, receptive vocabulary and syllable and rhyme awareness respectively. In other words, SES contributed to 71%, 68%, 31% and 51% of the total shared environmental influences on General Language, General Reading, receptive vocabulary and syllable and rhyme awareness respectively [71%= .22/(.22+.09); 68%= .17/(.17+.08); 31%= .14/(.14+.31); 51%= .25/(.25+.24)]. However, for orthographic skills, the models with and without SES had similar model fit (*p*>.05), and thus SES did not contribute significantly to the variances in orthographic skills.

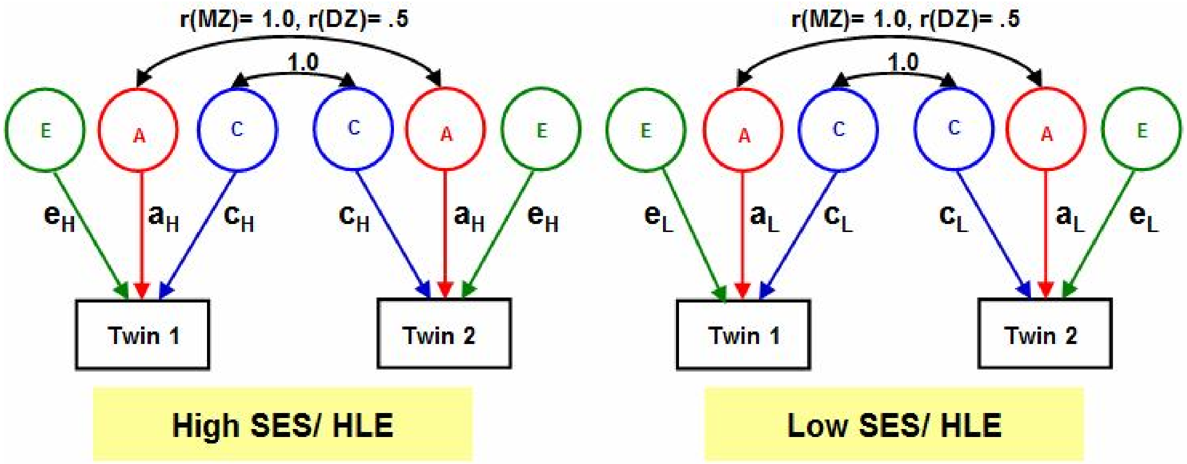
Reading resources and opportunities at home. The model fit dropped significantly when *Reading resources and opportunities at home* was removed from the model for General Language, General Reading, receptive vocabulary and syllable and rhyme awareness (*p*s <.05), and the *Reading resources and opportunities at home* paths were significant in these models, indicating *Reading resources and opportunities at home* was important in explaining the variances of these skills. Specifically, *Reading resources and opportunities at home* accounted for 22%, 16%, 22% and 17% of the total variances of General Language, General Reading, receptive vocabulary and syllable and rhyme awareness respectively. In other words, it contributed to 69%, 62%, 48% and 35% of the total shared environmental influences on

General Language, General Reading, receptive vocabulary and syllable and rhyme awareness respectively [69%= .22/(.22+.10); 62%= .16/(.16+.10); 48%= .22/(.22+.24); 35%= .17/(.17+.32)]. However, the models with and without *Reading resources and opportunities at home* had similar model fit for orthographic skills (p>.05), and so *Reading resources and opportunities at home* had a minimal role in orthographic skills.

Moderation effects of identified home environmental factors.

Next, moderation effects of identified home environmental factors on general language and reading abilities were examined by dividing the sample into two groups with mean SES or HLE (*Reading resources and opportunities at home*) as the cut-off and then testing the differences between these groups (i.e., low vs high SES; low vs high HLE [*Reading resources and opportunities at home*]). These split-group analyses yielded 126 MZ and 34 DZ pairs in the low SES group while 92 MZ and 46 DZ pairs in the high SES group, and 78 MZ and 25 DZ pairs in the low HLE group while 72 MZ and 40 DZ in the high HLE group. Data of the two groups were fitted to a model constraining genetic and environmental parameters to be equal for the low SES and the high SES groups (or the low and the high HLE [*Reading resources and opportunities at home*] groups) (homogenous model), and another model allowing different genetic and environmental parameter estimates for the two groups (heterogeneous model). The poorer fit of the homogenous model than the heterogeneous model indicates the two groups differ on the estimates, and thus moderation effects are demonstrated. Figure 4.2 depicts these models for testing moderating effects of SES/HLE.

Figure 4.2. Univariate ACE models for the two groups for testing moderating effects of SES/ HLE (Reading resources and opportunities at home).



The model comparison for testing moderating effects was conducted on a) SES and General Language; b) SES and General Reading; c) HLE (*Reading resources and opportunities at home*) and General Language; and d) HLE (*Reading resources and opportunities at home*) and General Reading. Results showed that the homogeneous and the heterogeneous models did not differ significantly on model fit for all models ($\Delta\chi^2[\Delta df = 3] =$ 3.33 for SES and General Language; 2.74 for SES and General Reading; 2.68 for *Reading resources and opportunities at home* and General Language; and 1.71 for *Reading resources and opportunities at home* and General Reading, $ps > .05$]. Thus, the same set of genetic and environmental parameters applied to both groups in the SES as well as the *Reading resources and opportunities at home* models. These results suggest SES and *Reading resources and opportunities at home* did not moderate the heritability of Chinese language and reading skills.

Discussion

This thesis examined the roles of family SES and HLE on Chinese children's language and reading abilities with a twin study design. Their relationships with general language and reading abilities were tested in two ways, including a) mediation of shared environmental influences and b) moderation of genetic influences. More specific skills including receptive vocabulary, syllable and rhyme awareness and orthographic skills, which showed significant shared environmental contributions, were also tested for the mediation effects on shared environmental influences. The unique contributions of this part of the thesis are threefold. First, it investigated both possible mediating and moderating roles of SES and HLE together in a single study. Second, it extended the examination of these links to cognitive skills, including syllable and rhyme awareness and orthographic skills. They provide more comprehensive understanding on how SES and HLE contribute to the individual differences in various language and reading skills. Lastly, by studying Chinese language and reading skills, it could shed light on whether these links are universal across different languages. Overall, the results showed that SES and HLE (Reading resources and opportunities at home) mediated the shared environmental contributions to general language and reading abilities as well as more specific skills, including receptive vocabulary and syllable and rhyme awareness, but not orthographic skills. However, they did not moderate genetic influences on general language and reading abilities.

SES and HLE as plausible environmental mediators

Results indicated SES and HLE (Reading resources and opportunities at home) had significant contributions to general language and reading abilities when genetic influences were controlled for, in which SES and HLE (Reading resources and opportunities at home) individually explained around one-fifths and one-sixths of the variances in General Language and General Reading respectively. SES and HLE (Reading resources and opportunities at home) mediated the majority (over three-fifths) of shared environmental influences. In addition, both of them explained a significant proportion of variances in receptive vocabulary knowledge, in which SES and HLE accounted for 14% and 22% of its total variances, and mediated around one-thirds and half of the shared environmental influences, respectively. Compared to past twin studies, both SES and HLE contributed to a higher portion of variances in language skills in this thesis. Specifically, Petrill et al. (2004) found SES accounted for 3 to 4% of the variances in expressive vocabulary and grammatical complexity and mediated 5 to 7% of the shared environmental influences. Also, Hart, Petrill, DeThorne, et al. (2009) found HLE accounted for 6 to 10% of the variances in expressive vocabulary and mediated 13 to 27% of the shared environmental variances. These findings are in line with the prediction that SES and HLE are even more important in children learning to read Chinese than their counterparts learning alphabetic scripts.

Also, results in this chapter showed the mediating roles of SES and HLE on syllable and rhyme awareness, which has not yet been examined in previous twin studies. SES and HLE (Reading resources and opportunities at home) explained 25% and 17% of its total variances, and mediated over half and one-thirds of the shared environmental influences, respectively. As demonstrated in past research, SES and HLE which included reading resources and opportunities was linked to children’s phonological sensitivity (e.g., Bowey, 1995; Burgess et al., 2002), and findings of this thesis have further strengthened this link with a genetically sensitive design. In other words, the relationships between SES and HLE and phonological skills persisted even when genetic influences were controlled for. It is interesting to find SES and HLE (Reading resources and opportunities at home) contributed to syllable and rhyme awareness but not orthographic skills, which suggest differential impacts of home environment on various cognitive skills. One possible reason is that orthographic skills are mostly learnt through explicit instruction in school, and thus home environment including SES and HLE have modest influences on the variances in orthographic skills. Another possible reason is that reading resources and opportunities at home as one aspect of HLE may have a weaker link with orthographic skills than the other HLE aspects. It could be in the similar vein with past findings that explicit instruction was more important than shared book reading in word reading development (Sénéchal & LeFevre, 2002), given both orthographic skills and word reading focus on processing prints. In this part of the thesis, *Reading resources and opportunities at home* had a similar or greater magnitude of phenotypic correlation with orthographic skills compared to those of the other HLE factors. Therefore, this possible mechanism might involve HLE factors beyond the five HLE factors examined. Further studies are needed to understand the potential mechanisms. Also, it was noted that the correlations between HLE (Reading resources and opportunities at home) and orthographic skills tended to be higher in MZ than DZ twins. These correlations did not significantly differ, and thus the real difference was not substantiated. However, if there was a real difference, the results would be spurious. Future research could confirm the findings.

This thesis indicated both SES and HLE (Reading resources and opportunities at home) did not moderate the heritability of general language and reading abilities, and thus has contributed to clarifying the mixed results demonstrated in past studies (e.g., no moderation, Turkheimer et al, 2003; Diathesis-stress model supported, Asbury et al., 2005; Bioecological model supported, Rowe et al., 1999). Therefore, it implies the same sets of genetic influences might function across the entire distribution of reading abilities, regardless of the variations in family SES and HLE (Reading resources and opportunities at home). As Friend and colleagues (2008, 2009) indicated SES moderation in both high and low reading ability groups, there is a possibility that SES moderation might be more obvious in the higher and the lower ends, than in the entire range of reading abilities. However, there has been a handful of research in this area and no converging finding has been obtained, and so these conclusions need further verification.

Limitations

There are three caveats in this part of the thesis. First, the measures employed might not fully characterize SES and HLE, and the genetic analyses focused on the *Reading resources and opportunities at home* domain only for HLE. In this thesis, the ways of tapping SES and HLE were similar to those in other studies (e.g., Burgess et al, 2002; Turkheimer et al., 2003), and *Reading resources and opportunities at home* linked to all language and reading variables of interest while other HLE domains did not, which justified the focus on this domain. However, the investigation of other potential domains and conceptualization of SES and HLE could provide a more comprehensive picture on their relationships. Second, the sample size in the split-group analyses was relatively small, and thus possible moderating effects of SES and HLE might not be detected. Lastly, to reiterate, the potential genotype and environment correlation might lead to spurious results.

Implications and conclusions

Despite these limitations, this thesis has contributed to the scarce research on how home environment influenced children’s language and reading skills in a twin study design. First, it has added to past research on alphabetic languages that SES and HLE mediated the shared environmental influences on language and reading skills, suggesting these links could be universal across languages. It should be noted that SES and HLE explained a higher portion of these variances in Chinese than those in alphabetic languages, highlighting the special roles of home environment in learning Chinese. These results suggest extra support should be given to low SES families, and enhancing HLE by providing support on reading resources and ways for parents to engage in reading with children, could be particularly helpful in children’s language and reading development. Second, findings of this thesis have extended the significant contributions of SES and HLE to phonological awareness at the syllable and rhyme levels when genetic influences were accounted for. However, these links were not shown in orthographic skills. So, SES and HLE could have differential links with various cognitive skills, and their mechanisms are promising topics for further investigation. Finally, there was a lack of moderating effect of SES and HLE on the heritability of general language and reading abilities. This finding implies the genetic and environmental influences on language and reading skills might be generally stable across the population with varying SES and HLE.

Chapter 5

Why do Chinese language and reading skills remain stable or change across time? The roles of genes and environments.

Overview

This part of the thesis examines the etiology of the stability and instability in Chinese language and reading skills in 292 Chinese twin pairs. Children were tested on tasks of Chinese word reading, receptive vocabulary, phonological memory, tone awareness, syllable and rhyme awareness, rapid automatized naming, morphological awareness and orthographic skills, as well as nonverbal reasoning and audiometric screening tests. They were tested again on the same tasks, except nonverbal reasoning, one year after the initial testing.

Results indicated that these language and reading skills were moderately to largely stable across time. Genetic analyses indicated their stability was mainly mediated by genetic influences, but some evidence showed shared environment exerted stronger influences on the stability of syllable and rhyme awareness. Also, findings showed new genetic and environmental factors came into play at Time 2 for Chinese word reading, and suggested new genetic influences and new shared environmental influences contributed to the changes of tone awareness and morphological awareness at Time 2 respectively. Some of these results have extended past research on alphabetic languages and hence suggest universal findings, and these include the genetic contributions to the stability of word reading and phonological awareness and the nonsigificant shared environmental contribution to rapid automatized naming. In addition, this thesis have examined a wider range of reading-related cognitive skills which were not studied in past longitudinal twin research focusing on normal range reading development (i.e., phonological memory, tone awareness, morphological awareness, and orthographic skills), and thus has contributed to a more comprehensive picture of the etiology of the stability in language and reading skills across time.

Research Background

Reading development is a continuous and cumulative process. Children have to master multiple skills progressively in learning to read, from linking meaning to print which emphasizes visual and contextual connections, to mastering print-sound correspondence, and finally reading words automatically and proficiently (Ehri, 2005). In spite of the similar developmental trajectory, individuals vary in reading abilities and these variations tend to emerge early in development and continue over time, and thus children with early delay are more likely to be lower achievers later on, while those who have early success in reading often continue to be higher achievers (Cunningham & Stanovich, 1997). In other words, developmental origins of subsequent reading achievement could be indicated in early reading ability, from the time children begin receiving reading instruction in school (i.e., reading performance in kindergarten or early elementary school grades; e.g., Juel, 1988; Torgesen & Burgess, 1998), and even from the preschool period (i.e., emergent literacy skills, such as letter knowledge; e.g., Lonigan, Burgess, & Anthony, 2000). Studying the underlying mechanisms of reading consistency is promising, as it is key to our understanding of the course of reading development and constructing effective learning environments for children. Past research has pursued these underlying mechanisms by identifying potential mediating factors, such as print exposure (e.g. Cunningham & Stanovich, 1997), and early cognitive correlates of reading, such as phoneme sensitivity (e.g., Muter, Hulme, Snowling, &

Stevenson, 2004), and naming speed (e.g., Schatschneider, Fletcher, Francis, Carlson, & Foorman, 2004). However, these factors and skills could be determined by more fundamental factors, such as genetic influences (e.g., Harlaar et al., 2007a). Nevertheless, little is known about how these fundamental factors including genetic influence in the stability of reading skills and related cognitive skills.

To date, scarce research has examined the roles that genes and environments play in the continuity of reading ability with a twin study design, and the findings so far have supported the genetic influences are important factors in the process (e.g., Byrne et al., 2005; Harlaar et al., 2007a; Petrill et al., 2007; Samuelsson et al., 2008). However, these longitudinal twin studies have been limited to alphabetic languages, mainly English, and a few reading-related cognitive skills only. Therefore, the present research has extended these past studies by investigating the relative genetic and environmental contributions to the continuity as well as the changes in Chinese reading and various reading related skills, including receptive vocabulary knowledge, syllable and rhyme awareness, rapid automatized naming, phonological memory, tone awareness, morphological awareness and orthographic skills.

Reading development

Learning to read is a continuous process of understanding visual symbols which represent speech (Ziegler & Goswami, 2005). Children have to master multiple skills in this process, including the awareness of phonological (e.g., Torgesen et al., 1992), morphological (e.g., Mahony et al., 2000), and orthographic (e.g., Cunningham et al., 2001) units of print. These skills evolve in the course of reading development, highlighting different reading strategies in each phase. Researchers like Ehri (1995) and Frith (1985) proposed reading advances from focusing on the contextual or visual features to utilizing print-sound correspondences and consolidated letter patterns. In general, early reading emphasizes contextual and visual features, such as treating a word as an unanalysed visual unit and focusing on certain more salient graphemic features (e.g., initial letters), and these connections are often arbitrary and learned by rote (Ehri, 2005). When children have gained more experience with print, understanding of the systemic print-sound correspondences develops and this knowledge is employed in reading (Ehri, 2005). Utilizing print-sound correspondence rules is an important milestone in reading development, as it provides a reliable way to memorize words and learning new words, and significantly reduces the memory load compared with memorizing each word as an unanalysed visual unit by rote (Chan & Siegel, 2001). While phonological awareness supports reading, reading experience also sensitizes children to phonological units, and thus they develop hand in hand. Converging evidence has shown the bidirectional relationships between phonological awareness and reading in alphabetic scripts (e.g., Wagner, Torgesen, & Rashotte, 1994), and even in scripts with relatively inconsistent print-sound correspondence, such as Chinese (Chow et al., 2005). Phonological awareness is important across scripts, but the phonological units in lexical representations and their related reading strategies differ depending on the print-sound correspondence consistency, as highlighted in the psycholinguistic grain size theory (Ziegler & Goswami, 2005). Specifically, while focusing on the small grain size, that is the phoneme, is sufficient for reading mastery in consistent scripts (e.g., German), additional correspondences for larger grain size units, including syllables, rimes, or the entire words, are required for more inconsistent scripts (e.g., English) (Ziegler & Goswami, 2005). In Chinese reading, a character and its components (strokes or radicals) do not represent phonemes but syllables instead, and thus reading in Chinese emphasizes larger psycholinguistic grain size units. It is in line with findings on Chinese reading, in which phonological awareness at the syllable and onset-rime levels contributes to Chinese word reading, but phonemic awareness has a diminished role (e.g., McBride-Chang & Ho, 2000; Siok & Fletcher, 2001).

Children learning to read Chinese gradually move from attending to visual features to phonological features of the characters, as with the case in English reading development of their English-speaking counterparts (Chan & Siegel, 2001). However, Chinese phonological features are relatively inefficient in aiding reading compared to those in English, because phonological cues are present in semantic-phonetic compound characters only and these cues could be misleading. Phonetic radicals in semantic-phonetic compound characters provide pronunciation cues to the character as a whole, and thus the print-sound correspondence in Chinese refers to the mapping between phonetic radicals and syllables. Reading a semantic-phonetic compound character is achieved by either deriving the pronunciation directly from the phonetic radical, or through analogy of characters in the same phonetic family (i.e., characters which share the same phonetic radical) (Ho, Ng & Ng, 2003). Phonetic radicals convey phonological information in three ways (Shu et al., 2003). First, they provide full phonological information (i.e., they possess the same pronunciation as that of the character; e.g., 桔 /gat1/ has the same pronunciation as the phonetic radical 吉 /gat1/). Second, they provide partial phonological information (i.e., their pronunciation is similar to but not exactly the same as that of the character; e.g., 結 /git3/ has the same onset but different rime and tone as the phonetic radical 吉 /gat1/). Third, they do not provide direct cues but making analogy with characters in the same phonetic family is useful in reading (e.g., 僵, 疆, 韁 which share the phonetic radical 畺 are pronounced as /goeng1/). Therefore, the utilization of phonemic principles requires considerable knowledge of words, and analytical skills in figuring out the amount of phonetic information that the phonetic radicals provide and the strategy (derivation or analogy) that is appropriate are essential in reading Chinese (see Chapter 1 for more thorough discussion).

Children as young as first graders can make use of the information in the phonetic radical in reading (Ho & Bryant, 1997). They develop their understanding of the phonemic principles in stages, from being aware of the function of phonetic radicals and pronouncing the characters as their phonetic radicals, to mastering more sophisticated use of partial phonological information that phonetic radicals convey to read semiregular characters (Chen, Shu, Wu, & Anderson, 2003). In general, children first develop rudimentary skills in utilizing phonological information in reading semantic-phonetic compounds, and overregularization could occur (Chan & Siegel, 2001). As they gain more reading experience, they develop skills in coping with the more complicated print-sound links, and thus are able to use partial phonological information in reading semiregular characters. These skills emerge when children become more experienced in reading, and even the initial stage needs the support of a vocabulary base of several hundred characters (Chen et al., 2003). More refined metalinguistic awareness is essential for a child to understand the more complicated link of the partial phonological information to the pronunciation of a character. For instance, tone awareness is a prerequisite to read tone-different characters, and onset-rime awareness is prominent in reading onset- or rime- different characters. Research has shown that children developed skills in reading tone-different characters before onset-different characters, using partial phonological information (Anderson, Li, Ku, Shu, & Wu, 2003).

The mastery of the phonemic principles is one major milestone of Chinese reading development, and the interpretation of character meaning from semantic radicals is another. Semantic radicals signify the semantic category which the characters belong to, and thus contribute to the semantic transparency of Chinese reading (e.g., the semantic radical 氵 which means water in 洋 [ocean /joeng4/]). Unlike the phonological links, there is no equivalence of these semantic connections in English (note that a Chinese character maps on a morpheme and is comparable to a morphemic unit in English word).

Similar to phonetic radicals, semantic radicals could stand alone themselves. However, among the total of approximately 200 semantic radicals, most of them are not a character and could not be pronounced on their own (Feldman & Siok, 1999). Shu et al. (2003) identified four kinds of semantic-phonetic compound characters according to their semantic transparency, including transparent, semitransparent, opaque and other. A transparent compound has a semantic radical providing a reliable cue to its meaning. It has the same or directly related meaning as its radical, or belongs to the semantic category which its semantic radical represents (e.g., 洋 [ocean /joeng4/] has a semantic radical 氵 meaning water). A semitransparent compound has indirect related meaning to its semantic radical, or its extended meaning is either directly or indirectly related to the meaning of the radical (e.g., 浮 [float /fau4/]). An opaque compound has meaning unrelated to its radical (e.g., 派 [distribute /paai3/]). Among the semantic-phonetic compound characters, 88% of them provide at least partial information on meaning (58% and 30% are semantically transparent and semitransparent respectively; Shu et al, 2003). Children as young as six were sensitive to the role of the semantic radical in representing character meaning, and also the legal position of the phonetic and semantic radicals in a character (Chan & Nunes, 1998). Some studies have shown that children learn the function of phonetic radicals before that of semantic radicals. For instance, children from grade one are able to use phonetic radicals for phonological cues, but not until grade three do they obtain meaning cues from semantic radicals (Ho, Ng, & Ng, 2003). However, other studies have shown the opposite (Chan & Nunes, 1998). Nevertheless, the knowledge of phonetic and semantic radials both emerges with reading experience and is key to reading development in Chinese.

This thesis does not test the knowledge on radicals directly, as they could be too difficult for young children, especially kindergarteners. When children are gaining experience with print and receiving explicit instruction on the function of radicals, they are at the same time getting sensitized to more general phonological and morphological units (e.g., syllables). Therefore, the emergence of the sense and knowledge of radicals could influence and be reflected in the development of reading and these more general skills (e.g., phonological awareness at the syllable and rhyme levels). The developmental etiology of Chinese reading and these more general cognitive skills is examined in this thesis.

Longitudinal twin studies

Though the genetic and environmental contributions to the development etiology of reading in alphabetic scripts, mainly English, is better known, these are not understood in a script which has distinctive linguistics characteristics, such as Chinese. To date, there have been four major twin projects investigating the developmental etiology of normal range development of reading and related skills in alphabetic scripts, including the Twins Early Development Study (TEDS) in the United Kingdom, the Western Reserve Reading Project (WRRP) and the Colorado Longitudinal Twin Study of Reading Disability (LTSRD) in the

United States, and the International Longitudinal Twin Study (ILTS) with samples in Australia, the United States and Scandinavia. It should be noted though some of their studies examined how different skills covary across time (e.g., Hayiou-Thomas et al., 2010), the developmental etiology of individual skills is the main scope of this thesis and is thus the focus in the following summary.

These projects have both similar and different foci of reading development. The TEDS examined children's general reading ability development assessed by teachers in ages 7 to 12 (Harlaar et al., 2007a; Harlaar et al., 2007b). They also tested reading comprehension ability with a web-based assessment and reading exposure at age 10. The WRRP and the ILTS focused on earlier stages of development, and tested not only reading ability, but also its related cognitive skills, including expressive vocabulary skills, phonological awareness and rapid automatized naming. While the WRRP tested twins every year from kindergarten or the first grade and the most recent paper reported growth in five waves (Hart, Petrill, DeThorne, et al., 2009; Hart, Petrill, Thompson, et al., 2009; Petrill et al., 2007), the ILTS followed children every year from preschool and the most recent paper reported their development up to grade two (Byrne et al., 2005, 2006, 2009; Samuelsson et al., 2008). The LTSRD included typically developing children as controls in their project. They examined normal range reading ability in older twins, with initial testing ages ranged from around 8 to 16, and follow-up testing around 5 to 6 years later (ages ranged from around 13 to 24) (Betjemann et al., 2008). As the reading levels they focused are beyond those examined in this thesis, their results are not included in the major reviews.

In general, past longitudinal twin studies have suggested three main findings concerning the roles of genes and environment in reading and its related skills. So far, these studies have examined reading, vocabulary, phonological awareness and rapid naming skills. First, reading and language skills tended to be stable across time and genetic factors played an important role in this stability. For instance, genetic factors contributed 58% to 93%[2] of the phenotypic correlations in word reading ability (Byrne et al., 2005; Harlaar et al., 2007a; Petrill et al., 2007; Samuelsson et al., 2008), and 39% to 52% of the phenotypic links in expressive vocabulary (Byrne et al., 2009; Hart, Petrill, DeThorne, et al., 2009; Petrill et al., 2007) across time. Studies on reading related skills have been limited, but converging evidence has suggested genetic mediation in the consistency of phonological awareness (Byrne et al., 2005, 2006; Petrill et al., 2007) and some evidence of genetic effects on the stability of rapid automatized naming (Byrne et al., 2005). It is also noted that these genetic influences tended to overlap across time. Substantial genetic correlations were found for word reading (.71 to .91 in Samuelsson et al., 2008), and teacher-assessed reading achievement (.65 to .80 in Harlaar et al., 2007b). Also, Harlaar et al. (2007a) found genetic influences on 7-year word reading accounted for 66% of the total genetic variance in 12-year word reading, and Petrill et al. (2007) showed genetic overlap at the two time points (kindergarten or grade one and one year afterwards) explained 53% of the total genetic variance at Time 2.

Second, shared environmental factors contributed to the stability of expressive vocabulary skills, but not rapid automatized naming, and mixed results have been demonstrated on reading and phonological awareness. For expressive vocabulary skills, shared environmental factors explained 40% to 56% of its phenotypic correlations across time (Byrne et al., 2009; Hart, Petrill, DeThorne, et al., 2009; Petrill et al., 2007). A high degree of these shared environmental factors were in common across time. Specifically, Petrill et al

---

[2] Figures on percentage of contribution in this section were computed from the path estimates provided in the papers if they were not explicitly given.

(2007) indicated the shared environmental overlap accounted for 95% of the total shared environmental variance at Time 2. Among these shared environmental influences, Hart, Petrill, DeThorne, et al. (2009) identified home literacy environment as an important factor contributed to this stability.

Lastly, new genetic factors emerged in rapid automatized naming ability. The new genetic influences contributed to around half (Byrne et al., 2005) to over four-fifths (Petrill et al., 2007) of the total genetic variances in rapid automatized naming at Time 2. Nevertheless, the impacts of genes and environments on the instability of other skills remain unclear. For instance, new sources of genetic influences in reading skills have been found in Harlaar et al. (2007b), Petrill et al. (2007) and the United States sample only in Samuelsson et al. (2008), but not the others (Byrne et al., 2005; Harlaar et al., 2007a).

Research questions

There have been a few longitudinal studies contributing to the better understanding of the genetic and environmental influences on the developmental etiology of reading in alphabetic scripts. However, some of their findings have been mixed, and there is a lack of a similar research on Chinese. This part of the thesis investigates how heredity and environment contribute to the developmental etiology of Chinese reading across a one-year time period in 292 Chinese twin pairs. The development of various reading and its related skills are examined, including word reading, vocabulary knowledge, phonological memory, tone awareness, phonological awareness at syllable and rhyme levels, rapid automatized naming, morphological awareness and orthographic skills.

There are three research questions derived from past research findings and characteristics of Chinese reading development. First, how stable are individual differences in various Chinese reading and related skills across time? Given the links between early and subsequent reading abilities demonstrated in past studies (e.g., Cunningham & Stanovich, 1997), it is predicted that the variations in reading and related skills will be rather stable over the one-year period. In other words, moderate to substantial correlations of performance across time are expected.

Second, to what extent do genes and environments contribute to the stability of performance? Past twin studies have found genetic contributions to word reading, vocabulary, and phonological skills, and environmental influences on vocabulary skills. Also, there is a lack of environmental influences on the continuity of rapid automatized naming skills. It is assumed these influences could be universal across scripts and so they are expected in Chinese reading as well. However, because of the relatively ambiguous print-sound correspondences in Chinese, diverse learning environments may be involved in the growth of syllable and rhyme awareness, and these environmental inputs could be important across time. Therefore, environmental influences on the stability of phonological awareness at syllable and rhyme levels are also expected. The genetic and environmental contributions to the other skills are left open because there is a lack of past research support and no strong reasons for particular predictions.

Third, to what extent do genes and environments contribute to the instability of performance? Past studies have shown new genetic factors affect rapid automatized naming, and the impacts of genes and environments on the instability of other skills remain unclear. As Harlaar et al. (2007b) and Petrill et al. (2007) highlighted new sources of genetic influences may come into play over time as reading skills rapidly develop in early stages of reading, it is thus predicted that new genetic factors will emerge in reading development. Also, as mentioned before, knowledge of phonetic and semantic radicals is useful to and emerges in Chinese reading development. When children are getting sensitized to this more difficult and advanced level of phonetic and semantic knowledge in reading, new genetic and environmental influences may also evolve. Putting them together, it is hypothesized that new genetic and/or environmental sources of influences could emerge for word reading, phonological awareness at syllable and rhyme levels, tone awareness and morphological awareness over time. Also, consistent with past findings on rapid automatized naming, it is expected that new genetic influences will be indicated. As this thesis tested the growth across a year, this period could be short for detecting new sources of genetic and environmental influences. Therefore, without a strong reason for expecting new genetic and environmental factors, the predictions lean towards a lack of independent genetic and environmental influences at Time 2 on the other skills. These hypotheses on genetic and environmental contributions to stability and instability across time are summarized in Table 5.1.

Table 5.1. A Summary of the Hypotheses on Genetic and Environmental Contributions to Stability and Instability across Time.

	Stability	Instability
Word reading	Gene (Byrne et al., 2005; Harlaar et al., 2007a; Petrill et al., 2007; Samuelsson et al., 2008)	Gene (Harlaar et al., 2007b; Petrill et al., 2007) Environment
Receptive vocabulary	Gene (Byrne et al., 2009; Hart, Petrill, DeThorne, et al., 2009; Petrill et al., 2007) Environment (Byrne et al., 2009; Hart, Petrill, DeThorne, et al., 2009; Petrill et al., 2007)	--
Phonological memory	--	--
Tone awareness	--	Gene and/or Environment
Syllable and rhyme awareness	Gene (Byrne et al., 2005, 2006; Petrill et al., 2007) Environment	Gene and/or Environment
Rapid automatized naming	NO Environment (Byrne et al., 2005; Petrill et al., 2007)	Gene (Byrne et al., 2005; Petrill et al., 2007)
Morphological awareness	--	Gene and/or Environment
Orthographic skills	--	--

Note. ‘Gene’ represents genetic contribution is predicted; ‘Environment’ represents environmental contribution is predicted; ‘No Environment’ represents a lack of environmental contribution is predicted; -- represents no particular prediction.

Results

Standardized scores adjusted for age and nonverbal reasoning were employed in all analyses.

Phenotypic stability across time

Phenotypic correlations between Time 1 and Time 2 performance of each variable after controlling for the effects of age and nonverbal reasoning are shown in Table 5.2. Performance on language and reading measures was moderately to highly correlated across

Time 1 and Time 2, ranging from .35 in orthographic skills to .79 in word reading across time (*p*s<.05). These correlations were computed from scores of a cotwin randomly selected in each pair.

Table 5.2. Descriptive Statistics and Phenotypic Correlations between Time1 and Time2 Scores Controlling for Age and Nonverbal Reasoning of Each Variable.

	Time 1			Time 2			Phenotypic Correlations
	N	*Mean*	*S.D.*	*N*	*Mean*	*S.D.*	
Age (in years)	584	6.70	1.89	584	7.64	1.89	
Word reading	583	81.77	63.13	584	108.29	58.39	.79**
Receptive vocabulary	584	52.58	17.04	584	59.85	13.72	.59**
Phonological memory	584	73.35	28.21	584	77.84	24.69	.64**
Tone awareness	584	7.22	3.04	584	7.74	3.32	.55**
Syllable and rhyme awareness	583	15.22	5.87	584	17.16	4.43	.57**
Rapid automatized naming	563	0.06	0.03	582	0.07	0.03	.71**
Morphological awareness	584	22.24	12.23	584	26.10	11.32	.50**
Orthographic skills	581	62.21	15.88	584	68.24	12.22	.35**

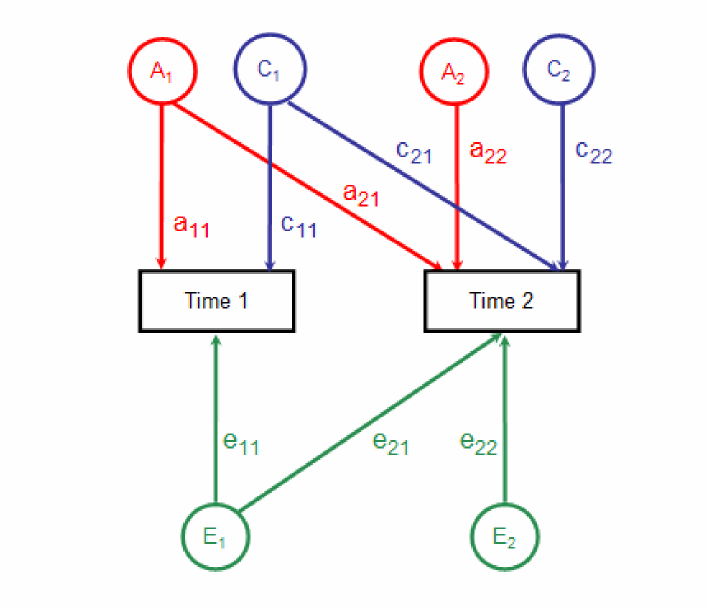
Note. **p<.01. Phenotypic correlations were computed from the scores of a twin randomly selected from each twin pair. *N* ranged from 283 to 291.

Contribution of genetic and environmental factors to stability and instability across time

Data were normalized to reduce the influences of skewness and extreme cases in model fitting. Specifically, the age- and nonverbal reasoning-adjusted scores were transformed into new standardized scores based on their cumulative frequencies for each variable. These scores were fitted to the Cholesky decomposition model by OpenMx in R statistical modeling package (Braun & Murdoch, 2007).

Cholesky decomposition model. The Cholesky decomposition model was employed to investigate the genetic and environmental contributions to the stability and instability of skills between Time 1 and Time 2 (see Figure 5.1).

Figure 5.1. Cholesky decomposition model concerning performance at two time points



The shared paths linking A1, C1, and E1 to Time 2 performance represented the genetic and environmental contributions to the stability. Also, the specific paths linking A2, C2 and E2 to Time 2 performance indicated the A, C, and E contributions to the instability across time. All models had satisfactory goodness-of-fit (*p*s>.05). The model fitting statistics and the path coefficients from the Cholesky decomposition models are shown in Tables 5.3 and 5.4 respectively. Results indicated that the nonshared environmental influences were largely specific to each of the time point, though small but significant nonshared environmental links were found. However, the nonshared environmental terms included measurement errors. Therefore, the genetic and shared environmental results are the main foci of interpretation.

Table 5.3. Model Fitting Statistics of the Cholesky Decomposition ACE models compared with the Saturated Models.

`	*-2LL* (Saturated)	*df* (Saturated)	*-2LL* (ACE)	*df* (ACE)	Δ-*2LL* (Δ*df* =17)	*p*
Word reading	2124.35	1137	2139.01	1154	14.66	.62
Receptive vocabulary	2879.78	1138	2891.75	1155	11.97	.80
Phonological memory	2699.90	1138	2711.58	1155	11.68	.82
Tone awareness	2971.24	1138	2983.32	1155	12.08	.80
Syllable and rhyme awareness	2938.73	1137	2959.94	1154	21.21	.22
Rapid automatized naming	2698.82	1116	2708.32	1133	9.50	.92
Morphological awareness	3008.68	1138	3023.20	1155	14.52	.63
Orthographic skills	3063.73	1135	3084.53	1152	20.80	.24

Note. Δ-*2LL* and Δ*df* are the differences between the Cholesky Decomposition ACE and the saturated models.

Table 5.4. Standardized Path Coefficients from a Cholesky Decomposition Model of Genetic (A); Shared-Environment (C); and Nonshared-Environment (E) Influences on Scores across Time Controlling for Age and Nonverbal Ability. (95% Confidence Intervals in Parentheses)

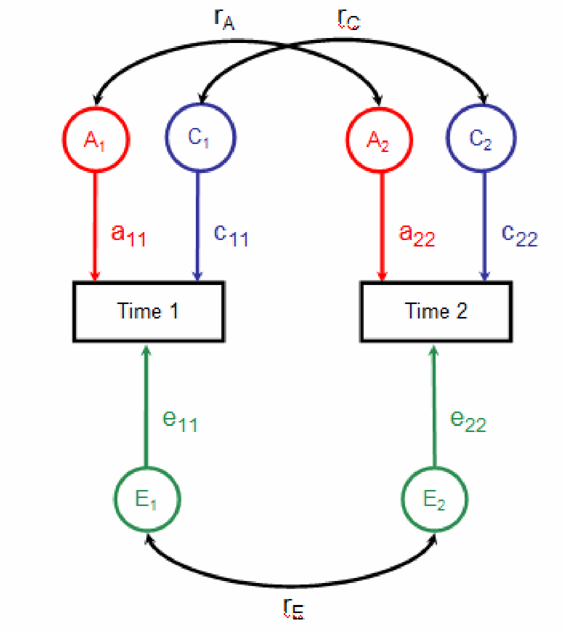
	Genetic factor		Shared environmental factor		Nonshared environmental factor	
	A1	A2	C1	C2	E1	E2
Word reading						
Time1	.84	r_A=.89[b]	.38		.37	
	(.66, 1.03)		(-.02, .79)		(.34, .41)	
Time2	.74[a]	.38[c]	.21[a]	.39[c]	.23	.25
	(.54, .93)	(.26, .51)	(-.32, .74)	(.24, .54)	(.19, .27)	(.23, .28)
Receptive vocabulary						
Time1	.43	r_A= .97	.63		.65	
	(.03, .83)		(.37, .89)		(.59, .71)	
Time2	.62	.15	.32	.32	.17	.60
	(.06, 1.17)	(-1.83, 2.14)	(-.04, .68)	(-.09, .73)	(.09, .25)	(.55, .66)
Phonological memory						
Time1	.78	r_A=.97	.36		.52	
	(.55, 1.00)		(-.10, .83)		(.47, .57)	
Time2	.55	.15	.47	.38	.12	.55
	(.29, .81)	(-.66, .95)	(-.04, .98)	(-.12, .88)	(.04, .19)	(.50, .60)
Tone awareness						
Time1	.72	r_A=.82	.00		.70	
	(.62, .81)		(-1.45, 1.45)		(.64, .76)	
Time2	.58	.40	.00	.00	.20	.68
	(.47, .70)	(.27, .53)	(-.77, .77)	(-.46, .46)	(.11, .29)	(.62, .74)
Syllable and rhyme awareness						
Time1	.33	r_A=1.00	.69	r_C=.78[b]	.65	
	(-.03, .68)		(.51, .87)		(.60, .71)	
Time2	.56	.00	.37	.30	.12	.67
	(.23, .89)	(-1.26, 1.26)	(.09, .65)	(-.10, .70)	(.03, .20)	(.61, .73)
Rapid automatized naming						
Time1	.72	r_A=.98	.29		.64	
	(.44, 1.00)		(-.34, .93)		(.58, .70)	
Time2	.71	.14	.00	.18	.32	.58
	(.38, 1.04)	(-.91, 1.20)	(-1.08, 1.09)	(-.90, 1.26)	(.23, .40)	(.53, .64)
Morphological awareness						
Time1	.51	r_A=1.00	.48		.71	
	(.15, .88)		(.12, .84)		(.65, .78)	
Time2	.52	.00	.22	.44	.17	.68
	(.16, .87)	(-1.02, 1.02)	(-.23, .66)	(.20, .68)	(.08, .26)	(.62, .74)
Orthographic skills						
Time1	.53	r_A=1.00	.46		.71	
	(.27, .79)		(.17, .75)		(.65, .77)	
Time2	.64	.00	.00	.35	.02	.69
	(.36, .92)	(-.61, .61)	(-.47, .47)	(-.14, .84)	(-.06, .11)	(.62, .75)

Note. r_A =genetic correlation; r_C= shared environmental correlation; Superscript a denotes the genetic and shared environmental contributions to the stability; superscript b denotes genetic and shared environmental correlations; superscript c denotes genetic and shared environmental contributions to the instability.

Genetic and environmental overlap. The genetic and environmental overlap was examined in three ways. First, the coefficients of the shared paths connecting A1, C1, and E1 to Time 2 performance revealed the genetic and environmental contributions to the stability (see Table 5.4 superscript a as an example). The coefficients of these genetic shared paths were significantly greater than zero in all measures, indicating significant genetic links between the two time points. However, the shared environmental link was significant for syllable and rhyme awareness only. These results suggested the stability of Chinese language and reading skills across time was mainly influenced by genetic factors, and both genetic and environmental influences played a role in the stability of syllable and rhyme awareness.

Second, among the variables with significant links, genetic and shared environmental correlations were computed to understand to what extent were genetic and environmental influences correlated across time (see Table 5.4 superscript b as an example). Specifically, the genetic or shared environmental correlation indicates the extent to which individual differences on scores of a variable at the two time points reflect the same genetic or shared environmental influences. Figure 5.2 depicts these correlations concerning performance of a variable across two time points.

Figure 5.2. A model on a variable across two time points with genetic correlation (r_A), shared environmental correlation (r_c), and nonshared environmental correlation (r_E) specified.



The genetic correlations on these measures were found to be substantial, ranging from .82 to 1.00. The shared environmental correlation on syllable and rhyme awareness was .78. Hence, 82% to 100% of the genetic influences at Time 1 overlapped with those at Time 2 for various Chinese language and reading skills, while 78% of the shared environmental factors were in common across time for syllable and rhyme awareness. Therefore, even though genetic influences were modest for some skills at each time point, such as syllable and rhyme awareness, they were largely the same genetic factors.

Third, the contributions of genetic and environmental factors to phenotypic correlations between the two time points were estimated. These estimates were derived from the products of the path coefficients from the genetic or environmental factors that loaded on performance at both time points (i.e., a_{11} x a_{21} and c_{11} x c_{21}; see Figure 5.1) divided by the phenotypic correlation. They are shown in Table 5.5. For instance, 79% of the phenotypic correlation between Time 1 and Time 2 word reading was mediated genetically. This estimate was the product of the A paths linking Time 1 and Time 2 word reading (.84 x .74= .62) divided by the phenotypic correlation (r_p=.79). Results indicated that substantial genetic contributions to word reading, phonological memory, tone awareness, rapid automatized naming, and orthographic skills, in which genetic influences mediated around or over two-thirds of the phenotypic correlations. Moderate genetic effects on stability were found on receptive vocabulary, syllable and rhyme awareness, and morphological awareness, in which over one-thirds of their phenotypic correlations were explained by genetic factors. The shared environmental contributions were smaller than the genetic influences for all measures, except syllable and rhyme awareness which 50% of its phenotypic correlation across time was explained by shared environmental factors. Shared environmental contributions were negligible to moderate in other measures (ranging from 0% to 34%).

Table 5.5. Genetic (A), Shared Environmental (C) and Nonshared Environmental (E) Contributions to the Stability of Performance across Time.

	r_p	Contribution of (product of A1, C1 or E1 path estimates)			% of r_p contributed by		
		A	C	E	A	C	E
Word reading	.79	.62	.08	.09	79	10	11
Receptive vocabulary	.58	.27	.20	.11	47	34	19
Phonological memory	.66	.43	.17	.06	65	26	9
Tone awareness	.56	.42	.00	.14	75	0	25
Syllable and rhyme awareness	.52	.18	.26	.08	35	50	15
Rapid automatized naming	.71	.51	.00	.20	72	0	28
Morphological awareness	.50	.27	.11	.12	54	22	24
Orthographic skills	.36	.34	.00	.02	94	0	6

Note. $r_p =$ Phenotypic correlation from the Cholesky decomposition model. r_p of some measures is slightly different from that computed from SPSS due to digit round-up and the correlation coefficients from SPSS were based on a twin in each pair.

Genetic and environmental specificity. The coefficients of the specific paths connecting A1, C1, and E1 to Time 2 performance represented the genetic and environmental contributions to the instability (see Table 5.4 superscript c as an example). New genetic sources specific to Time 2 word reading and tone awareness, and new shared environmental sources specific to Time 2 word reading and morphological awareness, were indicated. One possible explanation would be if new influences started to play a role once children reached a particular age. The sample spanned a wide age range, and so it was possible to obtain evidence from this point with a supplementary analysis, in which the sample was divided into 2 groups: younger group and older group, with mean age as the cut-off (7.64 years). Data of the two groups at time 1 were fitted to the standard univariate ACE model, constraining genetic and environmental parameters to be equal for the younger and older groups (homogenous model), and comparing with another model that allowed different genetic and environmental parameter estimates for the two groups (heterogeneous model). Parameter estimates and model fitting statistics of the age-split analyses are represented in Table 5.6.

Table 5.6. Parameter Estimates and Model Fitting Statistics of Age-Split Analyses.

	a^2	c^2	e^2	Model fitting comparison	$\Delta\chi^2$	Δdf	p
Word reading							
Heterogeneous model				Saturated vs. Heterogeneous	8.06	13	.84
Younger group	.64	.26	.09				
Older group	.77	.06	.17				
Homogeneous model	.64	.26	.09	Homogenous vs Heterogeneous	23.76	3	.00
Tone awareness							
Heterogeneous model				Saturated vs. Heterogeneous	13.43	13	.42
Younger group	.50	.00	.50				
Older group	.51	.00	.49				
Homogeneous model	.50	.00	.50	Homogenous vs Heterogeneous	5.13	3	.16
Morphological awareness							
Heterogeneous model				Saturated vs. Heterogeneous	13.11	13	.44
Younger group	.22	.36	.42				
Older group	.33	.00	.67				
Homogeneous model	.22	.36	.42	Homogenous vs Heterogeneous	26.71	3	.00

Note. There were 153 (105 MZ; 48 DZ) twin pairs in the younger group and 138 (109 MZ; 29 DZ) twin pairs in the older group. a^2 = additive genetic variance; c^2 = shared environmental variance; e^2 = nonshared environmental variance; $\Delta\chi^2$ and Δdf are the differences between the saturated and the heterogeneous models or between the homogenous and the heterogeneous models.

The homogeneous models yielded significantly worse model fit than the heterogeneous models for word reading and morphological awareness (*p*s<.05), indicating significant differences in the two groups. Specifically, greater genetic influences and smaller shared environmental influences were shown in the older group compared to the younger group. However, the homogeneous and the heterogeneous models did not differ significantly on the model fit for tone awareness (*p*>.05). Overall, the results of these age-split analyses further supported the existence of a new source of genetic influences for word reading. However, the results on shared environment were counter-intuitive, because for both word reading and morphological awareness, the c^2 term was larger for the young children than for the older children (in whom it was not significant). Yet the Cholesky decomposition model had indicated a new source of shared environmental influence at Time 2. This pattern of results is hard to explain but could indicate that the new sources of shared environmental influences for word reading and morphological awareness might emerge below 7.64 years of age. If this was the case, the changes would be picked up on the Cholesky decomposition models with the whole sample considered together, but not in the age-split analyses. (Note that a substantial number of children were below 7 years of age at both time 1 and time 2). Evidence on the new genetic factor on tone awareness was mixed, and so warrants further research to confirm the results.

Contributions of genetic and environmental overlap and specificity on Time 2 variance. To understand the extent to which the common and the new factors contribute to the variance of each variable at Time 2 comparatively, the proportion of variance shared between the two time points (A1 or C1 to time 2) and specific to Time 2 (A2 or C2 to time 2) was computed. These proportions were calculated by dividing the squared common path or the squared specific path by the sum of both of these squared paths, and are shown in Table 5.7. Results indicated that genetic factors that accounted for the variances at Time 1 measures contributed to the majority of the genetic variances of their corresponding Time 2 measures, ranging from 68% to all of the total genetic variances. Specifically, the genetic overlap contributed to all genetic variances at Time 2 for syllable and rhyme awareness, morphological awareness and orthographic skills, and over 90% of the genetic variances at Time 2 for receptive vocabulary, phonological memory and rapid automatized naming. Also, the genetic factors that exerted influences on Time 1 word reading and tone awareness accounted for over two-thirds of their genetic variances at Time 2. On the contrary, the contributions of the shared environmental overlap were smaller across measures, compared to those of the common genetic influences, ranging from 0% to 60% of the total shared environmental variances at Time 2. Specifically, the shared environmental influences that accounted for Time 1 receptive vocabulary, phonological memory and syllable and rhyme awareness also accounted for around or over half of their respective total shared environmental variance at Time 2. Furthermore, the common shared environmental factors explained around one-fifths of total shared environmental variances of Time 2 word reading and morphological awareness. However, the specific shared environmental factors accounted for all of the total shared environmental variance in rapid automatized naming and orthographic skills at Time 2.

Table 5.7. Proportions of Total Genetic, Shared Environmental and Nonshared Environmental Variances at Time 2 Accounted for by Overlap and Specificity.

	Genetic (A)			Shared environmental (C)			Nonshared environmental (E)		
	Total variance	Link (%)	Spec (%)	Total variance	Link (%)	Spec (%)	Total variance	Link (%)	Spec (%)
Word reading	.68	79	21	.20	22	78	.12	46	54
Receptive vocabulary	.41	94	6	.20	50	50	.39	7	93
Phonological memory	.33	93	7	.37	60	40	.32	5	95
Tone awareness	.50	68	32	.00	--	--	.50	8	92
Syllable and rhyme awareness	.31	100	0	.23	60	40	.46	3	97
Rapid automatized naming	.53	96	4	.03	0	100	.44	23	77
Morphological awareness	.27	100	0	.24	20	80	.49	6	94
Orthographic skills	.41	100	0	.12	0	100	.47	0	100

Note. Link = proportion of variance shared between the two time points; Spec= proportion of variance specific to Time 2; A/C/E total variance = sum of squared standardized paths of A/C/E common and A/C/E specific factors. A/C/E link = squared standardized A/C/E common path divided by A/C/E total variance. A/C/E spec = squared standardized A/C/E specific path divided by A/C/E total variance.

Nested models. Four nested models with either the path linking Time 1 and Time 2 performance (A1 or C1 to Time 2) or the path representing a new factor at Time 2 (A2 or C2 to Time 2) dropped were fitted to each variable. The model fit of these nested models was compared to that of the full model to further clarify the roles of genes and environments across time. The model fitting statistics of the nested models are shown in Appendix 2C. When the genetic link across time was dropped (see Appendix 2C superscript d), there was a significant decrease in model fit for all variables (*p*s<.05), except for syllable and rhyme awareness. However, the removal of the new genetic path reduced model fit significantly for word reading only (*p*<.05) (see Appendix 2C superscript e). Conversely, when either the shared environmental link across time or the new shared environmental path was dropped, the model fit was not significantly poorer than the full model for all variables (*p*s>.05) (see Appendix 2C superscripts f and g respectively). It was noted that the nested model with the shared environmental link across time removed yielded a marginally significant decrease in model fit for syllable and rhyme awareness (*p*=.07).

Overall, consistent with the results of the Cholesky decomposition models, these findings have suggested genetic factors play a major role in the stability of Chinese language and reading abilities, and contribute to the instability of Chinese word reading. However, while Cholesky decomposition models indicated both genetic and shared environmental factors contributed to the stability of syllable and rhyme awareness, the nested model analyses suggested shared environment might be relatively more important. Also, the new genetic influences on tone awareness and the new shared environment effects on morphological awareness were picked up on Cholesky decomposition models but not the nested model analyses. Thus, further research is needed to confirm the findings.

Discussion

This part of the thesis examined the etiology of the stability and instability in Chinese language and reading skills in 292 Chinese twin pairs. There are two key findings on Chinese language and reading skills. First, results indicated their stability was mainly mediated by genetic influences. However, the stability of syllable and rhyme awareness was influenced by both genetic and environmental factors, with some evidence suggested shared environment exerted stronger influences. Second, findings in this chapter demonstrated new genetic and environmental factors exerted influences on some skills. Specifically, the instability of Chinese reading skills was mediated by both genetic and environmental factors. Also, findings suggested new genetic influences and new shared environmental influences contributed to the changes of tone awareness and morphological awareness at Time 2 respectively, but the results on tone awareness have to be substantiated by further research. Some of these results have extended past research on alphabetic languages and hence suggest universal findings, but unique results in terms of contrastive difference across languages, and skills which have not yet been studied in past longitudinal twin research (i.e., phonological memory, tone awareness, morphological awareness, and orthographic skills), have also been demonstrated in this thesis.

Genetic and environmental overlap

At the phenotypic level, Time 1 and Time 2 performance was moderately to strongly correlated in all language and reading measures, indicating the stability of various facets of language and reading skills across time, which was in line with the prediction. To what extent was the stability explained by genetic and shared environmental factors? Results form genetic analyses showed the stability of Chinese language and reading skills was mainly mediated by genetic influences, and shared environmental mediation was found for syllable and rhyme awareness only. They are consistent with our hypotheses, except the environmental influences on vocabulary skills were not shown.

Word reading. Word reading performance was strongly correlated across time after controlling the effects of age and nonverbal reasoning, indicating its stability across time. Results showed that the majority of this phenotypic stability was mediated by genetic factors (79% for genetic factors compared to 10% for shared environment), and strong overlap of these genetic factors across the two time points (89%). These results suggested genetic influences played a prominent role in word reading stability, and most of these genetic factors which exerted effects at Time 1 also had impacts at Time 2. The genetic mediation in word reading consistency was demonstrated across different methods of analyses, and so the evidence was robust in this thesis.

These findings are in line with past longitudinal studies which found genetic contributions to word reading consistency across time in samples of different ages, from preschoolers to teenagers (e.g., Betjemann et al., 2008; Byrne et al., 2005; Harlaar et al., 2007a; Petrill et al., 2007; Samuelsson et al., 2008). Apart from the various samples' age ranges, these studies were conducted in different countries, including Australia, Scandinavia, the United Kingdom, and the United States, and thus the genetic contributions might not only be robust across ages, but also universal across countries. Samuelsson et al. (2008) compared children in three of these countries, including Australia, Scandinavia, and the United States, and found significant genetic influences on the reading consistency in all three countries, and thus supported the notion of universal genetic contributions. This thesis has extended the evidence to Chinese reading, and has further strengthened the essential role of genetic factors in the reading performance stability across time.

Compared to the genetic influences, shared environmental factors had a diminished role in word reading stability. This thesis found nonsigificant shared environmental links between Time 1 and Time 2 word reading. It was in contrast to the substantial shared environmental overlap on reading ability shown in Harlaar et al. (2007b). Nevertheless, reading skills were assessed by teachers' ratings, and twins in the same classroom were even assessed by the same teacher in their research, and thus the shared environmental link might be inflated. Also, a significant shared environmental link between preschool print knowledge and kindergarten word reading skills was indicated in Byrne et al. (2005). Though print knowledge was a pre-requisite of word reading skills, its link with word reading might differ from that of word reading itself across time, and so this finding could not be directly compared with the results of this thesis. In general, longitudinal studies focused on word reading skills assessed by psychometric tests failed to show shared environmental contributions (e.g., Harlaar et al., 2007a; Petrill et al., 2007), and so the results of this thesis chime with these past studies.

Vocabulary and rapid automatized naming. Performance at the two time points was moderately and strongly correlated for receptive vocabulary knowledge and rapid automatized naming respectively, when age and nonverbal reasoning was accounted for. Both of them showed genetic influences on their stability, but nonsignificant contribution of shared environment to their continuity. Specifically, genetic influences contributed to a greater portion of their phenotypic correlations than shared environmental factors, in which 47% and 72% of the phenotypic correlations were mediated by genetic factors in receptive vocabulary and rapid automatized naming respectively (compared to 34% and 0% explained by shared environmental influences for receptive vocabulary and rapid automatized naming respectively). Also, almost all of their genetic variances overlapped at Time 1 and Time 2 (97% for receptive vocabulary and 98% for rapid automatized naming), indicating the majority of genetic factors exerted effects at Time 1 continuously influenced the performance at Time 2.

Past twin studies on children’s vocabulary skills focused on expressive vocabulary, and they found that both genetic and shared environmental influences mediated its stability across time (e.g., Byrne et al., 2009; Hart, Petrill, DeThorne, et al., 2009; Petrill et al, 2007). Therefore, this thesis has extended the evidence on the prominence of genetic factors on expressive vocabulary stability to receptive vocabulary knowledge. However, the impacts of shared environment showed a different picture. One possibility is that shared environment might have differential effects on various facets of vocabulary skills. Thus, it contributes to the continuity of vocabulary knowledge at the expressive level only, but not the receptive level. The other possibility is that it might operate in different ways across languages, and thus mediates the stability of vocabulary skills in alphabetic languages, but not in Chinese. Nevertheless, any conclusions of this kind might be premature at this stage, and further longitudinal twin research on receptive vocabulary skills in alphabetic languages and expressive vocabulary in Chinese could clarify the mechanisms.

Though results of this thesis and previous twin research (Byrne et al., 2005) showed genetic influences contributed to rapid automatized naming consistency, Petrill et al. (2007)

did not demonstrate this link. As Petrill et al. (2007) noted, the contrastive findings might be a consequence of different rapid automatized naming measures employed, but how these different domains affect the genetic contributions remains unclear. This thesis has shed light on the underlying mechanisms. This thesis which included digit naming only indicated a greater genetic link path estimate than Byrne et al. (2005) which tested colour and letter naming as well (.72 compared to .59; though it should be noted that their confidence intervals overlapped owing to the relatively small sample size for longitudinal genetic analyses in both studies), and Petrill et al. (2007) focused on number and letter naming found nonsignificant genetic contributions. Hence, these findings imply that rapid automatized naming measurement which involved more reading-related aspects, such as letter naming, could reduce the effects of genetic influences demonstrated, because these reading-related aspects rely more on explicit instruction from parents and teachers.

Despite the mixed evidence on genetic contributions, consistent evidence on the lack of shared environmental mediation in its stability has been obtained in this thesis and past twin research (Byrne et al., 2005; Petrill et al., 2007). Given the limited research evidence on rapid automatized naming, more research would be helpful in understanding its developmental etiology.

Phonological awareness. Phonological awareness at the syllable and rhyme level was moderately linked across time (.51). While genetic factors exerted relatively greater influences on the stability of most of the language and reading skills than shared environmental factors, phonological awareness at the syllable and rhyme levels showed a different picture. Results of the Cholesky decomposition models showed both genetic and environmental contributions, but the nested model indicated a marginally poorer fit when shared environmental link was dropped and nonsignificant change of model fit when genetic link was removed. They chimed with the results of greater contribution of shared environmental factors to the phenotypic correlations across time (50% for shared environment compared to 35% for genetic factors). Also, the genetic factors completely overlapped across time, showing that all genetic influences at Time 2 were in common to those at Time 1, and there was no new genetic factor unique to Time 2. Regarding shared environment, 78% of its influences overlapped across time. Overall, the stability of syllable and rhyme awareness was influenced by both genetic and environmental factors, with some evidence suggested shared environment exerted stronger influences. Past twin studies have shown the genetic mediation in the stability of phonological awareness, but evidence on the shared environmental mediation has been mixed (significant shared environmental link in Petrill et al., 2007; nonsignificant shared environmental link in Byrne et al., 2005; Byrne et al., 2006). Therefore, genetic factors played an important role in the continuity of different facets of phonological awareness, including rhyme and phoneme in alphabetic languages, and syllable and rhyme in Chinese, but their manifestation relative to that of shared environment could vary when different facets of phonological awareness or languages were considered. The environmental variance in phonological awareness development in Chinese tends to be greater than those in English, given its ambiguous print-sound correspondence and geographical diversity in use of the different instruction methods (e.g., Pinyin in mainland China, Zhu Yin Fu Hao in Taiwan, and lack of systematic teaching on Cantonese phonetics in Hong Kong), highlighting the relative importance of shared environment in Chinese phonological awareness.

Other language and reading skills. Moderate correlations in performance across time were demonstrated in phonological memory, tone awareness, morphological awareness, and orthographic skills, after controlling the effects of age and nonverbal reasoning. The majority of their phenotypic stability was mediated by genetic factors (65%, 75%, 54% and 94% for phonological memory, tone awareness, morphological awareness, and orthographic skills respectively), and there was strong overlap of these genetic factors across the two time points (97%, 82%, 100% and 100% for phonological memory, tone awareness, morphological awareness, and orthographic skills respectively). These results indicated the importance of genetic contributions in stability and most of these genetic factors which exerted effects at Time 1 maintained their impacts at Time 2. Genetic factors completely overlapped across time for morphological awareness and orthographic skills, indicating no new sources of genetic influences emerged at Time 2. These findings have shed light on the developmental etiology of skills which have not yet been examined in past longitudinal twin studies, and thus broaden our understanding of language and reading development.

Mechanism of genetic and shared environmental overlap. In general, the important role of genetic factors on the consistency of various language and reading skills has been underscored by this thesis and past research. Genetic factors might exert effects on language and reading skills in two ways. First, the same sets of genes continue to express in different ages or developmental stages, and thus they have direct impacts on language and reading skills across age. Second, genes expressed at an early age have impacts on later language and reading skills through indirect pathways, such as early-expressed genes determined neural network changes, and these changes could influence language and reading learning abilities at a later age.

Overall, shared environment played a less important role than genetic factors on language and reading skills. It contributed to expressive vocabulary as shown in past studies, and to phonological awareness as demonstrated in this thesis and some previous research only. Some research has identified environments, including family socio-economic status and chaos, which contributed to the stability of expressive vocabulary and grammatical complexity over the genetic influences (Petrill et al., 2004).

Genetic and environmental specificity

Despite the moderate to strong phenotypic correlations of performance across time, they were not unity and thus some degree of instability was shown. So, to what extend do new genetic or shared environmental factors contribute to changes in performance? This thesis showed that new genetic and/or shared environmental factors contributed to changes in performance of word reading, tone awareness, and morphological awareness across time, but results on tone awareness need further research to confirm. However, this thesis did not find new sources of genetic or shared environmental influences on the other language and reading skills, including receptive vocabulary, phonological memory, syllable and rhyme awareness, rapid automatized naming, and orthographic skills. They are partly in line with our hypotheses, as some findings are different from the predictions, such as new sources of influences were not indicated in phonological awareness at the syllable and rhyme levels, and new genetic factors were not found in rapid automatized naming.

Word reading. This thesis found that both specific genetic and shared environmental factors exerted effects on word reading. Findings of the Cholesky decomposition model and the age-split analysis supported the effects of these new factors on the instability of word reading. Results showed that the specific genetic factors which were unique at Time 2 explained 21% of the total genetic variance in Time 2 reading, despite the relatively large contribution of the genetic link to the total genetic variance (79%). In other words, while genetic influences contributed to the stability of word reading across time, new genetic factors exerted smaller but significant effects which yielded changes in word reading ability. The specific genetic influences have been demonstrated in some twin studies (e.g., Petrill et al., 2007; Harlaar et al., 2007b), but not in others (e.g., Byrne et al., 2005; Harlaar et al., 2007a). Also, Samuelsson et al. (2008) indicated significant specific genetic effects in the United States sample only, but not among the samples in Australia and Scandinavia. Overall, the genetic effects on reading instability have been demonstrated, but their role in reading instability is less clear than on reading stability.

On the other hand, this thesis showed some evidence on shared environmental contribution to the change of word reading performance across time, which was in line with Petrill et al. (2007). Specifically, results showed that the specific shared environmental factors at Time 2 explained the majority (78%) of the total shared environmental variance in Time 2 reading (compared to 22% for shared environmental link), highlighting the fact that different shared environmental factors influenced word reading skills at the two time points. However, other studies did not find new shared environmental factors on reading ability (e.g., Byrne et al. 2005; Harlaar et al., 2007b; Samuelsson et al., 2008). It is interesting to note that new genetic and shared environmental influences were indicated in the United States samples only in previous twin studies, except in Harlaar et al. (2007b) which tested the United Kingdom sample. However, Harlaar et al. (2007b) examined the link of teacher-rated reading performance at 7, 9 and 10 years of age, and thus the new sources of influences might only be the result of different teachers taking part in the assessments. Therefore, emerging evidence suggests the impacts of new genetic and shared environmental factors contributing to the instability of word reading might be country-specific and highly depend on the country's policy and curriculum of literacy instruction. This thesis has extended the evidence on new genetic and shared environmental influences to Chinese children learning to read Chinese. Furthermore, the age-split analysis in this thesis suggested, if shared environmental influences existed, they might emerge below 7.64 year of age. It agreed with previous research findings, as Petrill et al. (2007) showed new shared environmental factors emerged between around age 6 to 7.

Tone awareness and morphological awareness. This thesis found that specific genetic and specific shared environmental factors contributed to the instability of tone awareness and morphological awareness respectively. Regarding tone awareness, the new genetic factors explained 32% of its total genetic variance at Time 2, in spite of the comparatively large contribution of the genetic link (68%). In other words, while genetic factors at Time 1 continued their influences at Time 2, some new genetic influences emerged and contributed to the change in tone awareness. However, the supplementary age-split analysis did not confirm these Cholesky decomposition model findings, and thus warrants further research to substantiate the findings.

On the contrary, morphological awareness did not show specific genetic influences, but the new shared environmental factors accounted for 80% of its total shared environmental variance at Time 2. Hence, different sets of shared environmental factors influenced morphological awareness across time. These results were strengthened by the age-split analysis, and suggested the new sources of shared environmental influences might emerge under 7.64 years of age, as with what was found on word reading.

Mechanism of genetic and shared environmental specificity. During the course of language and reading development, the cognitive demands vary quantitatively and qualitatively. As proficiency increases, some cognitive processes may have a diminished role, whereas some qualitatively different processes come into play (Stuart & Coltheart, 1988). Children not only have to polish up the skills that they have mastered, but also acquire more advanced skills. For instance, a child who has mastered a certain level of word reading ability has to keep learning more words, and at the same time develops skills on comprehending sentences or even passages to become a proficient reader. Reading comprehension is closely linked to word reading ability (Verhoeven & van Leeuwe, 2008), but demands more sophisticated skills, such as text integration skills and metacognitive monitoring (Oakhill, Cain & Bryant, 2003). Thus, the genetic and environmental factors which link to these more sophisticated skills, which might not have an impact on early-stage reading, start to influence reading skills later on. Similar processes might happen in the development of word reading itself. At the initial stage of Chinese word reading, characters are encoded as a holistic unit and learning relies on rote memorization (Anderson et al., 2003). Children then begin to aware of the phonemic principles and are able to make use of the phonetic components to read characters, which happens when they have more experience in reading and have mastered several hundred characters (Chen et al., 2003). However, at this stage, the utilization of the phonemic principles is limited to reading regular characters (i.e., characters with phonetic radicals which contain full information about pronunciation, including onset, rime and tone), but then more sophisticated skills are progressively acquired, and they are able to make use of partial phonetic information (i.e., characters which have a onset-, rime-, or tone- different phonetic radical) to learn and read characters (Anderson et al, 2003). Research found children at around age 9 might still struggle to fully understand the phonemic principles (Chan & Nunes, 1998). The emergence of these awareness and skills might be linked to new genetic and environmental factors which come to play in word reading later on. This thesis did not show new genetic or environmental factors on syllable and rhyme awareness. These findings underscore a possibility that the understanding of the phonemic principles might influence reading skills qualitatively, but not more general phonological awareness at the syllable and rhyme levels.

The new sources of shared environmental influences on morphological awareness could be attributable to the explicit teaching of morphological skills. This usually happens when the demand on word reading increases, and the child needs to progress from relying solely on rote memorization to utilizing metalinguistics skills for better learning. While phonetic information as discussed before is helpful, morphology also provides useful aids for supporting the acquisition and retention of words (Sandra, 1994). This applies particularly well to Chinese, which involves extensive lexical compounding, and semantic radicals as character components provide hints on the character meanings. A whole-character approach is employed in Chinese reading instruction in Hong Kong, but semantic radicals are explicitly taught in school, mainly as a reference unit for looking up Chinese characters in dictionaries after the child has mastered the basic level of word reading (Ho et al., 2003). Several studies have established the effects of morphology training on children's word reading ability (Fu & Huang, 2000; Nagy et al., 2002).

Other skills did not show new sources of genetic and shared environmental effects. As our study examined language and reading growth in a one-year period, it may be long enough for picking up new genetic and environmental influences. Also, it could speculate that the development of these skills might not involve qualitative changes in cognitive demands, and thus the same sets of genetic and shared environmental factors continue to play a dominant role. However, it is not necessary. The lack of new sources of influences might also reflect that there are qualitative changes of cognitive demands, but the same sets of influences contribute to these new demands as well as those required in earlier stages.

Limitations

When interpreting the results of this thesis, it is important to note that the sample spanned a relatively wide age range (from age 3 to 11). Changes in some subjects might be greater than the others, owing to the comparatively more contrastive learning environment across the two time points. For instance, children who promoted from kindergarten grade 3 to primary school grade 1 might encounter greater changes in instruction, as more intensive and formal language and reading curriculum was introduced in primary school. Thus, these findings indicated the general patterns of individual variations across time, rather than the changes at particular points in the language and reading development. Also, another potential problem regarding the wide age range is that the age ranges overlapped substantially across the one-year period (i.e., 3 to 11 years at Time 1 and 4 to 12 years at Time 2), and thus interpretation on the longitudinal findings should be cautious. Therefore, further research could aim to reducing the sample's age range and focusing on particular reading stages, such as the developmental changes from being taught literacy informally to formally, and from learning to read to reading to learn. In addition, the sample size was relatively small for longitudinal genetic analyses, and thus some significant effects might not be detected. So, cautious interpretations of nonsignificant results should be taken, and further studies are warranted to confirm the findings.

The twin design employed in this thesis nicely separated genetic and environmental effects and so the relative roles of genetic and environment on the stability and instability could be clearly shown. However, another potential source of stability and instability, the genotype-environment correlation, has not been addressed. For instance, a child who is genetically at risk of language difficulty, may also has impoverished language learning environment at home (passive genotype-environment correlation), evokes less conversation with his or her peers (evocative), and is less interested in seeking out language and reading learning opportunities (active). These genetic and environmental factors could jointly contribute to the consistently poorer language ability of the child across time. This issue has been discussed more thoroughly in Chapter 4.

Implications and conclusions

To conclude, this thesis has extended past twin research on the developmental etiology of language and reading abilities in alphabetic languages to Chinese. There are several important implications. First, this thesis has indicated the important role of genetic factors not only in the developmental stability of all Chinese language and reading skills examined, but also in the instability of Chinese word reading and tone awareness. On one hand, the genetic mediation on the consistency suggests early identification of at-risk children could be possible by targeting at parents who have problems with language or reading themselves. On the other hand, the genetic contributions to the instability imply why some children who have typical language and reading development in early stages begin to show difficulties at some point. These findings are also interesting to molecular genetics studies by suggesting the target language and reading skills for further investigation in the Chinese populations.

Second, environmental factors had significant contributions to the stability of phonological awareness at the syllable and rhyme levels, and the instability of word reading and morphological awareness. These results suggest early training on phonological awareness could be particularly useful, and highlight the fact that children could be sensitive to different instruction methods which aid word reading and morphological awareness learning in the course of reading development. Some evidence has shown that the new sources of environmental influences might emerge under 7.64 years of age for both word reading and morphological awareness. It is promising to identify these environmental factors. These findings provide insights to educators and parents on structuring effective Chinese language and literacy learning environment for children.

Third, by comparing this thesis and past twin research on alphabetic languages, the universal or specific factors of language and reading acquisition across languages could be better understood. Some of these results have extended past research on alphabetic languages and hence suggest universal findings. For instance, results of this thesis have added to strong research evidence which supported the genetic contributions to the stability of word reading.

Besides, there were genetic influences on phonological awareness and nonsigificant shared environmental contribution to rapid automatized naming. This thesis has also demonstrated novel findings. Specifically, this thesis has extended the findings of the genetic influences on expressive vocabulary to receptive vocabulary, and has suggested new sources of genetic and environmental influences on word reading could be country-specific. Also, this thesis informs further longitudinal research by demonstrating the genetic influences on the stability of phonological memory, tone awareness, morphological awareness and orthographic skills, the potential genetic contributions to the instability of tone awareness, and the shared environmental mediation of the instability of morphological awareness, in which the developmental etiology of these skills has not been examined in past longitudinal twin studies. Models on language and reading development across languages should take special account of these universal and specific factors contributing to their development.

Chapter 6

Parent-rated communicative ability and motivation for learning text: Do genetic factors contribute to their individual differences and links with reading abilities?

Overview

This part of thesis investigates the genetic and environmental contributions to parent-rated communicative ability and motivation for learning text and their links with reading skills in Chinese twin pairs aged 3 to 11. Parents' ratings on twenty positive items of Children's Communication Checklist – Second Edition (CCC-2; Bishop, 2003) and four items on motivation for learning text were obtained for 307 and 306 twin pairs respectively.

Results indicated significant genetic and shared environmental influences for both communicative ability and motivation for learning text. Specifically, there were stronger shared environmental influences than genetic effects on communicative ability and comparable genetic and environmental contributions to motivation. Also, genetic factors contributed to the link between motivation and word reading. This thesis has enhanced our understanding of the genetic perspective on communicative ability and motivation evaluated in naturalistic contexts. In addition, it has extended the genetic perspective on reading and related cognitive skills to reading-related behaviours and beliefs.

Research Background

Children's language and reading development has been better understood with the help of a variety of assessments in clinical and laboratory settings. While these assessments provide direct and objective means to tap children's language and reading levels, they are less sensitive to the language used in naturalistic contexts and thus are not optimal methods on their own (Feldman et al., 2005). In unfamiliar clinical and laboratory settings, children often behave differently than they do in daily life and assessment sessions have to be short in consideration of children's attention span (Sachse & Von Suchodoletz, 2008). Also, some skills are particularly difficult to be tested by psychometric tests in these settings, such as pragmatic skills (Bishop, 1998). For these reasons, skills which are highly dependent on the situations that the children encounter and cover broad aspects of abilities, such as communicative skills, are particularly hard to be appropriately evaluated by psychometric measures in these restricted settings. In a similar vein, children's motivation is difficult to be accurately reflected in laboratory contexts. As parents have extensive experience with their children in diverse contexts, they are the best informants to evaluate children's abilities and behaviours, and thus parental reports are useful means to assess these dynamic constructs (Feldman et al., 2005). In past research, children's communicative ability and motivation for reading have been successfully assessed by parents (e.g. Norbury, Nash, Baird, & Bishop, 2004; Zhou & Salili, 2008).

To date, the genetic correlates of language and reading skills of typically developing children have been examined using psychometric measures (e.g., Byrne et al., 2009; Hart, Petrill, Thompson, et al., 2009), though some studies employed parents' or teachers' reports (e.g., Dionne et al., 2003; Harlaar et al., 2007b). However, very little is known about the genetic correlates of communicative ability and motivation observed in naturalistic situations. Therefore, the first aim of this part of the thesis is to investigate the genetic and environmental influences on parent-rated communicative ability and motivation with a twin study design. In addition, robust research evidence has supported the links between children's motivation and reading skills (e.g., Wang & Guthrie, 2004), and some findings have suggested the relationships between communicative ability and reading skills (e.g., Wooster & Carson, 1982). The study of the link between communicative ability and reading skills is also motivated by the independent etiology of Chinese language and reading abilities demonstrated (see Chapter 3). It is interesting to know if this independency extends to communicative ability observed in naturalistic situations. So, the second aim of this part of the thesis is to explore the genetic and environmental contributions to the links between a) communicative ability and reading abilities; and b) motivation and reading abilities.

Communicative ability

Effective communication requires adequate linguistics skills and sensitivity to the others' needs and characteristics (Genesee, Tucker, & Lambert, 1975). To use and interpret language in various contexts appropriately, knowledge and awareness of language structure, such as vocabulary, syntax and morphology, as well as pragmatics, are important (Norbury et al., 2004). These highlight adequate knowledge of language structure is essential but not sufficient for good communication, because if individuals lack of pragmatic skills and are unable to use language in different ways in response to a variety of situations, they could encounter communication difficulties. These challenges are evident in pragmatic language impairment (PLI). Though standardized language tests help understand children's knowledge and awareness of language structure, they are less sensitive to children's pragmatic skills, and thus are often not effective in detecting PLI children who encounter communicative difficulties in daily life.

The Children's Communication Checklist (CCC; Bishop, 1998) was developed to assess aspects of communicative impairment that could not be accurately identified by standardized language tests. Its revised version, the Children's Communication Checklist-2 (CCC-2; Bishop, 2003) is a 70-item checklist for evaluating children's communicative behaviours with a format most suited for parents. The CCC-2 has a total of 10 subscales, in which four evaluate pragmatic skills (inappropriate initiation, stereotyped language, use of context and non-verbal communication), the other four assess a wider range of language functions other than pragmatics (speech, syntax, semantics and coherence), and the remaining two reflect autistic-type behaviours (social relations, and interests). The subscales of the CCC-2 are mostly consistent with the CCC, except a semantics subscale was added, the syntactic subscale was enhanced, and the coherence subscale was recatergorized in the CCC-2. So, the CCC-2 assesses a wider range of language functions apart from pragmatics. Each subscale has two positive items showing communication strengths and five negative items showing weaknesses. The general communication composite of the CCC-2 (i.e., the sum of the first eight subscales listed above) can effectively discriminate communication impaired children from their typically developing counterparts, and its social-interaction deviance composite score (i.e., the sum of speech, syntax, semantics and coherence subscales minus the sum of inappropriate initiation, non-verbal communication, social relations and interests subscales) can identify children who perform in the normal range of language measures but show impairment on pragmatic skills (Norbury et al., 2004). The utility of both the CCC and the CCC-2 in identifying children with communication impairment have been well documented in different populations, including English (e.g., Norbury et al., 2004), Dutch (e.g., Ketelaars, Cuperus, van Daal, Jansonius, & Verhoeven, 2009), and Norwegian children (e.g., Helland, Biringer, Helland, & Heimann, 2009). Apart from this, the CCC has recently been successfully employed to study the development of communicative ability in typically developing Finnish-speaking children (Yliherva, Loukusa, Väisänen, Pyper, & Moilanen, 2009).

So far, the genetic influences on communication skills assessed by the CCC or the CCC-2 have been explored in a twin study only (Bishop, Laws, Adams, & Norbury, 2006). Bishop et al. (2006) found high heritability and negligible environmental influences on most of the CCC subscales among 196 six-year-old twin pairs (130 pairs with at least one twin in the pair had low language skills at four years of age and 66 pairs with no indication of language difficulties at four years of age). Three subscales, namely speech, syntax, and coherence, discriminated well between at-risk and low-risk children of language impairment, and they were highly influenced by genetic factors. These findings chimed with past studies which demonstrated strong genetic influences on specific language impairment based on psychometric testing (e.g., Bishop, 2002). This thesis extends this line of research in two major ways, particularly by a) assessing communicative ability of typically developing Chinese children with positive items of CCC-2 (positive items are used given the emphasis on communication strength in typically developing children); and b) examining genetic and environmental contributions to parent-rated communicative ability and its link with reading abilities.

Motivation for learning written text

Learning to read requires effort and persistence, and so motivation is key to reading engagement and success (Wigfield, Guthrie, Tonks, & Perencevich, 2004). In general, motivation for reading is multifaceted which involves self-efficacy belief, purposes for learning to read (intrinsic and extrinsic motivation, and goal orientations), and social goals (Wigfield & Guthrie, 1997). Among these dimensions of reading motivation, self-efficacy, intrinsic motivation, and extrinsic motivation are often the main foci of reading research (e.g., Wigfield et al., 2004). Reading self-efficacy is the belief in one's own capabilities to carry out reading tasks; intrinsic motivation refers to performing reading tasks out of genuine interest; external motivation refers to completing reading tasks for external rewards (Wigfield et al., 2004). Robust evidence has shown self-efficacy (e.g., Shell, Colvin, & Bruning, 1995; Shell, Murphy, & Bruning, 1989) and intrinsic motivation (e.g., Law, 2009; Wang & Guthrie, 2004; Wigfield & Guthrie, 1997) were positively linked to reading engagement and achievement, but the contribution of extrinsic motivation was less clear (positive links, e.g., Wigfield & Guthrie, 1997; negative links, e.g., Law, 2009; Wang & Guthrie, 2004). So, the self-efficacy and intrinsic motivation aspects are focused in this thesis.

Learning environments play important roles in motivation (e.g., Baker, Scher, & Mackler, 1997; Ginsburg & Bronstein, 1993). For instance, parental model of reading behaviour, number of books and years of character teaching at home, were linked to children's intrinsic motivation (Zhou & Salili, 2008). Also, instructional methods influence reading motivation. Specifically, the usage of stimulating tasks in classroom fostered reading motivation and achievement (Guthrie et al., 2006), and Concept Oriented Reading Instruction enhanced children's intrinsic motivation for reading and reading self-efficacy, but Strategy Instruction did not (Wigfield et al., 2004). The environments studied have been extended from home and school environments to broader contexts, such as culture. While some studies found the links between motivation and reading achievement could vary across cultures (Chiu &

Chow, in press), others showed these links were common across children in different cultures, particularly the American and the Chinese children (Wang & Guthrie, 2004). Therefore, different levels of environments might have differential impacts on motivation.

To date, the impacts of various environments have been examined, but the effects of genetic factors on children's reading motivation have not yet been investigated. One twin study has examined children's reading exposure, which often has a close relationship with reading motivation. In this study, Harlaar et al. (2007a) found modest genetic influences and moderate shared environmental effects on reading exposure at 10 years of age, and the majority of these genetic and shared environmental variances overlapped with word reading at 7 years. After the genetic and environmental links between 7-year word reading and 10-year reading exposure were controlled, shared environmental factors mediated the link between 10-year reading exposure and 12-year word reading. The authors concluded that genetic and environmental influences on reading ability at early age contributed to later tendency to seek out reading experiences which in turn determined further reading ability through environmental processes. This line of research takes into account of the genetic perspective on behavioural and psychological aspects of reading other than the cognitive level. This is a promising area of research as past twin studies have emphasized the cognitive aspects of reading, such as phonological awareness and rapid automatized naming (e.g., Samuelsson et al., 2005). Given very little is known in this area and the importance of motivation in reading development, this thesis explores the genetic and environmental contributions to parent-rated motivation and its link with reading abilities.

Research questions

Overall, this part of the thesis investigates children's communicative ability and motivation for learning text rated by parents and their links with reading abilities assessed by psychometric tests in 307 Chinese twin pairs. There are three research questions. First, how do heredity and environment contribute to children's communicative ability and motivation? It is hypothesized that communicative ability in Chinese children is influenced by genetic factors, as evidenced in English children (Bishop et al., 2006). Also, given the important roles of learning environments in children's motivation, and the relatively weaker genetic than shared environmental contributions to reading exposure, it is predicted shared environment will have an impact on motivation. Second, what are the phenotypic relationships between communicative ability and motivation, and reading abilities in Chinese? Word reading ability, and more general reading skills which constitute of word reading and related cognitive abilities (i.e., General Reading, see chapter 3), are examined. It is expected that communicative ability and motivation will significantly associate with reading skills even after the effects of age and nonverbal reasoning on reading abilities are accounted for. Lastly, how do heredity and environment contribute to these links? This thesis is the first to explore this research question, and thus the predictions are left open. For motivation, the expectation leans towards considerable genetic overlap between motivation and reading skills, based on the findings of reading exposure (Harlaar et al., 2007a). It should be noted that the same informant evaluated both twins in each pair on their communicative ability and motivation, which might yield higher shared environmental estimates in the genetic analyses, and this should be taken into consideration when interpreting the results.

Method

Children’s communicative ability and motivation for learning text were obtained by parent-reported questionnaire. Ratings on communicative ability and motivation were obtained from 307 and 306 families respectively. The scores of word reading and General Reading were included in the following analyses (see Chapter 2 for participant and word reading task details).

Measures of communicative ability and motivation for learning text

Communicative ability. Parental report of children’s communicative ability was obtained from 20 positive items adapted from the Children’s Communication Checklist - Second Edition (CCC-2; Bishop, 2003). These items were translated into Chinese by Wong (2007) and each of them was a statement describing an aspect of children’s communicative behaviour. Parents were asked to judge whether the child have already or have not yet acquired the behaviour on a 4-point Likert scale (0 = not yet acquired; 3 = fully acquired). A composite score indicated the child’s general communicative ability. The maximum score was 60 and its Cronbach’s α was .92.

Motivation for learning text. Parental report of children’s motivation in learning written Chinese was obtained from four statements, including the child 1) likes reading Chinese books; 2) likes learning Chinese; 3) believes learning Chinese is easy; 4) is confident about learning Chinese. Parents were asked to judge to what extent they agreed or disagreed each of these statements on a 7-point Likert scale (1 represented totally disagree; 7 represented totally agree). The two former statements are related to the intrinsic motivation aspect whereas the two latter statements are related to the self-efficacy aspect of motivation.

These four statements were considered together and represented more general motivation in this thesis. A composite score indicated the child's general motivation for learning written Chinese. The maximum score was 28 and its Cronbach's α was .92.

Results

Table 6.1 shows the descriptive statistics on the raw scores of communicative ability and motivation.

Table 6.1. Descriptive Statistics on Communicative Ability and Motivation Raw Scores.

	N	*M*	*S.D.*	z_{skewness}	z_{kurtosis}	*r* with age
Communicative ability	614	42.93	10.36	-5.49**	-1.43	.21**
Motivation	612	19.57	5.64	-5.34**	-1.55	.02

Note. ***p*<.01. *N* for correlation with age is 307 and 306 for communicative ability and motivation respectively; *z* score with an absolute value exceeding 1.96 indicates significant skewness or kurtosis.

Both communicative ability and motivation were negatively skewed (i.e., scores were clustered at the higher end of the distribution), but had satisfactory kurtosis. Standardized scores were employed in all analyses. Correlations of age, and communicative ability and motivation were computed from scores of one cotwin randomly selected in each pair. Motivation was not correlated with age (r= .02, p> .05), and thus its standardized scores were not controlled for age. Communicative ability was significantly correlated with age (r= .21, p< .05). For all communicative, language and reading measures, their standardized scores were adjusted for age as well as nonverbal reasoning. Data were then normalized to reduce the influences of skewness and extreme cases in model fitting. Specifically, the scores were transformed into new standardized scores based on their cumulative frequencies for each variable.

Next, the scores of MZ and DZ twins were compared. Descriptive statistics of communicative ability and motivation by zygosity are shown in Table 6.2.

Table 6.2. Descriptive Statistics on Communicative Ability and Motivation by Zygosity.

	MZ			DZ		
	N	*M*	*S.D.*	*N*	*M*	*S.D.*
Communicative ability (age- and nonverbal reasoning- adjusted)	447	-0.12	0.98	166	0.34	0.95
Motivation	446	-0.02	0.99	166	0.07	1.03

Note. A child failed to complete the nonverbal reasoning measure.

Results of independent sample T-tests on the scores of a cotwin randomly selected in each twin pair showed MZ twins and DZ twins did not differ significantly on both communicative ability ($t[304]= -1.02$, $p>.05$) and motivation ($t[305]= .-02$, $p>.05$). There was a trend of higher ratings on communicative ability of DZ twins than MZ twins, though the contrast was not significant. To minimize the possible effects of this difference on model fitting, the communicative ability scores were standardized by zygosity. Communicative ability and motivation significantly correlated with word reading and General Reading (see Table 6.3; r ranged from .12 to .17, ps<.05). Though these correlations were significant, they were modest and the variances of overlap ranged from 1% to 3%.

Table 6.3. Correlations between Communicative Ability and Motivation, and Reading Skills on Scores of a Cotwin Randomly Selected in Each Twin Pair.

Variables	Communicative ability (age- and nonverbal reasoning-adjusted)		Motivation	
	N	*r*	*N*	*r*
Word reading	307	.12*	306	.13**
General Reading	290	.17**	289	.16**

Note. **p*<.05; ***p*<.01.

Genetic analyses

The communicative ability and motivation scores were first fitted to the full ACE model, and then nested AE and CE models, and also the Cholesky decomposition models, by OpenMx in R statistical modeling package (Braun & Murdoch, 2007). The correlations of cotwins' scores by zygosity are shown in Table 6.4.

Univariate genetic analyses

Results indicated that all models provided a satisfactory goodness-of-fit (*p*s> .05). Significant genetic and shared environmental influences were found for both communicative ability and motivation. Specifically, strong shared environmental effects were indicated in communicative ability, explaining 77% of its variance. Also, comparable genetic and shared environmental influences were shown in motivation, each explained nearly half of the variances in motivation. The parameter estimates are shown in Table 6.4.

Table 6.4. Twin Correlations by Zygosity and Genetic Model Parameter Estimates of Communicative Ability and Motivation. (95% Confidence Intervals in Parentheses)

	Twin Correlations		ACE Models					
Variable	MZ	DZ	a^2	c^2	e^2	χ^2 (df=6)	*p*	AIC
Communicative ability (age- and nonverbal reasoning- adjusted)	.95	.86	.18 (.12, .24)	.77 (.64, .89)	.05 (.05, .06)	0.54	1.00	-92.56
Motivation	.95	.73	.47 (.34,.59)	.48 (.33,.63)	.05 (.05,.06)	2.92	.82	-38.16

Note. MZ= 223 pairs; DZ= 83 pairs. a^2 = additive genetic variance; c^2 = shared environment variance; e^2 = nonshared environment variance; AIC = Akaike's information criterion.

Next, the model fit of the nested AC and AE models was compared to that of the full ACE models to further clarify the roles of genes and environments. All nested models had poorer fit than the full model, confirming the important roles of genes and shared environments in communicative ability and motivation ($\Delta \chi^2 = 47.98, 32.72, 13.47, 100.41$, $\Delta df = 1$, *p*s<.05, for communicative ability AE and CE models, and motivation AE and CE models respectively).

Bivariate genetic analyses

Genetic Cholesky decomposition model. The Cholesky decomposition model was employed to investigate the genetic and environmental links between communicative ability/ motivation and word reading/ General Reading. Four models were tested, including 1) communicative ability and word reading; 2) communicative ability and General Reading; 3) motivation and word reading; 4) motivation and General Reading. In each of the models, two sets of ACE terms were specified. The first set (A1, C1 and E1) linked to both variables (e.g., communicative ability and word reading in model 1), and the second set (A2, C2, and E2) contributed to word reading or General Reading only (see Figure 6.1). All models had satisfactory goodness-of-fit (see Table 6.5)

Figure 6.1. Four independent Cholesky decomposition models on communicative ability, motivation and reading abilities.

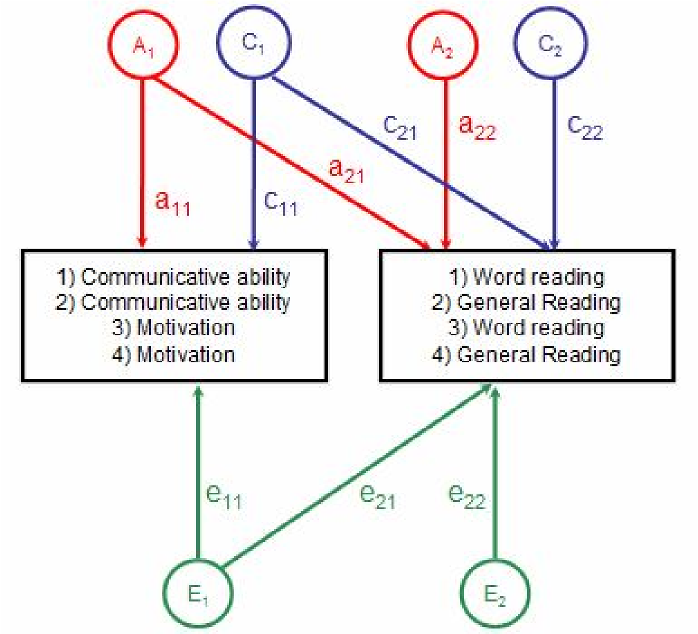


Table 6.5. Model Fitting Statistics of the Cholesky Decomposition ACE Models Compared with the Saturated Models.

`	*-2LL* (saturated)	*df* (saturated)	*-2LL* (ACE)	*df* (ACE)	Δ-*2LL* (Δ*df* =17)	*p*
Communicative ability (age- adjusted)						
Word reading	2402.74	1209	2415.93	1226	13.19	.72
General Reading	2560.91	1180	2568.22	1197	7.31	.98
Communicative ability (age- and nonverbal reasoning- adjusted)						
Word reading	2535.42	1207	2548.41	1224	12.99	.74
General Reading	2628.27	1179	2634.72	1196	6.45	.99
Motivation						
Word reading	2580.03	1206	2587.39	1223	7.36	.98
General Reading	2677.63	1178	2683.42	1195	5.79	.99

Note. Δ-*2LL* and Δ*df* are the differences between the Cholesky decomposition ACE and the saturated models.

Communicative ability. The path coefficients from the Cholesky decomposition models are shown in Table 6.6. The genetic and shared environmental influences on communicative ability (A1 and C1) were significant, but they did not significantly link to word reading or General Reading. However, word reading or General Reading were influenced by specific genetic and shared environmental factors (A2 and C2) which were independent from those on communicative ability. All coefficients of nonshared environmental paths, which included measurement errors, were significant, except the nonshared environmental link with General Reading. To further examine whether controlling for nonverbal reasoning effects yielded these nonsignificant links, the same model was tested with all scores accounting for age only. Results showed nonsignificant genetic and shared environmental links.

Motivation. Significant genetic and shared environmental influences on motivation (A1 and C1) were indicated. Also, the genetic link between motivation and word reading was significant (A1 to word reading), and their genetic correlation was found to be .31. However, other genetic and environmental factors on motivation did not link to the reading variables. Word reading and General Reading were influenced by specific genetic and shared environmental factors (A2 and C2) which were independent from those on motivation. All coefficients of nonshared environmental paths, which included measurement errors, were significant. The path coefficients from the Cholesky decomposition models are shown in Table 6.6.

Table 6.6. Standardized Path Coefficients from a Cholesky Decomposition Model of Genetic (A); Shared-Environmental (C); and Nonshared-Environmental (E) Influences on Communicative Ability, Motivation, Word Reading and General Reading. (95% Confidence Intervals in Parentheses)

	Paths					
	A1	A2	C1	C2	E1	E2
Communicative ability (age-adjusted)						
Communicative ability	.44		.88		.21	
	(.34,.53)		(.78,.98)		(.19,.23)	
Word reading (age-adjusted)	.15	.84	.08	.42	.06	.30
	(-.10,.40)	(.68,.99)	(-.09,.24)	(.09,.75)	(.02,.10)	(.27,.33)
Communicative ability	.44		.88		.21	
	(.34,.53)		(.78,.98)		(.19,.23)	
General Reading (age-adjusted)	.18	.67	.11	.49	.01	.51
	(-.09,.46)	(.44,.90)	(-.06,.28)	(.19,.79)	(-.06,.08)	(.46,.56)
Communicative ability (age- and nonverbal reasoning- adjusted)						
Communicative ability	.43		.88		.23	
	(.32,.54)		(.78,.98)		(.21,.25)	
Word reading	.14	.81	.05	.41	.10	.37
	(-.14,.41)	(.64,.98)	(-.12,.22)	(.06,.77)	(.05,.15)	(.34,.40)
Communicative ability	.43		.88		.23	
	(.32,.53)		(.78,.98)		(.21,.25)	
General Reading	.14	.64	.12	.51	.04	.54
	(-.15,.43)	(.39,.88)	(-.05,.29)	(.21,.81)	(-.03,.11)	(.49,.59)
Motivation						
Motivation	.68		.69		.23	
	(.56, .81)		(.53,.85)		(.21,.25)	
Word reading	.26	.78	-.09	.40	.12	.36
	(.03,.49)	(.62,.95)	(-.36,.18)	(.01,.80)	(.07,.17)	(.33,.40)
Motivation	.68		.69		.23	
	(.55,.81)		(.54,.84)		(.21,.25)	
General Reading	.17	.64	.05	.51	.08	.54
	(-.07,.41)	(.40,.89)	(-.22,.31)	(.20,.81)	(.01,.15)	(.48,.59)

To understand the extent to which the common and the specific genetic factors contribute to the variance of word reading comparatively, the proportions of genetic variances overlapped between the motivation and word reading (A1 to word reading) and specific to word reading (A2 to word reading) were computed. These proportions were calculated by dividing the squared common genetic path or the squared specific genetic path by the sum of both of these squared paths. Results indicated that genetic factors that explained the variances in motivation accounted for 10% of the genetic variances in word reading ($10\% = .26^2/.26^2+.78^2$).

Nested models. Eight nested models with the genetic link (A1 to word reading or General Reading), the shared environmental link (C1 to word reading or General Reading), the specific genetic factor (A2 to word reading or General Reading), or the specific shared environmental factor (C2 to word reading or General Reading) dropped were fitted for communicative ability and motivation separately (see Appendix 2D). The model fit of these nested models was compared to that of the full model to further clarify the roles of genes and environments. When the genetic link was dropped, there was a significant decrease in model fit for motivation and word reading only ($p<.05$). Also, the removal of the specific genetic path reduced model fit significantly for all variables (*p*s<.05). Conversely, when either the shared environmental link or the specific shared environmental path was dropped, the model fit was not significantly poorer than the full model for all variables (*p*s>.05). These results highlighted the genetic overlap between motivation and word reading, and also the roles of specific genetic factors in influencing word reading and General Reading which were independent from those on parent-rated communicative ability and motivation.

Discussion

This part of the thesis investigated children’s communicative ability and motivation and their links with reading abilities with a twin study design in 307 Chinese twin pairs. There are two major findings. First, results showed significant genetic and shared environmental contributions to both communicative ability and motivation, in which these effects were comparable for motivation but influences of shared environment were stronger than genetic factors for communicative ability. Second, there was significant genetic mediation of the link between motivation and word reading, but independent genetic and environmental influences on motivation and General Reading, and on communicative ability and both word reading and General Reading were demonstrated. These findings have enhanced our understanding of the extent to which genetic and environmental factors determine communicative ability and motivation for learning text and the mechanisms of their links with reading skills. Because past twin research on children’s reading emphasized cognitive aspects, and this thesis adds to this research area by examining motivation. Also, children’s communicative ability and motivation were assessed by parental reports which provide better account of children’s performance in naturalistic situations.

Genetic and environmental contributions to communicative ability and motivation

Results indicated genetic and shared environmental factors had significant influences on both communicative ability and motivation. For communicative ability, the majority of its variance was explained by shared environment (77%). After these strong shared environmental effects were accounted for, genetic factors contributed to a significant 18% of its variance. For motivation, genetic and shared environmental factors each explained around half of its variance. These findings confirmed the hypothesis on the genetic contributions to communicative ability and the environmental influences on motivation, except the genetic influences on communicative ability were smaller than expected. The moderate to substantial shared environmental influences could reflect the effects of ratings made by the same informant for each twin pair. The informants may have systematic bias in giving consistent ratings to a twin pair which leads to inflation of cotwins' correlations and in turn shared environmental estimates. This rater bias could not be teased apart from the genuine effects of shared environment with the design employed in this thesis, and therefore cautious result interpretation is warranted. In future, this concern could be addressed by obtaining ratings for a twin pair from different informants, such as each twin in a pair separately rated by their mother and father. Van der Valk, van den Oord, Verhulst, and Boomsma (2001) investigated children's problem behaviours using separate maternal and paternal ratings on each twin in a pair. In their study, they found both common and unique components in parents' rating on their children's behaviours. The unique genetic, shared environmental and nonshared environmental components each explained around 8% of the variances in problem behaviours, and these components reflected rater bias. The common genetic, shared environmental and nonshared environmental factors which were not confounded by rater bias contributed to the majority (around 75% altogether) of the total variances in children's behaviours. Therefore, though rater bias existed, their effects were relatively small. In a similar vein, Bishop et al., (2006) tested the influences of rater bias on communicative difficulties by comparing same-teacher and different-teacher ratings on twin pairs, and indicated the evaluation given by the same teacher involved rater bias. Therefore, it is promising to extend this thesis by including separate maternal and paternal ratings of twin pairs in further research.

It is interesting to note that significant genetic influences on communicative ability and motivation were indicated after the environmental influences were controlled. The genetic contribution to motivation chimed with the finding on reading exposure (Harlaar et al., 2007a), though it was higher for motivation indicated in this thesis. Therefore, apart from reading and related cognitive skills, other reading behaviours and beliefs could also be genetically predisposed. Regarding communicative ability, the heritability of communicative skills in this thesis was much lower compared to that in Bishop et al. (2006). One possible reason is Bishop et al. (2006) emphasized communicative impairment, included both typically developing and language impaired children, and administered all positive and negative items of CCC, while this thesis tested typically developing children and assessed the communication strengths only. Another possible reason is Bishop et al. (2006) included ratings from different informants, such as parents, same teachers and different teachers, whereas this thesis included parents only and so was more prone to rater bias which could inflate shared environmental and underestimate genetic contributions. However, the indication of genetic effects on communicative ability and motivation suggests this rater bias, if existed, did not completely account for the results of this thesis. It may also actually reflect genuine differences in communicative ability between Chinese and English children. These findings on motivation and communicative ability are unique in two ways. First, it showed that reading-related constructs other than those of the cognitive level, could be determined by genetic influences. Second, parent-assessed language and reading-related skills could pick up the genetic effects as language and reading psychometric tests did.

It should be noted that the genetic terms could be influenced by zygosity-dependent contrast effects. These effects occur when raters who evaluate both twins in a pair tend to evaluate monozygotic twins as more similar to one another than they do on dizygotic twins. These effects could be tested, again, with a design with ratings obtained from different informants. If the genetic terms specific to the informants are significant in same-informant ratings only but not different-informant ones, the genetic effects could be due to the zygosity-dependent contrast factor. Bishop et al. (2006) indicated nonsignificant specific genetic terms on communicative difficulties evaluated by different teachers, showing the existence of zygosity-dependent contrast effects. Therefore, these effects should be aware of when interpretating the genetic findings.

The links of communicative ability and motivation, and reading skills

Communicative ability and motivation were significantly linked to word reading and general reading abilities, which were consistent with the hypothesis. These links were modest after accounting for the effects of age and nonverbal reasoning on reading skills (*r* ranged from .12 to .17). Similar correlation magnitude between motivation and reading was demonstrated in past research, though some studies showed moderate correlations between them (see Morgan & Fuchs, 2007). The modest links may be due to the skewed ratings towards the high end for both communicative ability and motivation (e.g., Law, 2009). Nevertheless, these results indicated the tendency that children with higher communicative ability or motivation were likely to be better readers even after the effects of their age and nonverbal reasoning skills were considered.

The genetic analyses of this thesis indicated significant genetic overlap between motivation and word reading, and also the important roles of specific genetic factors, which were independent from those on parent-rated communicative ability and motivation, in influencing word reading and General Reading. Also, independent genetic and shared environmental factors contributed to communicative ability, and word reading and General Reading, showing different developmental etiology of communicative ability and reading skills. These findings corresponded to the genetic and environmental independency of general language and general reading skills in Chinese (see chapter 3). In this thesis, communicative ability was conceptualized as a broader construct than general language skills, as it constituted not only wide aspects of language skills (as tapped by tasks of various language structures) but also pragmatics. So, even this broader aspect of verbal skills had different etiology with reading abilities in Chinese.

Motivation overlapped genetically with word reading, in which genetic factors that contributed to the variance in motivation explained 10% of the genetic variance in word reading. This finding chimed with past research on reading exposure (Harlaar et al., 2007a), and this line of research underscores the genetic perspective on the relationships between reading-related bahaviours and beliefs and reading skills. Note that the genetic variances overlapped for motivation and word reading, but were independent for motivation and general reading skills. In other words, motivation had common genetic etiology with word reading only, but not general reading abilities. Therefore, the genetic links were specific to word reading, and could not be generalized to reading abilities in broader aspects or even other reading-related cognitive skills. In contrast, there was no shared environmental links between motivation and both specific and general domains of reading abilities. In sum, the link between motivation and word reading was mediated by genetic factors instead of environmental ones. Children who are more motivated to learn written Chinese are likely to perform better on word reading because of genetic processes. Note that motivation and reading develop hand in hand. Children with better reading skills tend to have higher motivation in reading, and in turn higher motivation leads to further improvement in reading skills. So, the gap between originally good and poor readers widens in the course of development, showing the Matthew effects (Stanovich, 1986). Follow-up longitudinal research on motivation and reading could better reflect their developmental nature.

Limitations

This part of the thesis has three major caveats. First, communicative ability and motivation of a twin pair were rated by the same informant. Therefore, findings could be confounded with rater bias. Second, it administered only the 20 positive items of the CCC-2. Though the validity of the CCC-2 has been empirically proven, this short form with only the positive items needs further validation. Lastly, motivation was evaluated by general statements on children's beliefs, but statements with more concrete reading-related situations could yield comparatively more objective and specific evaluation (e.g., Motivation for reading questionnaire, Wigfield & Guthrie, 1997). Different results may be obtained with motivation tapped in different ways.

Implications and conclusions

Despite these possible caveats, this thesis has enhanced our understanding of the extent to which genetic and environmental factors determine communicative ability and motivation and the mechanisms of their links with reading skills. Specifically, its contributions are fourfold. First, this thesis has extended the utility of the CCC-2 to Chinese communicative ability in typically developing Chinese children, and has demonstrated the etiology of communicative ability. There were substantial shared environmental influences and significant but modest genetic effects on communicative ability. Second, it has extended the genetic perspective on reading and related cognitive skills to reading-related behaviours and beliefs. Results showed comparable genetic and environmental contributions to motivation. Third, this thesis has helped understand the underlying mechanisms of how motivation is related to reading skills. Its results found the link between motivation and word reading was mediated by genetic factors but not environmental influences. Lastly, this thesis has fostered the knowledge of the genetic perspective on communicative ability and motivation evaluated in naturalistic contexts.

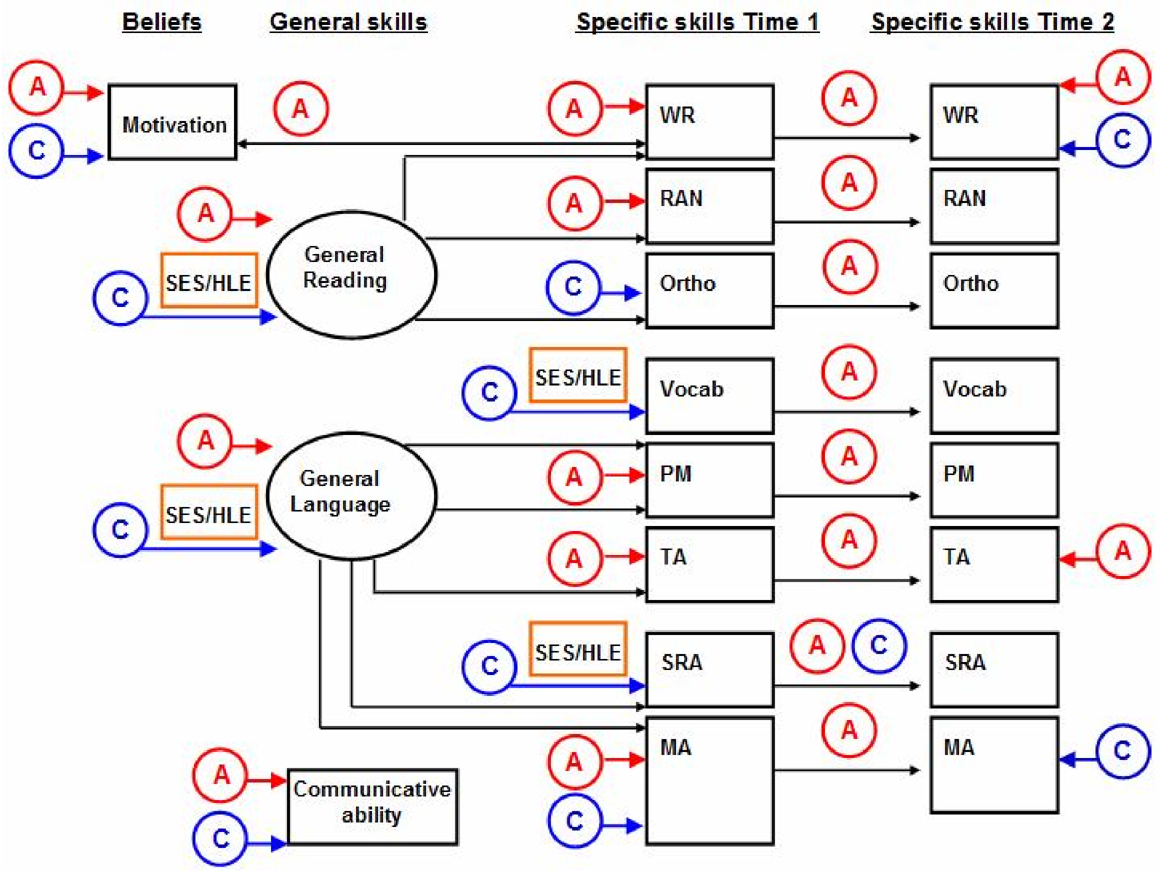
Chapter 7

General Discussion

Summary of results

This thesis investigated the genetic and environmental contributions to Chinese language and literacy skills with a two-wave longitudinal design using a sample of 312 Chinese twin pairs aged 3 to 11 in Hong Kong. Four domains of genetic and environmental influences were examined, including a) their relative contributions to various language and reading skills and the links between general language and general reading abilities (Chapter 3); b) testing socioeconomic status (SES) and home literacy environment (HLE) as plausible environmental mediators and/or moderators of these influences (Chapter 4); c) their contributions to the stability and change of various language and reading skills across time (Chapter 5); and d) their influences on parent-rated communicative ability and motivation for learning text (Chapter 6). The major findings are summarized in Figure 7.1.

Figure 7.1. A Summary of major findings in this thesis.



Note. A= genetic influences; C= shared environmental influences; WR= word reading; RAN= rapid automatized naming; Ortho= orthography skills; Vocab= receptive vocabulary; PM= phonological memory; TA= tone awareness; SRA= syllable and rhyme awareness; MA= morphological awareness; SES= socioeconomic status; HLE= home literacy environment.

Figure 7.1 illustrates several major findings. To begin with, genetic and environmental factors had differential influences on the individual differences in various Chinese language and reading skills (see Specific skills Time 1 column). Specifically, genetic factors played a more important role in word reading, rapid automatized naming, phonological memory and tone awareness. Shared environmental effects were more prominent in orthographic skills, receptive vocabulary, and syllable and rhyme awareness, and these effects on receptive vocabulary and syllable and rhyme awareness were mediated by family socioeconomic status and home literacy environment (reading resources and opportunities at home). Also, results of this thesis suggested comparable effects of genetic and shared environmental factors on morphological awareness.

Genetic factors which influenced these skills at Time 1, regardless of the strengths of their contributions, continued to exert their effects at Time 2 (see Specific skills Time 2 column). In other words, genetic influences maintained the stability of each of these skills across a one-year time period. Though the continuity of all these skills was mediated by genetic processes, shared environment was also important for the stability of syllable and rhyme awareness. After a one-year period, new genetic and shared environmental factors emerged for word reading, and new shared environmental influences were found for morphological awareness. Results also suggest new genetic factors emerged for tone awareness. Therefore, both stable and dynamic features of genetic and environmental influences were demonstrated.

More general language and literacy skills were also considered. Results of this thesis showed general language and general reading abilities shared neither genetic nor shared environmental influences, indicating the possibility of different etiology in their development (see General skills column). Under this context, it is interesting to find SES and HLE (reading resources and opportunities at home) mediated the shared environmental effects of both general language and general reading abilities. These findings, when put together, suggest the domains under SES and HLE might have differential effects on each of these skills. For instance, richer reading resources enhance the opportunities for parents to explicitly teach and discuss print with their children, and these foster children's reading development, whereas language interaction between parent and child stimulated by these reading resources and opportunities is more important for children’s language development (e.g., Chow et al., 2005; Sénéchal, 2006). Also, SES and HLE (reading resources and opportunities at home) were plausible mediators of environmental influences, but they did not moderate genetic influences on general language and general reading abilities. This implies the same sets of genetic influences might function across the entire reading abilities, regardless of the SES and HLE (reading resources and opportunities at home) variations. The different genetic and environmental origins were found not only between language and reading skills, but also between communicative ability rated by parents and reading skills. Parent-rated communicative ability was influenced by strong shared environmental influences and significant but modest genetic effects.

So far these results concerned the cognitive aspects of reading. Findings of this thesis also extended the genetic perspective to motivation for learning text assessed in naturalistic contexts (see Beliefs column). Moderate genetic and shared environmental contributions were indicated for parent-rated motivation for learning text. Also, the link between motivation and word reading was mediated by genetic processes. So, apart from cognitive skills relevant to reading, reading beliefs and behaviours and their links with reading ability could also be genetically predisposed.

Possible universal genetic and environmental contributions

This thesis has indicated some findings on Chinese language and literacy skills which are consistent with past studies on alphabetic languages. Chinese is characteristically distant to alphabetic languages, and thus these findings suggest genetic and environmental influences which are likely to be universal across languages. Overall, three aspects of universal influences are suggested. First, genetic factors exerted greater impacts on word reading, phonological memory, and rapid automatized naming, whereas shared environments had stronger effects on vocabulary knowledge. Second, home environments, including SES and HLE, mediated the shared environmental influences on language and reading skills. Third, genetic factors contributed to the stability of word reading and phonological awareness, and shared environments did not significantly influence the continuity of rapid automatized naming across time.

Novel findings on Chinese language and literacy skills

In contrast, some findings are more language- or country- specific. Results showed that individual differences in Chinese phonological awareness at syllable and rhyme levels and orthographic skills were more affected by shared environments than by genes, but past studies highlighted the genetic influences on these skills in alphabetic languages. Also, there was no genetic link between general language and general reading abilities in Chinese, and this is contrastive to past findings on alphabetic languages. Furthermore, results indicated new sources of genetic and environmental influences on Chinese word reading, and these new sources of influences were demonstrated in the United States samples in past research. Therefore, they could be country-specific.

This thesis has also demonstrated some novel findings of which the aspects of language and literacy skills and their relationships with other variables have not been studied in past twin research. First, genetic factors exerted greater influences than shared environments on Chinese tone awareness. Second, SES and HLE contributed to phonological awareness at the syllable and rhyme levels, but not orthographic skills, when genetic influences were accounted for. Third, the developmental etiology of phonological memory, tone awareness, morphological awareness and orthographic skills was indicated. Specifically, results showed genetic influences on the stability of phonological memory, tone awareness, morphological awareness and orthographic skills, potential genetic contributions to the instability of tone awareness, and shared environmental mediation of the instability of morphological awareness. Fourth, substantial shared environmental influences and significant but modest genetic effects on communicative ability in typically developing children were indicated. Lastly, motivation for learning text was influenced by comparable genetic and environmental effects, and was linked to word reading skills through genetic processes.

Limitations

Findings of this thesis have to be interpreted with the consideration of several caveats. First, this thesis has investigated the mediating roles of SES and HLE by assuming negligible genotype-environment correlation, and SES and HLE as pure environmental variables. However, these family environment variables may not be pure measures of shared environment, because, for instance, a parent's SES may be affected by their genes which they also share with their children. In this case, the family measures have the genetic influences subsumed, and thus could lead to spurious findings of environmental mediation (Purcell & Koenen, 2005; Turkheimer et al., 2005).

Second, this thesis included sample with a relatively large age range. There are potential problems in interpreting the longitudinal findings as the age ranges overlapped substantially across the one-year period (i.e., 3 to 11 years at Time 1 and 4 to 12 years at Time 2), and informing models of reading development as the sample covered different stages of reading. Therefore, further research could aim to reducing the sample's age range and focusing on particular reading stages, such as the developmental changes from being taught literacy informally to formally, and from learning to read to reading to learn.

Third, this thesis has showed the overlap between motivation and reading skills, but has not studied the developmental nature of motivation and reading, such as the Matthew effects (Stanovich, 1986). Motivation and reading develop hand in hand. Children with better reading skills tend to have higher motivation in reading, and in turn higher motivation leads to further improvement in reading skills, making a wider gap between originally good and poor readers. Follow-up longitudinal research on the same sample with motivation and reading data collected at different time points could better reflect the developmental nature of motivation and reading.

Fourth, the dual-route model has been discussed in Chapter 1 as a framework for illustrating the similarities and differences between English word and Chinese character reading. However, this model was not experimentally tested in the present thesis. Also, it should be noted that there are alternative models which could be relevant in Chinese reading, such as the "triangle" connectionist approach (Seidenberg & McClelland, 1989). Past research has demonstrated the frequency, consistency and regularity effects on Chinese reading under the connectionist framework (e.g., Yang, McCandliss, Shu, & Zevin, 2009). The twin study design could be employed to investigate models of reading, and previous twin studies have tested the dual-route model in English reading (e.g. Bates et al., 2007). Therefore, extending this thesis to involve irregular-word and novel-word reading in Chinese, and examine frequency, consistency and regularity effects would better inform the underlying mechanisms of Chinese reading.

Lastly, this thesis included same-sex twin pairs only, but did not correct for potential sex effects. However, there are potential sex effects owing to the likeness of better language and reading abilities in girls than boys (e.g., Logan & Johnston, 2009), and sharing of a common sex in each twin pairs which could lead to higher resemblance between members in each twin pair (e.g., McGue & Bouchard, 1984). To test the size of sex differences on the reading and language measures in this thesis, R^2 between sex and age-adjusted scores were computed, and they ranged from .00 to .01. McGue and Bouchard (1984) noted a larger R^2 indicated a stronger influence of age-sex adjustment. For instance, when R^2 was .10, age-sex adjustment made little difference (e.g., age-sex adjusted intraclass twin correlation was .40 while the unadjusted one was .46). Therefore, sex-adjustment would make little difference to the findings in this thesis. However, given the potential sex effects, further studies could take sex effects into account.

Future research direction

Findings of this thesis are cornerstones for further understanding of the genetic and environmental origins of language and literacy development. The universality of genetic and environmental influences inferred from these findings suggests very similar etiology of development in these skills across languages. The next step of research is to identify these influences so as to confirm if they are the same sets of genes and/or environments across languages. It could be possible that both word reading in Chinese and in alphabetic languages is highly influenced by genetic factors, but they could be different sets of genes. It could also be possible that the sets of genes influencing word reading in alphabetic languages could exert effects on Chinese word reading, and vice versa. Therefore, identifying these influences is prominent in understanding the roots of these universal findings. Similarly, the environments exerting influences on vocabulary should be better known. While molecular genetic studies have located genes, such as KIAA0319, which linked to normal variations in reading skills of alphabetic languages (e.g., Paracchini et al., 2008), this thesis together with very few twin studies have put effort on pinpointing plausible environments, such as family SES, chaos, and HLE, influencing language and reading skills (e.g., Hart, Petrill, DeThorne, et al., 2009; Petrill et al., 2004). Recently, researchers have started to examine the effects of assigning twins to the same or different classrooms on their literacy achievement (e.g., Byrne et al., 2010). These studies have paved the way for pursuing what genetic and environmental factors are prominent in language and literacy development. Further studies are needed to replicate these research findings. Also, they can cover broader aspects of environmental influences, including school literacy environments.

On the other hand, the novel findings on Chinese language and literacy skills encourage further investigation of why there are such differences across languages. Cross-cultural studies including parallel language and literacy measures in Chinese and alphabetic languages, and testing their links with the same sets of genetic markers and environments could better show the reasons. These findings also motivate the construction of learning environments which could be the most efficient for learning a particular language. For instance, Chinese language and reading development involves different genetic environmental and processes, and so different activities tailor-made to language or reading learning could be the most efficient in Chinese acquisition. Special attention should be paid to these specific influences in theory development and policy making.

In addition, this thesis and past research (e.g., Harlaar et al., 2007a) have suggested the genetic origins of reading-related behaviours and beliefs, including motivation for learning text and reading exposure, and their links with literacy skills. Comparing to cognitive aspects of reading, the etiology of these behavioural and belief domains is much less understood. However, both strands are prominent in understanding why children differ greatly on language and literacy skills. Future studies which substantiate the existing findings, and also explore wider behavioural and belief aspects, such as causal attributions and outcome expectancy, are promising.

Conclusion

Language and literacy abilities are complex, multi-dimensional, and quantitatively distributed. Therefore, it is unlikely that a single determinant could fully explain their variability. Research has pursued the quest for their etiology from different perspectives, including genetic, cognitive, behavioural and belief, and developmental aspects. This thesis has put these perspectives together, and demonstrated how cognitive skills and motivation were linked to language and reading abilities, to what extent genes and environments contributed to their individual differences and links, and how genes and environments mediated the stability and change in language and literacy skills across time. This thesis has contributed to our knowledge of the origins and development of language and literacy skills, and helps paving the road for achieving the ultimate goals of fully exploiting children's potentials and for supporting at risk children to surpass their genetic propensities.

References

Anderson, R. C., Li, W., Ku, Y.-M., Shu, H., & Wu, N. (2003). Use of partial information in learning to read Chinese characters. *Journal of Educational Psychology, 95*(1), 52-57.

Asbury, K., Wachs, T. D., & Plomin, R. (2005). Environmental moderators of genetic influence on verbal and nonverbal abilities in early childhood. *Intelligence, 33*(6), 643-661.

Badian, N. A. (2001). Phonological and orthographic processing: Their roles in reading prediction. *Annals of Dyslexia, 51*, 179-202.

Baker, L., Scher, D., & Mackler, K. (1997). Home and family influences on motivations for reading. *Educational Psychologist, 32*(2), 69-82.

Barker, T. A., Torgesen, J. K., & Wagner, R. K. (1992). The role of orthographic processing skills on five different reading tasks. *Reading Research Quarterly, 27*(4), 334-345.

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

Bates, T. C., Castles, A., Luciano, M. Wright, M. J., Coltheart, M., et al. (2007). Genetic and environmental bases of reading and spelling: A unified genetic dual route model. *Reading and Writing, 20*(1-2), 147-171.

Bates, E., Dale, P. S., & Thal, D. (1995). Individual differences and their implications for theories of language development. In P. Fletcher & B. MacWhinney (Eds.), *The handbook of child language* (pp.96-151). Oxford: Blackwell Publishing.

Betjemann, R. S., Willcutt, E. G., Olson, R. K., Keenan, J. M., DeFries, J. C., & Wadsworth, S. J. (2008). Word reading and reading comprehension: Stability, overlap and independence. *Reading and Writing, 21*(5), 539-558.

Bishop, D. V. M. (1998). Development of the Children's Communication Checklist (CCC): A method for assessing qualitative aspects of communicative impairment in children. *Journal of Child Psychology and Psychiatry, 39*(6), 879-891.

Bishop, D. V. M. (2002). The role of genes in the etiology of specific language impairment. *Journal of Communication Disorders, 35*(4), 311-328.

Bishop, D. V. M. (2003). *The Children's Communication Checklist – 2nd Edition (CCC-2).* London: Psychological Corporation.

Bishop, D. V. M., Laws, G., Adams, C., & Norbury, C. F. (2006). High heritability of speech and language impairments in 6-year-old twins demonstrated using parent and teacher report. *Behavior Genetics, 36*(2), 173-184.

Bleker, O. P., Breur, W., & Huidekoper, B. L. (1979). A study of birth weight, placental weight and mortality of twins as compared to singletons. *British Journal of Obstetrics and Gynaecology, 86*(2), 111–118.

Bouchard T. J., & McGue, M. (1981). Familial studies of intelligence: A review. *Science, 212*(4498), 1055-1059.

Bowey, J. A. (1995). Socioeconomic status differences in preschool phonological sensitivity and first-grade reading achievement. *Journal of Educational Psychology, 87*(3), 476-487.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read: A causal connection. *Nature, 301*(5899), 419-421.

Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual Review of Psychology, 53*(1), 371-399.

Braun, W. J., & Murdoch, D. J. (2007). *A first course in statistical programming with R.* New York: Cambridge University press.

Bronfenbrenner, U., & Ceci, S. J. (1994). Nature-nurture reconceptualized in developmental perspective: A bioecological model. *Psychological Review, 101*(4), 568-586.

Brooks, A., Fulker, D. W., & DeFries, J. (1990). Reading performance and general cognitive ability: A multivariate genetic analysis of twin data. *Personality and Individual Differences, 11*(2), 141-146.

Brownell, R. (2000). *Expressive One-Word Picture Vocabulary Test Manual.* Novato, CA: Academic Therapy Publications.

Burgess, S. R., Hecht, S. A., & Lonigan, C. J. (2002). Relations of the home literacy environment (HLE) to the development of reading-related abilities: A one-year longitudinal study. *Reading Research Quarterly, 37*(4), 408-426.

Bus, A. G., van Ijzendoorn, M. H., & Pellegrini, A. D. (1995). Joint book reading makes for success in learning to read: A meta-analysis on intergenerational transmission of literacy. *Review of Educational Research, 65*(1), 1-21.

Bus, A., & Out, D. (2009). Unraveling genetic and environmental components of early literacy: A twin study. *Reading and Writing, 22*(3), 293-306.

Butterworth, B., & Wengang, Y. (1991). The universality of two routines for reading: Evidence from Chinese dyslexia. *Proceedings of the Royal Society: Biological Sciences, 246,* 91-95.

Byrne, B., Coventry, W. L., Olson, R. K., Samuelsson, S., Corley, R., Willcutt, E. G., et al. (2009). Genetic and environmental influences on aspects of literacy and language in early childhood: Continuity and change from preschool to grade 2. *Journal of*

Neurolinguistics, 22(3), 219-236.

Byrne, B., Coventry, W. L., Olson, R. K., Wadsworth, S. J., Samuelsson, S., Petrill, S. A., et al. (2010). “Teacher effects” in early literacy development: Evidence from a study of twins. *Journal of Educational Psychology, 102*(1), 32-42.

Byrne, B., Delaland, C., Fielding-Barnsley, R., Quain, P., Samuelsson, S., Hoien, T., et al. (2002). Longitudinal twin study of early reading development in three countries: Preliminary results. *Annals of Dyslexia, 52,* 49-73.

Byrne, B., Olson, R. K., Samuelsson, S., Wadsworth, S., Corley, R., DeFries, J. C., et al. (2006). Genetic and environmental influences on early literacy. *Journal of Research in Reading, 29*(1), 33-49.

Byrne, B., Samuelsson, S., Wadsworth, S., Hulslander, J., Corley, R., Defries, J. C., et al. (2007). Longitudinal twin study of early literacy development: Preschool through grade 1. *Reading and Writing, 20*(1-2), 77-102.

Byrne, B., Wadsworth, S., Corley, R., Samuelsson, S., Quain, P., DeFries, J. C., et al. (2005). Longitudinal Twin Study of Early Literacy Development: Preschool and Kindergarten Phases. *Scientific Studies of Reading, 9*(3), 219-235.

Castles, A., & Coltheart, M. (1993). *Varieties of developmental dyslexia. Cognition, 47*(2), 149-180.

Chan, C. K. K., & Siegel, L. S. (2001). Phonological processing in reading Chinese among normally achieving and poor readers. *Journal of Experimental Child Psychology, 80*(1), 23-43.

Chan, D .W., Ho, C. S.-H., Tsang, S.-M., Lee, S.-H., & Chung, K. K. H. (2007). Prevalence, gender ratio and gender differences in reading-related cognitive abilities among Chinese children with dyslexia in Hong Kong. *Educational Studies, 33*(2), 249-265.

Chan, L., & Nunes, T. (1998). Children's understanding of the formal and functional characteristics of written Chinese. *Applied Psycholinguistics, 19*(1), 115-131.

Chan, L., & Wang, L. (2003). Linguistic awareness in learning to read Chinese: A comparative study of Beijing and Hong Kong children. In C. McBride-Chang & H.-C. Chen (Eds.), *Reading development in Chinese children* (pp. 91-106). Westport, CT: Praeger Press.

Chen, X., Hao, M., Geva, E., Zhu, J., & Shu, H. (2009). The role of compound awareness in Chinese children's vocabulary acquisition and character reading. *Reading and Writing, 22*(5), 615-631.

Chen, X., Shu, H., Wu, N. N., & Anderson, R. C. (2003). Stages in learning to pronounce Chinese characters. *Psychology in the Schools, 40*(1), 115-124.

Cheung, H., & Ng, L. (2003). Chinese reading development in some major Chinese societies: An introduction. In C. McBride-Chang & H.-C. Chen (Eds.), *Reading development in Chinese children* (pp. 3-17). Westport, CT: Praeger Press.

Cheung, H., Chan, M., & Chong, K. (2007). Use of orthographic knowledge in reading by Chinese-English bi-scriptal children. *Language Learning, 57*(3), 469-505.

Chia, K. S., Lee, J. J. M., Cheung, P., Cheung, K. H., Seielstad, M., Wilcox, M. M., et al. (2004). Twin births in Singapore: A population-based study using the national birth registry. *Annals Academy of Medicine Singapore, 33*(2), 195-199.

Chiu, M. M., Chow, B., W.-Y. (in press). Culture, motivation, and reading achievement: High school students in 41 countries. *Learning and Individual Differences.*

Chow, B. W.-Y., & McBride-Chang, C. (2003). Promoting language and literacy development through parent-child reading in Hong Kong preschoolers. *Early Education and Development, 14*(2), 233-248.

Chow, B. W.-Y., McBride-Chang, C., & Burgess, S. (2005). Phonological processing skills and early reading abilities in Hong Kong Chinese kindergartners learning to read English as a second language. *Journal of Educational Psychology, 97*, 81-87.

Chow, B. W.-Y., McBride-Chang, C., Cheung, H., & Chow, C. S.-L. (2008). Dialogic reading and morphology training in Chinese children: Effects on language and literacy. *Developmental Psychology, 44*(1), 233-244.

Colledge, E., Bishop, D. V. M., Koeppen-Schomerus, G., Price, T. S., Happé, F. G. E., Eley, T. C., et al. (2002). The structure of language abilities at 4 years: A twin study. *Developmental Psychology, 38*(5), 749-757.

Coltheart, M. (2005). Modeling reading: the dual route approach. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: a handbook* (pp.6-23). Oxford: Blackwell Publishing.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review, 108*(1), 204-256.

Conway, D., Lytton, H., & Pysh, F. (1980). Twin-singleton language differences. *Canadian Journal of Behavioral Science, 12,* 264-271.

Cunningham, A. E., Perry, K. E., & Stanovich, K. E. (2001). Converging evidence for the concept of orthographic processing. *Reading and Writing, 14*(5-6), 549-568.

Cunningham, A. E., & Stanovich, K. E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33,* 934-945.

Davis, C. J., Knopik, V. S., Olson, R. K., Wadsworth, S. J., & DeFries, J. C. (2001). Genetics and environmental influences on rapid naming and reading ability. *Annals of Dyslexia, 51,* 231-247.

Deacon, S. H., & Kirby, J. R. (2004). Morphological awareness: Just "more phonological"? The roles of morphological and phonological awareness in reading development. *Applied Psycholinguistics, 25*(2), 223-238.

Dionne, G., Dale, P. S., Boivin, M., & Plomin, R. (2003). Genetic evidence for bidirectional effects of early lexical and grammatical development. *Child Development, 74*(2), 394-412.

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test – 4th Edition.* Minneapolis, MN: Pearson Assessments.

Ehri, L. C. (1995). Phases of development in learning to read words by sight. *Journal of Research in Reading, 18*(2), 116-125.

Ehri, L. C. (2005). Development of sight word reading: Phases and findings. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp.135-154). Oxford: Blackwell Publishing.

Feldman, H. M., Dale, P. S., Campbell, T. F., Colborn, D. K., Kurs-Lasky, M., Rockette, H. E., et al. (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child Development, 76*(4), 856-868.

Feldman, L. B., & Siok, W. W. T. (1999). Semantic radicals in phonetic compounds: Implications for visual character recognition in Chinese. In J. Wang, A. W. Inhoff, & H.-C. Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp.19-35). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Friend, A., DeFries, J. C., & Olson, R. K. (2008). Parental education moderates genetic influences on reading disability. *Psychological Science, 19*(11), 1124-1130.

Friend, A., DeFries, J. C., Olson, R. K., Pennington, B., Harlaar, N., Byrne, B., et al. (2009). Heritability of high reading ability and its interaction with parental education. *Behavior Genetics, 39*(4), 427-436.

Frith, U. (1985). Beneath the surface of developmental dyslexia. In K. Patterson, J. Marshall, & M. Coltheart (Eds.), *Surface dyslexia, neuropsychological and cognitive studies of phonological reading* (pp. 301-330). London: Erlbaum.

Fu, C.-L., & Huang, H.-S. (2000). The effects of metalinguistic awareness training of the low-achievers at elementary schools. *Chinese Journal of Psychology, 42(1)*, 101-112.

Fu, G. S. (1987). The Hong Kong bilingual. In R. Lord & N. L. Cheng (Eds.), *Language education in Hong Kong* (pp.27-50). Hong Kong: The Chinese University Press.

Gathercole, S. E., Willis, C., & Baddeley, A. D. (1991). Differentiating phonological memory and awareness of rhyme: Reading and vocabulary development in children. *British Journal of Psychology, 82*(3), 387-406.

Gayán, J., & Olson, R. K. (2003). Genetic and environmental influences on individual differences in printed word recognition. *Journal of Experimental Child Psychology, 84*(2), 97-123.

Genesee, F., Tucker, G. R., & Lambert, W. E. (1975). Communication skills of bilingual children. *Child Development, 46*(4), 1010-1014.

Ginsburg, G. S., & Bronstein, P. (1993). Family factors related to children's intrinsic/extrinsic motivational orientation and academic performance. *Child Development, 64*(5), 1461-1474.

Griffin, E. A., & Morrison, F. J. (1997). The unique contribution of home literacy environment to differences in early literacy skills. *Early Child Development and Care, 127-128*, 233-243.

Guthrie, J. T., Wigfield, A., Humenick, N. M., Perencevich, K. C., Taboada, A., & Barbosa, P. (2006). Influences of stimulating tasks on reading motivation and comprehension. *Journal of Educational Research, 99*(4), 232-245.

Harlaar, N., Dale, P. S., & Plomin, R. (2005). Telephone testing and teacher assessment of reading skills in 7-year-olds: II. Strong genetic overlap. *Reading and Writing, 18*(5), 401-423.

Harlaar, N., Dale, P. S., & Plomin, R. (2007a). Reading exposure: a (largely) environmental risk factor with environmentally-mediated effects on reading performance in the primary school years. *Journal of Child Psychology and Psychiatry, 48*(12), 1192-1199.

Harlaar, N., Dale, P. S., & Plomin, R. (2007b). From Learning to Read to Reading to Learn: Substantial and Stable Genetic Influence. *Child Development, 78*(1), 116-131.

Harlaar, N., Hayiou-Thomas, M. E., Dale, P. S., & Plomin, R. (2008). Why do preschool language abilities correlate with later reading? A twin study. *Journal of Speech, Language, and Hearing Research, 51*(3), 688-705.

Hart, S. A., Petrill, S. A., DeThorne, L. S., Deater-Deckard, K., Thompson, L. A., Schatschneider, C., et al. (2009). Environmental influences on the longitudinal covariance of expressive vocabulary: Measuring the home literacy environment in a genetically sensitive design. *Journal of Child Psychology and Psychiatry, 50*(8), 911-919.

Hart, S. A., Petrill, S. A., Thompson, L. A., & Plomin, R. (2009). The ABCs of math: A genetic analysis of mathematics and its links with reading ability and general cognitive ability. *Journal of Educational Psychology, 101*(2), 388-402.

Haworth, C. M., Kovas, Y., Harlaar, N., Hayiou-Thomas, M. E., Petrill, S. A., Dale, P. S., et al. (2009). Generalist genes and learning disabilities: A multivariate genetic analysis of low performance in reading, mathematics, language and general cognitive ability in a sample of 8000 12-year-old twins. *Journal of Child Psychology and Psychiatry, 50*(10), 1318-1325.

Hayiou-Thomas, M. E., Harlaar, N., Dale, P. S., & Plomin, R. (2006). Genetic and environmental mediation of the prediction from preschool language and nonverbal ability to 7-year reading. *Journal of Research in Reading, 29*(1), 50-74.

Hayiou-Thomas, M. E., Harlaar, N., Dale, P. S., & Plomin, R. (2010). Preschool speech, language skills, and reading at 7, 9, and 10 years: Etiology of the relationship. *Journal of Speech, Language, and Hearing Research, 53,* 311–332.

Hecht, S. A., Burgess, S. R., Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (2000). Explaining social class differences in growth of reading skills from beginning kindergarten through fourth-grade: The role of phonological awareness, rate of access, and print knowledge. *Reading and Writing, 12*(1-2), 99-127.

Helland, W. A., Biringer, E., Helland, T., & Heimann, M. (2009). The usability of a Norwegian adaptation of the Children's Communication Checklist Second Edition

(CCC-2) in differentiating between language impaired and non-language impaired 6- to 12-year-olds. *Scandinavian Journal of Psychology, 50*(3), 287-292.

Ho, C. S.-H., & Bryant, P. (1997). Learning to read Chinese beyond the logographic phase. *Reading Research Quarterly, 32*(3), 276-289.

Ho, C. S.-H., Chan, D. W., Chung, K. K. H., Lee, S.-H., & Tsang, S.-M. (2007). In search of subtypes of Chinese developmental dyslexia. *Journal of Experimental Child Psychology, 97*(1), 61-83.

Ho, C. S.-H., Chan, D. W.-O., Tsang, S. M., & Lee, L. H. (2000). *The Hong Kong Test of Specific Learning Disabilities in Reading and Writing (HKT-SpLD).* Hong Kong, China: Chinese University of Hong Kong and Education Department, HKSAR Government.

Ho, C. S.-H., Leung, M.-T., & Cheung, H. (submitted). Early difficulties of Chinese preschool children at familial risk for dyslexia: Deficits in oral language, phonological processing skills, and print-related skills. Manuscript submitted for publication.

Ho, C. S.-H., Ng, T.-T., & Ng, W.-K. (2003). A "radical" approach to reading development in Chinese: The role of semantic radicals and phonetic radicals. *Journal of Literacy Research, 35*(3), 849-878.

Hoff-Ginsberg, E. (1998). The relation of birth order and socioeconomic status to children's language experience and language development. *Applied Psycholinguistics, 19*(4), 603-629.

Hohnen, B., & Stevenson, J. (1999). The structure of genetic influences on general cognitive, language, phonological, and reading abilities. *Developmental Psychology, 35*(2), 590-603.

Hong Kong Department of Education. (1990). *Syllabus for Chinese language (primary 1 - 6).* Hong Kong: Curriculum Development Institute, Hong Kong Department of Education.

Hong Kong Education Bureau (2008). *小學中國語文建議學習重點 (試用)*. Hong Kong: Curriculum Development Institute, Hong Kong Education Bureau.

Hoosain, R. (1991). *Psycholinguistic implications for linguistic relativity: A case study of Chinese.* Hillsdale, HJ: Erlbaum.

Hu, C.-F., & Catts, H. W. (1998). The role of phonological processing in early reading ability: What we can learn from Chinese. *Scientific Studies of Reading, 2*(1), 55-79.

Hulme, C., Hatcher, P. J., Nation, K., Brown, A., Adams, J., & Stuart, G. (2002). Phoneme awareness is a better predictor of early reading skill than onset-rime awareness. *Journal of Experimental Child Psychology, 82*(1), 2-28.

Imaizumi, Y. (2003). A comparative study of zygotic twinning and triplet rates in eight countries, 1972-1999. *Journal of Biosocial Science, 35*(2), 287-302.

Jackson, N. E., & Coltheart, M. (2001). *Routes to reading success and failure.* New York: Psychology Press.

Juel, C. (1988). Learning to read and write: A longitudinal study of 54 children from first through fourth grades. *Journal of Educational Psychology, 80*(4), 437-447.

Keenan, J. M., Betjemann, R. S., Wadsworth, S. J., DeFries, J. C., & Olson, R. K. (2006). Genetic and environmental influences on reading and listening comprehension. *Journal of Research in Reading, 29*(1), 75-91.

Kendler, K. S., & Baker, J. H. (2007). Genetic influences on measures of the environment: A systematic review. *Psychological Medicine: A Journal of Research in Psychiatry and the Allied Sciences, 37*(5), 615-626.

Kendler, K.S., Neale, M. C., Kessler, R. C., Heath, A. C., & Eaves, L. J. (1993). A test of the equal-environment assumption in twin studies of psychiatric illness. *Behavior Genetics, 23*(1), 21-27.

Ketelaars, M. P., Cuperus, J. M., van Daal, J., Jansonius, K., & Verhoeven, L. (2009). Screening for pragmatic language impairment: The potential of the children's communication checklist. *Research in Developmental Disabilities, 30*(5), 952-960.

Kovas, Y., Hayiou-Thomas, M. E., Oliver, B., Dale, P. S., Bishop, D. V. M., & Plomin, R. (2005). Genetic influences in different aspects of language development: The etiology of language skills in 4.5-year-old twins. *Child Development, 76*(3), 632-651.

Kremen, W. S., Jacobson, K. C., Xian, H., Eisen, S. A., Waterman, B., Toomey, R., et al. (2005). Heritability of word recognition in middle-aged men varies as a function of parental education. *Behavior Genetics, 35*(4), 417-433.

Ku, Y.-M., & Anderson, R. C. (2003). Development of morphological awareness in Chinese and English. *Reading and Writing, 16*(5), 399-422.

Larney, R. (2002). The relationship between early language delay and later difficulties in literacy. *Early Child Development and Care, 172*(2), 183-193.

Lau, Y.-H. J., & McBride-Chang, C. (2005). Home Literacy and Chinese Reading in Hong Kong Children. *Early education and development, 16*(1), 5-22.

Law, Y.-K. (2009). The role of attribution beliefs, motivation and strategy use in Chinese fifth-graders' reading comprehension. *Educational Research, 51*(1), 77-95.

Li, H., Corrie, L. F., & Wong, B. K. M. (2008). Early teaching of Chinese literacy skills and later literacy outcomes. *Early Child Development and Care, 178*(5), 441-459.

Li, H., & Rao, N. (2000). Parental influences on Chinese literacy development: A comparison of preschoolers in Beijing, Hong Kong, and Singapore. *International Journal of Behavioral Development, 24(1),* 82-90.

Li, H., & Rao, N. (2005). Curricular and instructional influences on early literacy attainment: Evidence from Beijing, Hong Kong and Singapore. *International Journal of Early Years Education, 13*(3), 235-253.

Lin, D., McBride-Chang, C., Aram, D., Levin, I., Cheung, R. Y. M., Chow, Y. Y. Y., & Tolchinsky, L. (in press). Maternal mediation of writing in Chinese children. *Language and Cognitive Processes.*

Liu, I. M., Chuang, C. J., & Wang, S. C. (1975). *Frequency count of 40,000 Chinese words.* Taipei: Lucky Books.

Logan, S., & Johnston, R. (2009). Gender differences in reading ability and attitudes: Examining where these differences lie. *Journal of Research in Reading, 32*(2), 199-214.

Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent-variable longitudinal study. *Developmental Psychology, 36*(5), 596-613.

Luke, B., & Keith, L. G. (1992). The contribution of singletons, twins and triplets to low birth weight, infant mortality and handicap in the United States. *Journal of Reproductive Medicine, 37*(8), 661-666.

Lytton, H., Conway, D., & Sauvé, R. (1977). The impact of twinship on parent-child interaction. *Journal of Personality and Social Psychology, 35*(2), 97-107.

Mahony, D., Singson, M., & Mann, V. (2000). Reading ability and sensitivity to morphological relations. *Reading and Writing, 12*(3-4), 191-218.

Manis, F. R., Seidenberg, M. S., & Doi, L. M. (1999). See Dick RAN: Rapid naming and the longitudinal prediction of reading subskills in first and second graders. *Scientific Studies of Reading, 3*(2), 129-157.

Mann, V. A., & Liberman, I. Y. (1984). Phonological awareness and verbal short-term memory. *Journal of Learning Disabilities, 17*(10), 592-599.

Mather, P. L., & Black, K. N. (1984). Hereditary and environmental influences on preschool twins. *Developmental Psychology*, *20*(2), 303–308.

McBride-Chang, C. (2004).*Children literacy development.* London: Edward Arnold/Oxford Press.

McBride-Chang, C., Bialystok, E., Chong, K .Y., & Li, Y. (2004). Levels of phonological awareness in three cultures. *Journal of Experimental Child Psychology, 89*, 93-111.

McBride-Chang, C., Cheung, H., Chow, B. W.-Y., Chow, C. S.-L., & Choi. L. (2006). Metalinguistic skills and vocabulary knowledge in Chinese (L1) and English (L2). *Reading and Writing, 19*(7), 695-716.

McBride-Chang, C., Cho, J.-R., Liu, H., Wagner, R. K., Shu, H., Zhou, A., et al. (2005). Changing models across cultures: Associations of phonological awareness and morphological structure awareness with vocabulary and word recognition in second graders from Beijing, Hong Kong, Korea, and the United States. *Journal of Experimental Child Psychology*, *92*(2), 140-160.

McBride-Chang, C., & Ho, C. S.-H. (2000). Developmental issues in Chinese children's character acquisition. *Journal of Educational Psychology, 92*(1), 50-55.

McBride-Chang, C., & Kail, R. V. (2002). Cross-cultural similarities in the predictors of reading acquisition. *Child Development, 73*(5), 1392-1407.

McBride-Chang, C., Shu, H., Zhou, A., Wat, C. P., & Wagner, R. K. (2003). Morphological awareness uniquely predicts young children's Chinese character recognition. *Journal of Educational Psychology, 95*(4), 743-751.

McGrath, L. M., Pennington, B. F., Willcutt, E. G., Boada, R., Shriberg, L. D., & Smith, S. D. (2007). Gene x environment interactions in speech sound disorder predict language and preliteracy outcomes. *Development and Psychopathology, 19,* 1047-1072.

McGue, M., & Bouchard T. J. (1984). Adjustment of twin data for the effects of age and sex. *Behavior Genetics, 14*(4), 325-343.

Merriman, C. (1924). The intellectual resemblance of twins. *Psychological Monographs, 33,* 1-58.

Molfese, V. J., Modglin, A., & Molfese, D. L. (2003). The role of environment in the development of readings skills: A longitudinal study of preschool and school-age measures. *Journal of Learning Disabilities, 36*(1), 59-67.

Morgan, P. L., & Fuchs, D. (2007). Is there a bidirectional relationship between children's reading skills and reading motivation? *Exceptional Children, 73*(2), 165-183.

Munsinger, J., & Douglass, A. (1976). The syntactic abilities of identical twins, fraternal twins, and their siblings. *Child Development, 47*(1), 40–50.

Muter, V., Hulme, C., Snowling, M. J., & Stevenson, J. (2004). Phonemes, rimes, vocabulary, and grammatical skills as foundations of early reading development: Evidence from a longitudinal study. *Developmental Psychology, 40*(5), 665-681.

Nagy, W. E., & Anderson, R. C. (1999). Metalinguistic awareness and literacy acquisition in different languages. In D. A. Wagner, R. L. Venetzky, & B. Street (Eds.), *Literacy: An international handbook* (pp. 155-160). Boulder, CO: Westview Press.

Nagy, W.E., Kwo-Kealoha, A., Wu, X., Li, W., Anderson, R. C., & Chen, X. (2002). The role of morphological awareness in learning to read Chinese. In W. Li, J. S. Gaffney, & J. L. Packard (Eds.), *Chinese children's reading acquisition.* London: Kluwer Academic Publishers.

Nation, K., & Hulme, C. (1997). Phonemic segmentation, not onset-rime segmentation, predicts early reading and spelling skills. *Reading Research Quarterly, 32*(2), 154-167.

Nation, K., & Snowling, M. J. (2004). Beyond phonological skills: broader language skills contribute to the development of reading. *Journal of Research in Reading, 27*(4), 342-356.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2006). *Mx: Statistical modeling manual (7th ed.)*. Richmond, VA: Department of Psychiatry, Medical College of Virginia.

Norbury, C. F., Nash, M., Baird, G., & Bishop, D. V. M. (2004). Using a parental checklist to identify diagnostic groups in children with communication impairment: a validation of the Children's Communication Checklist - 2. *International Journal of Language & Communication Disorders, 39*(3), 345-364.

Oakhill, J. V., Cain, K., & Bryant, P. E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and Cognitive Processes, 18*(4), 443-468.

Oliver, B. R., Dale, P. S., & Plomin, R. (2005). Predicting literacy at age 7 from preliteracy at age 4 - A longitudinal genetic analysis. *Psychological Science, 16*(11), 861-865.

Paracchini, S., Steer, C. D., Lyn-Buckingham, L., Morris, A. P., Ring, S., Scerri, T., et al. (2008). Association of the KIAA0319 dyslexia susceptibility gene with reading skills in the general population. *The American Journal of Psychiatry, 165*(12), 1576-1584.

Payne, A. C., Whitehurst, G. J., & Angell, A. L. (1994). The role of home literacy environment in the development of language ability in preschool children from low-income families. *Early Childhood Research Quarterly, 9*(3-4), 427-440.

Petrill, S. A., Deater-Deckard, K., Thompson, L. A., DeThorne, L. S., & Schatschneider, C. (2006). Reading skills in early readers: Genetic and shared environmental influences. *Journal of Learning Disabilities, 39*(1), 48-55.

Petrill, S. A., Deater-Deckard, K., Thompson, L. A., Schatschneider, C., Dethorne, L. S., & Vandenbergh, D. J. (2007). Longitudinal genetic analysis of early reading: The Western Reserve Reading Project. *Reading and Writing, 20*(1-2), 127-146.

Petrill, S. A., Pike, A., Price, T., & Plomin, R. (2004). Chaos in the home and socioeconomic status are associated with cognitive development in early childhood: Environmental mediators identified in a genetic design. *Intelligence, 32*(5), 445-460.

Plomin, R., & Kovas, Y. (2005). Generalist genes and learning disabilities. *Psychological Bulletin, 131*(4), 592-617.

Plomin, R., & Neiderhiser, J. M. (1992). Genetics and experience. *Current Directions in Psychological Science, 1*(5), 160-163.

Plomin, R., Reiss, D., Hetherington, E. M., & Howe, G. W. (1994). Nature and nurture: Genetic contributions to measures of the family environment. *Developmental Psychology, 30*(1), 32-43.

Pollard, R. (1995). Ethnic comparison of twinning rates in California. *Human Biology, 67*(6), 921-931.

Purcell, S., & Koenen, K. C. (2005). Environmental mediation and the twin design. *Behavior Genetics, 35*(4), 491-498.

Raven, J. C., Court, J. H., & Raven, J. (1995). *Raven's Coloured Progressive Matrices.* UK: Oxford Psychologists Press.

Raz, I. S., & Bryant, P. (1990). Social background, phonological awareness and children's reading. *British Journal of Developmental Psychology, 8*(3), 209-225.

Rescorla, L. (2005). Age 13 language and reading outcomes in late-talking toddlers. *Journal of Speech, Language, and Hearing Research, 48*(2), 459–472.

Rice, M. L., Smith, S. D., & Gayan, J. (2009). Convergent genetic linkage and associations to language, speech and reading measures in families of probands with Specific Language Impairment. *Journal of Neurodevelopmental Disorders, 1*(4), 264-282.

Rowe, D. C., Jacobson, K. C., & Van den Oord, E. J. C. G. (1999). Genetic and environmental influences on vocabulary IQ: Parental education level as moderator. *Child Development, 70*(5), 1151-1162.

Rutter, M (2006). *Gene and behavior. Nature-nurture interplay explained.* Oxford: Blackwell Publishing.

Rutter, M., & Maughan, B. (2002). School effectiveness findings 1979-2002. *Journal of School Psychology, 40*(6), 451-475.

Rutter, M., Thorpe, K., Greenwood, R., Northstone, K., & Golding, J. (2003). Twins as a natural experiment to study the causes of mild language delay: I: Design; twin-singleton differences in language, and obstetric risks. *Journal of Child Psychology and Psychiatry, 44*(3), 326-341.

Sachse, S., & Von Suchodoletz, W. (2008). Early identification of language assessment or parent report? *Journal of Developmental and Behavioral Pediatrics, 29*(1), 34-41.

Samuelsson, S., Byrne, B., Olson, R. K., Hulslander, J., Wadsworth, S., Corley, R., et al. (2008). Response to early literacy instruction in the United States, Australia, and Scandinavia: A behavioral-genetic analysis. *Learning and Individual Differences, 18*(3), 289-295.

Samuelsson, S., Byrne, B., Quain, P., Wadsworth, S., Corley, R., DeFries, J. C., et al. (2005). Environmental and genetic influences on prereading skills in Australia, Scandinavia, and the United States. *Journal of Educational Psychology, 97*(4), 705-722.

Samuelsson, S., Olson, R., Wadsworth, S., Corley, R., Defries, J. C., Willcutt, E., et al. (2007). Genetic and environmental influences on prereading skills and early reading and spelling development in the United States, Australia, and Scandinavia. *Reading and Writing, 20*(1-2), 51-75.

Sandra, D. (1994). The morphology of the mental lexicon: Internal word structure viewed from a psycholinguistic perspective. *Language and Cognitive Processes, 9*(3), 227-269.

Scarborough, H. S., & Dobrich, W. (1994). On the efficacy of reading to preschoolers. *Developmental Review, 14*(3), 245-302.

Scarr, S. (1992). Developmental theories for the 1990s: Development and individual differences. *Child Development, 63*(1), 1-19.

Scarr, S., & McCartney, K. (1983). How People Make Their Own Environments: A Theory of Genotype → Environment Effects. *Child Development, 54*(2), 424-435.

Schatschneider, C., Fletcher, J. M., Francis, D. J., Carlson, C. D., & Foorman, B. R. (2004). Kindergarten prediction of reading skills: A longitudinal comparative analysis. *Journal* *of Educational Psychology, 96*(2), 265-282.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed developmental model of word recognition and naming. *Psychological Review, 96*(4), 523-568.

Sénéchal, M. (2006). Testing the home literacy model: Parent involvement in kindergarten is differentially related to grade 4 reading comprehension, fluency, spelling, and reading for pleasure. *Scientific Studies of Reading, 10*(1), 59-87.

Sénéchal, M., & LeFevre, J.-A. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development, 73*(2), 445-460.

Share, D. L. (2008). On the Anglocentricities of current reading research and practice: The perils of overreliance on an "outlier" orthography. *Psychological Bulletin, 134*, 584-615.

Shell, D. F., Colvin, C., & Bruning, R. H. (1995). Self-efficacy, attribution, and outcome expectancy mechanisms in reading and writing achievement: Grade-level and achievement-level differences. *Journal of Educational Psychology, 87*(3), 386-398.

Shell, D. F., Murphy, C. C., & Bruning, R. H. (1989). Self-efficacy and outcome expectancy mechanisms in reading and writing achievement. *Journal of Educational Psychology, 81*(1), 91-100.

Shu, H., Chen., X., Anderson, R. C., Wu, N., Xuan, Y. (2003). Properties of school Chinese: Implications for learning to read. *Child development, 74*(1), 27-47.

Shu, H., Peng, H., & McBride-Chang, C. (2008). Phonological awareness in young Chinese children. *Developmental Science, 11*(1), 171-181.

Siegel, L. S. (2008). Morphological awareness skills of English language learners and children with dyslexia. *Topics in Language Disorders, 28*(1), 15-27.

Siok, W. T., & Fletcher, P. (2001). The role of phonological awareness and visual-orthographic skills in Chinese reading acquisition. *Developmental Psychology, 37*(6), 886-899.

Snow, C. E., Burns, M. S., & Griffin, P. (Eds.). (1998). *Preventing reading difficulties in young children*. Washington, D.C.: National Academy Press.

Snowling, M. J., & Hulme, C. (2005). *The science of reading: A Handbook.* Oxford: Blackwell Publishing.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly, 21*(4), 360-407.

Storch, S. A., & Whitehurst, G. J. (2002). Oral language and code-related precursors to reading: Evidence from a longitudinal structural model. *Developmental Psychology, 38*(6), 934-947.

Stuart, M., & Coltheart, M. (1988). Does reading develop in a sequence of stages? *Cognition, 30*(2), 139-181.

Thomas, C. J. (1905). Congenital ‘word blindness’ and its treatment. *Ophtalmoscope, 3,* 380–385.

Thorpe, K., Rutter, M., & Greenwood, R. (2003). Twins as a natural experiment to study the causes of mild language delay: II: Family interaction risk factors. *Journal of Child Psychology and Psychiatry, 44*(3), 342-355.

Tong, S., Caddy, D., & Short, R. V. (1997). Use of dizygotic to monozygotic twinning ratio as a measure of fertility. *Lancet, 349*(9055), 843-845.

Torgesen, J. K., & Burgess, S. R. (1998). Consistency of reading-related phonological processes throughout early childhood: Evidence from longitudinal-correlational and instructional studies. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in* *beginning literacy* (pp. 161-188). Mahwah, NJ: Lawrence Erlbaum Associates.

Torgesen, J. K., Morgan, S. T., & Davis, C. (1992). Effects of two types of phonological awareness training on word learning in kindergarten children. *Journal of Educational Psychology, 84*(3), 364-370.

Torgesen, J. K., Wagner, R. K., Rashotte, C. A., Burgess, S., & Hecht, S. (1997). Contributions of phonological awareness and rapid automatic naming ability to the growth of word-reading skills in second- to fifth-grade children. *Scientific Studies of Reading, 1*(2), 161-185.

Tunmer, W. E., & Nesdale, A. R. (1985). Phonemic segmentation skill and beginning reading. *Journal of Educational Psychology, 77*(4), 417-427.

Turkheimer, E, D'Onofrio, B. M. Maes, H. H., & Eaves, L. J. (2005). Analysis and interpretation of twin studies including measures of the shared environment. *Child Development, 76*(6), 1217-1233.

Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science, 14*(6), 623-628.

Umek, L. M., Podlesek, A., & Fekonja, U. k. (2005). Assessing the home literacy environment: Relationships to child language comprehension and expression. *European Journal of Psychological Assessment, 21*(4), 271-281.

Van der Valk, J. C., van den Oord, E. J. C. G., Verhulst, F. C., & Boomsma, D. I. (2001). Using parental ratings to study the etiology of 3-year-old twins' problem behaviors: Different views or rater bias? *Journal of Child Psychology and Psychiatry, 42*(7), 921-931.

Van Hulle, C. A., Goldsmith, H., & Lemery, K. S. (2004). Genetic, environmental, and gender effects on individual differences in toddler expressive language. *Journal of Speech, Language, and Hearing Research, 47*(4), 904-912.

Vellutino, F., R., & Fletcher, J. M. (2005). Developmental dyslexia. In M. J. Snowling & C. Hulme (Eds.), *The science of reading: A Handbook* (pp.362-378). Oxford: Blackwell Publishing.

Verhoeven, L., & Van Leeuwe, J. (2008). Prediction of the development of reading comprehension: A longitudinal study. *Applied Cognitive Psychology, 22*(3), 407-423.

Viding, E., Spinath, F. M., Price, T. S., Bishop, D. V. M., Dale, P. S., & Plomin, R. (2004). Genetic and environmental influence on language impairment in 4-year-old same-sex and opposite-sex twins. *Journal of Child Psychology and Psychiatry, 45*(2), 315-325.

Wadsworth, S. J., DeFries, J. C., Fulker, D. W., Olson, R. K., et al. (1995). Reading performance and verbal short-term memory: A twin study of reciprocal causation. *Intelligence, 20*(2), 145-167.

Wagner, R. K., & Torgesen, J. K. (1987). The nature of phonological processing and its causal role in the acquisition of reading-skills. *Psychological Bulletin, 101*(2), 192-212.

Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1994). Development of reading-related phonological processing abilities: New evidence of bidirectional causality from a latent variable longitudinal study. *Developmental Psychology, 30*(1), 73-87.

Wagner, R. K., Torgesen, J. K., Rashotte, C. A., Hecht, S. A., Barker, T. A., Burgess, S. R., et al. (1997). Changing relations between phonological processing abilities and word-level reading as children develop from beginning to skilled readers: A 5-year longitudinal study. *Developmental Psychology, 33*(3), 468-479.

Wang, J. H.-Y., & Guthrie, J. T. (2004). Modeling the effects of intrinsic motivation, extrinsic motivation, amount of reading, and past reading achievement on text comprehension between U.S. and Chinese students. *Reading Research Quarterly, 39*(2), 162-186.

Weigel, D. J., Martin, S. S., & Bennett, K. K. (2006). Contributions of the home literacy environment to preschool-aged children's emerging literacy and language skills. *Early Child Development and Care, 176*(3-4), 357-378.

Whitehurst, G. J., Falco, F. L., Lonigan, C., Fischel, J. E., DeBaryshe, B. D., Valdez-Menchaca, M. C., & Caulfield, M. (1988). Accelerating language development through picture-book reading. *Developmental Psychology, 24*(4), 552-559.

Wigfield, A., & Guthrie, J. T. (1997). Relations of children's motivation for reading to the amount and breadth or their reading. *Journal of Educational Psychology, 89*(3), 420-432.

Wigfield, A., Guthrie, J. T., Tonks, S., & Perencevich, K. C. (2004). Children's motivation for reading; Domain specificity and instructional influences. *Journal of Educational Research, 97*(6), 299-309.

Wong, K. Y. N. (2007). *Functional Communication skills of Cantonese-speaking children with autism and specific language impairment: performance on Bishop’s Children’s Communication Checklist- 2*. Unpublished master’s thesis, University of Sheffield, United Kingdom.

Wood, C. (2002). Parent-child preschool activities can affect the development of literacy skills. *Journal of Research in Reading, 25*(3), 241-258.

Wooster, A. D., & Carson, A. (1982). Improving reading and self-concept through communication and social skills training. *British Journal of Guidance & Counselling,* *10*(1), 83-87.

Yang, J. F., McCandliss, B. D., Shu, H., & Zevin, J.D. (2009). Simulating language-specific and language-general effects in a statistical learning model of Chinese reading. *Journal of Memory and Language, 61,* 238-257.

Yliherva, A., Loukusa, S., Väisänen, R., Pyper, A., & Moilanen, I. (2009). Development of communication skills in Finnish pre-school children examined by the Children's Communication Checklist (CCC). *Child Language Teaching & Therapy, 25*(2), 235-249.

Zhou, H., & Salili, F. (2008). Intrinsic reading motivation of Chinese preschoolers and its relationships with home literacy. *International Journal of Psychology, 43*(5), 912-916.

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

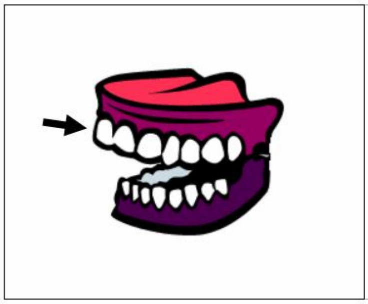
Appendix 1: Measures

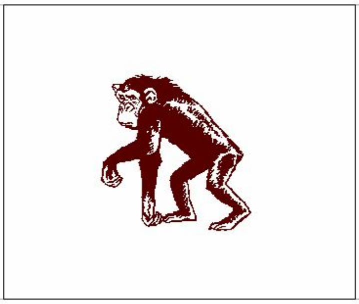
Appendix 1A. An example item of the rhyme awareness task.

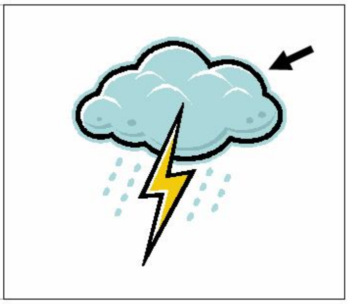
Target syllable: ‘人’ /*jan4*/ (human)

Options (from left to right): ‘牙’ /*ngaa4*/ (tooth), ‘猴’ /*hau4*/ (monkey) and ‘雲’ /*wan4*/ (cloud)

Target answer: ‘雲’ /*wan4*/ (cloud)







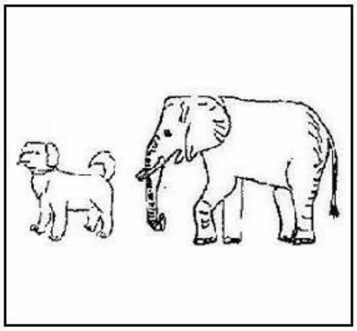
Appendix 1B. The speeded naming task.

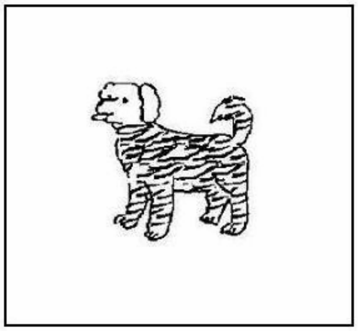
2	5	4	9	7
4	2	7	5	9
5	7	9	4	2
2	4	5	7	9
7	9	2	4	5
9	5	7	2	4

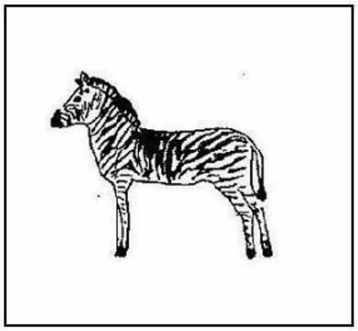
Appendix 1C. An example item of the receptive morphological awareness task.

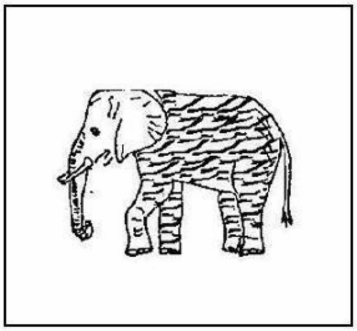
Target item: a striped elephant (斑象 /*baan1 zoeng6*/)

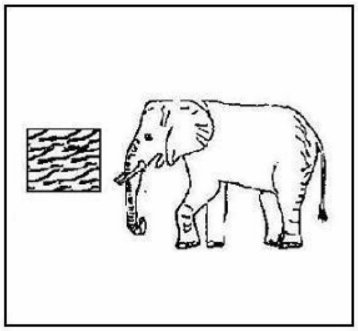
Options (from top left to bottom right): a dog and an elephant (狗+象 / *gau2*/ + /*zoeng6*/); a striped dog (斑狗 /*baan1 gau2*/); a zebra (斑馬/*baan1 maa5*/); a striped elephant (斑象 /*baan1 zoeng6*/); stripes and an elephant (斑+象/*baan1*/ + /*zoeng6*/)









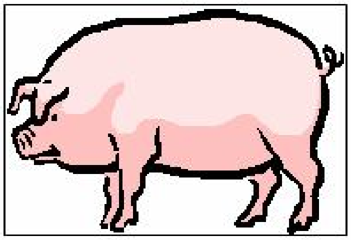


Appendix 1D. An example item of the tone awareness task.

Target tone sound: a Cantonese lexical first tone sound (i.e., high-level tone sound)

Options (from left to right): a letter (信 /*seon3*/), a lock (鎖 /*so2*/) and a pig (豬 /*zyu1*/)

Target answer: a pig (豬 /*zyu1*/)



Appendix 1E. Questionnaire.

[English translation in blankets]

各位家長，多謝填寫此份研究問卷。 請在 ______ 上填寫答案或在 □ 內填上「✓」。如兩位孩子在以下問題有不同的答案，請以孩子 1(雙胞胎中哥哥/姐姐)的情況作答並在該問題的旁邊加上孩子 2 的答案。

[Dear parents, thank you very much for filling in this questionnaire. Please fill in the blank or put a 「✓」in the box □ for the most suitable option. In cases when the situations of the two children are different, please indicate the ratings on the scales provided based on Twin 1's (i.e., the elder twin's) situations, and label Twin 2's situations underneath the scales of the particular questions.]

背景資料
[Background information]

A1) 雙胞胎孩子 1　姓名：____________________　性別：□男　□女
[Twin 1　Name: ___　Sex:　Male/　Female]
雙胞胎孩子 2　姓名：____________________　性別：□男　□女
[Twin 2　Name: ___　Sex:　Male/　Female]

A2) 雙胞胎的出生日期：_______年____月____日
[Twin's date of birth:　__year __month　__day]
填寫問卷日期：_______年____月____日
[Date of completing this questionnaire:　__year __month　__day]

A3) 父親的教育程度：
[Father's education level:]
□小學或以下　□中學：_________年級　□大專或大學　□碩士或以上
[Primary school or below/ Secondary school: ___grade/ Tertiary/　Postgraduate]

A4) 母親的教育程度：
[Mother's education level:]
□小學或以下　□中學：_________年級　□大專或大學　□碩士或以上
[Primary school or below/ Secondary school: ___grade/ Tertiary/　Postgraduate]

A5) 家庭每月平均總收入：
[Family average monthly income:]

□$5,000 或以下 [HK$5000 or below]	☐ $5,001-$10,000 [HK$5001-10000]	☐ $10,001-$15,000 [HK$10001-15000]	☐ $15,001-$20,000 [HK$15001-20000]
☐ $20,001-$25,000 [HK$20001-25000]	☐ $25,001-$30,000 [HK$25001-30000]	☐ $30,001-$35,000 [HK$30001-35000]	☐ $35,001 或以上 [HK$35001 or above]

家庭語文環境
[Home literacy environment]

請根據你對雙胞胎的觀察，在選項前面的 □內填上「✓」。
[Please select the most suitable option according to your observation by putting a 「✓」 in the box □]

B1) 孩子**每月**平均到圖書館的次數：
[Your child's library visit frequency per month]
☐ 沒有 □1 次 □2 次 □3 次 □4 次 □5 次或以上
[None/ once/ twice/ 3 times/ 4 times/ 5 times or above]

B2) 你**每日**平均陪伴孩子做家課和溫習的時間：
[Duration of working with your child on his/her homework per day]
☐ 沒有 □15 分鐘 □半小時 □1 小時 □2 小時 □3 小時或以上
[None/ 15 minutes/ Half an hour/ 1 hour/ 2 hours/ 3 hours or above]

B3) 你和你的配偶**每日**平均看書籍, 報紙或雜誌的時間：
[Duration of you and your spouse reading books, newspapers or magazines per day]
☐ 沒有 □15 分鐘 □半小時 □1 小時 □2 小時 □3 小時或以上
[None/ 15 minutes/ Half an hour/ 1 hour/ 2 hours/ 3 hours or above]

B4) 你從何時開始與孩子一起看**中文**書籍?
[When did you start reading Chinese books with your child?]
□1 歲或之前 □2 歲 □3 歲 □4 歲 □5 歲或以後
[1 year old or before/ 2 years old/ 3 years old/ 4 years old/ 5 years old or after]

B5) 你從何時開始教導孩子**認中文字**?
[When did you start teaching your child to recognize Chinese characters?]
□1 歲或之前 □2 歲 □3 歲 □4 歲 □5 歲或以後
[1 year old or before/ 2 years old/ 3 years old/ 4 years old/ 5 years old or after]

B6) 你從何時開始教導孩子**寫中文字**?
[When did you start teaching your child to write Chinese characters?]
□1 歲或之前 □2 歲 □3 歲 □4 歲 □5 歲或以後
[1 year old or before/ 2 years old/ 3 years old/ 4 years old/ 5 years old or after]

B7) 孩子共有多少本**中文**書籍?
[How many Chinese books does your child own?]
☐ 0-1 本 □2-5 本 □6-10 本 □11-20 本 □21-40 本 □41 本或以上
[0-1/ 2-5/ 6-10/ 11-20/21-40/ 41 or more]

B8) 孩子**每星期**平均看多少本**中文**書籍？
[How many Chinese books does your child read per week?]
☐ 沒有 □1本 □2本 □3本 □4本 □5本或以上
[None/ 1/ 2/ 3/ 4/ 5 or more]

B9) 你**每星期**平均與孩子一起看**中文**書籍的時間：
[Duration of reading Chinese books with you child per week]
□沒有 □半小時 □1小時 □2小時 □3小時 □4小時或以上
[None/ Half an hour/ 1 hour/ 2 hours/ 3 hours/ 4 hours or above]

B10) 你們會不會討論與書中內容有關的問題？
[Do you and your child discuss issues related to the book content?]
□從不 □很少 □間中 □經常 □每次都會
[Never/ Seldom/ Sometimes/ Always/ Every time]

B11) 你**每星期**平均教導孩子認**中文**字的時間：
[Duration of teaching your child to recognize Chinese characters per week]
□沒有 □半小時 □1小時 □2小時 □3小時 □4小時或以上
[None/ Half an hour/ 1 hour/ 2 hours/ 3 hours/ 4 hours or above]

B12) 你**每星期**平均教導孩子寫**中文**字的時間：
[Duration of teaching your child to write Chinese characters per week]
□沒有 □半小時 □1小時 □2小時 □3小時 □4小時或以上
[None/ Half an hour/ 1 hour/ 2 hours/ 3 hours/ 4 hours or above]

B13) 你鼓勵孩子在日常生活中(例如: 告示牌,食物標籤)學習認**中文**字嗎?
[Do you encourage your child to learn characters through daily life experiences (e.g., notice boards, food labels)?]
□從不 □很少 □間中 □經常
[Never/ Seldom/ Sometimes/ Always]

B14) 你會以識字卡或海報幫助孩子學**中文**嗎?
[Do you use flash cards or posters to aid your child’s Chinese learning?]
□從不 □很少 □間中 □經常
[Never/ Seldom/ Sometimes/ Always]

B15) 你會以兒歌或童謠幫助孩子學**中文**嗎?
[Do you use song or nursery rhymes to aid your child’s Chinese learning?]
□從不 □很少 □間中 □經常
[Never/ Seldom/ Sometimes/ Always]

B16) 孩子**每星期**平均參與有關**中文**學習的課外活動(如寫作坊、會話班)的時間：
[Duration of your child’s participation in Chinese learning extracurricular activities (e.g., writing workshops, language playgroups) per week]
□沒有 □半小時 □1 小時 □2 小時 □3 小時 □4 小時或以上
[None/ Half an hour/ 1 hour/ 2 hours/ 3 hours/ 4 hours or above]

B17) 孩子**每星期**平均看有關**中文**語文教育的電視節目和影碟的時間：
[Duration of your child’s educational TV and DVD watching for Chinese learning per week]
□沒有 □半小時 □1 小時 □2 小時 □3 小時 □4 小時或以上
[None/ Half an hour/ 1 hour/ 2 hours/ 3 hours/ 4 hours or above]

B18) 孩子**每星期**平均使用有關**中文**語文教育的電腦軟件的時間：
[Duration of your child’s use of education software for Chinese learning per week]
□沒有 □半小時 □1 小時 □2 小時 □3 小時 □4 小時或以上
[None/ Half an hour/ 1 hour/ 2 hours/ 3 hours/ 4 hours or above]

孩子學習中文的動機
[Child’s motivation for learning Chinese text]

請根據你的觀察，圈出最合適的數字。
[Please circle the most suitable number according to your observation]

		十分不同意[#]						十分同意[##]
C1)	孩子喜歡閱讀**中文**書籍。 [Your child likes reading Chinese books]	1	2	3	4	5	6	7
C2)	孩子喜歡學習**中文**。 [Your child likes learning Chinese]	1	2	3	4	5	6	7
C3)	孩子認為學習**中文**很容易。 [Your child thinks learning Chinese is easy]	1	2	3	4	5	6	7
C4)	孩子對學習**中文**很有信心。 [You child is confident about learning Chinese]	1	2	3	4	5	6	7

[#] [Totally disagree]
[##] [Totally agree]

Appendix 2A. Model Fitting Statistics of Nested Models in Chapter 3.

		Nested models	Comparing nested with full ACE models	
Variables	Parameters	χ^2 (df =4)	$\Delta\chi^2$ (Δdf =1)	p
Word reading	AE	1.87	1.15	.28
Receptive vocabulary	CE	1.55	0.30	.58
Phonological memory	AE	4.81	1.46	.23
Tone awareness	AE	1.35	0.00	1.00
Syllable and rhyme awareness	CE	1.15	0.29	.59
Rapid automatized naming	AE	0.62	0.54	.46
Morphological awareness	AE	2.66	1.31	.25
	CE	3.07	1.72	.19
	E	71.72 (df =5)	69.06 (compared with AE model)	.00
	E	71.72 (df =5)	68.65 (compared with CE model)	.00
Orthographic skills	CE	0.45	0.11	.74
General Language	AE	3.26	3.10	.08
General Reading	AE	2.50	2.00	.16

Note. $\Delta\chi^2$ and Δdf are the differences between the full and the nested models.

Appendix 2B. Model Fitting Statistics Comparing the ACE Mediation Models with Saturated Models in Chapter 4.

Variables	$\Delta\chi^2$ (Δdf =13)	p
SES		
General Language	10.83	.63
General Reading	8.88	.78
Receptive Vocabulary	15.01	.31
Syllable and rhyme awareness	10.70	.64
Orthographic skills	7.27	.89
HLE		
Full model		
General Language	12.81	.46
General Reading	5.85	.95
Receptive Vocabulary	10.61	.64
Syllable and rhyme awareness	8.48	.81
Orthographic skills	4.75	.98

Note. $\Delta\chi^2$ and Δdf are the differences between the ACE mediation models and the saturated models.

Appendix 2C. Model Fitting Statistics of Nested Models in Chapter 5.

	-2LL (full)	*df* (full)	*-2LL* (nested)	*df* (nested)	Δ-*2LL* (Δ*df* =1)	*p*	*-2LL* (nested)	*df* (nested)	Δ-*2LL* (Δ*df* =1)	*p*
			Drop T1/T2 Genetic Link[d]				Drop T1/T2 Genetic Spec[e]			
Word reading	2139.01	1154	2191.90	1155	52.89	.00	2152.74	1155	13.73	.00
Receptive vocabulary	2891.75	1155	2896.05	1156	4.30	.04	2891.76	1156	0.01	.92
Phonological memory	2711.58	1155	2727.81	1156	16.23	.00	2711.61	1156	0.03	.86
Tone awareness	2983.32	1155	2994.17	1156	10.85	.00	2984.96	1156	1.64	.20
Syllable and rhyme awareness	2959.94	1154	2962.43	1155	2.49	.11	2959.94	1155	0.00	1.00
Rapid automatized naming	2708.32	1133	2719.75	1134	11.43	.00	2708.34	1134	0.02	.89
Morphological awareness	3023.20	1155	3027.08	1156	3.88	.05	3023.20	1156	0.00	1.00
Orthographic skills	3084.53	1152	3095.97	1153	11.44	.00	3084.53	1153	0.00	1.00
			Drop T1/T2 Shared Env Link[f]				Drop T1/T2 Shared Env Spec[g]			
Word reading	2139.01	1154	2139.32	1155	0.31	.58	2139.68	1155	0.67	.41
Receptive vocabulary	2891.75	1155	2893.66	1156	1.91	.17	2892.31	1156	0.56	.45
Phonological memory	2711.58	1155	2713.03	1156	1.45	.23	2711.77	1156	0.19	.66
Tone awareness	2983.32	1155	2983.32	1156	0.00	1.00	2983.32	1156	0.00	1.00
Syllable and rhyme awareness	2959.94	1154	2963.31	1155	3.37	.07	2960.40	1155	0.46	.50
Rapid automatized naming	2708.32	1133	2708.32	1134	0.00	1.00	2708.38	1134	0.06	.81
Morphological awareness	3023.20	1155	3023.74	1156	0.54	.46	3024.55	1156	1.35	.25
Orthographic skills	3084.53	1152	3084.76	1153	0.23	.63	3085.49	1153	0.96	.33

Note. Δ-*2LL* and Δ*df* are the differences between the full and the nested models. Superscripts d, e, f, and g denote the nested models with genetic link, genetic specificity, shared environmental link, or shared environmental specificity was dropped respectively.

Appendix 2D. Model Fitting Statistics of Nested Models in Chapter 6.

	χ^2	*df*	$\Delta\chi^2$ (Δdf =1)	*p*
Communicative ability and Word reading				
Full model	2548.41	1224	--	--
Drop Genetic Link	2549.35	1225	0.94	.33
Drop Genetic spec	2580.73	1225	32.32	.00
Drop Shared Env Link	2548.68	1225	0.27	.60
Drop Shared Env Spec	2549.58	1225	1.17	.28
Communicative ability and General Reading				
Full model	2634.72	1196	--	--
Drop Genetic Link	2635.63	1197	0.91	.34
Drop Genetic spec	2643.82	1197	9.10	.00
Drop Shared Env Link	2636.44	1197	1.72	.19
Drop Shared Env Spec	2637.08	1197	2.36	.12
Motivation and Word reading				
Full model	2587.39	1223	--	--
Drop Genetic Link	2592.27	1224	4.88	.03
Drop Genetic spec	2638.07	1224	50.68	.00
Drop Shared Env Link	2587.87	1224	0.48	.49
Drop Shared Env Spec	2588.23	1224	0.84	.36
Motivation and General Reading				
Full model	2683.42	1195	--	--
Drop Genetic Link	2685.29	1196	1.87	.17
Drop Genetic spec	2693.04	1196	9.62	.00
Drop Shared Env Link	2683.54	1196	0.12	.73
Drop Shared Env Spec	2685.61	1196	2.19	.14

Note. $\Delta\chi^2$ and Δdf are the differences between the full and the nested models.

Appendix 3: Mx/ OpenMx scripts

Appendix 3A. Univariate ACE model Mx script.

```
#NGroups 4

G1: model parameters
Calculation
Begin Matrices;
X Lower 1 1 Free        !Additive genetic path coefficient
Y Lower 1 1 Free        !Shared environmental path coefficient
Z Lower 1 1 Free        !Nonshared environmental path coefficient
H Full 1 1              !Fixed scalar in model specification =.5 (half) for DZ’s genetic
relatedness
End Matrices;
Label Row X add_gen
Label Row Y comm_env
Label Row Z spec_env
Matrix H .5
Start .6 All
Begin Algebra;
 A = X*X';              !Additive genetic variance
 C= Y*Y';               !Shared enviornmental variance
 E = Z*Z';              !Nonshared environmental variance
End Algebra;
End;

G2: monozygotic twin pairs
Data NI=2 NO=226        !Number of input variables=2; number of observations=226
CMatrix Full            !Variance/covariance matrix of observed MZ twins’ word reading
scores
0.9932285598040967      0.8444281956518909
0.8444281956518909      0.9887910780778233
Labels  CWR1 CWR2           !Labels for Chinese word reading
Matrices = Group 1
Covariances  A+C+E | A+C_           !Model for MZ variance/covariance
             A+C   | A+C+E /
Option RSiduals
End ;

G3: dizygotic twin pairs
Data NI=2 NO=84             !Number of input variables=2; number of observations=84
CMatrix Full                !Covariance matrix of observed DZ twins’ word reading scores
0.924188896063557       0.5166999785214661
0.5166999785214661      1.082343268868148
Labels  CWR1 CWR2           !Labels for Chinese word reading
```

```
Matrices = Group 1

Covariances  A+C+E | H@A+C+_     !Model for DZ variance/covariance
               H@ A+C     | A+C+E /
Start .6 all
Option RSiduals
End

G4:  Calculation of standardized solution
Calculation
Matrices= Group 1
I Iden 1 1
Begin Algebra;
S= \SQRT(\V2D(\D2V(A+C+E)));    !Standardized total variance
J= S~*A*S~;              !Standardized additive genetic variance
K= S~*C*S~;              !Standardized shared environmental variance
L= S~*E*S~;              !Standardized nonshared environmental variance
End Algebra;
intervals J 1 1          !Confidence interval for standardized additive genetic variance
intervals K 1 1          !Confidence interval for standardized shared environmental variance
intervals L 1 1          !Confidence interval for standardized nonshared environmental
variance
End ;
```

Appendix 3B. Cholesky decomposition model Mx script.

```
 #define nvar=2                    !Number of variables= 2
 G1: MODEL PARAMETERS
 Data Calc NGroups=4
 Begin Matrices;
 X Lower nvar nvar Free    !Additive genetic path coefficients
 Y Lower nvar nvar  Free   !Shared environment path coefficients
 Z Lower nvar nvar  Free   !Nonshared environment path coefficients
 End Matrices;
 Begin Algebra;
 A= X*X' ;              !Additive genetic variances
 C= Y*Y' ;              !Shared environmental variances
 E= Z*Z' ;              !Nonshared environmental variances
 End Algebra;
 Bound  -1 1 X 1 1 - X nvar nvar
 Bound  -1 1 Y 1 1 - Y nvar nvar
 Bound -1 1 Z 1 1 - Z nvar nvar
 Start All .1
 End

 G2:  MZ twin pairs
 Data NInput_vars=4 NObservations=210
 !Number of input variables=4; number of observations=210
 CMatrix Full
!Variance/covariance matrix of observed MZ twins' general language and general reading
scores
1.0018019478612434       0.26673253315718126      0.6445218907177825       0.24003436671727524
0.26673253315718126      0.9466321351438346       0.20450357825519727      0.6821112602821744
0.6445218907177825       0.20450357825519727      0.9755374388886991       0.2760844373520133
0.24003436671727524      0.6821112602821744       0.2760844373520133       1.0061893183868253
Labels GL.a GR.a GL.b GR.b
 Matrices= Group 1
 Covariances (A+C+E | A+C _          !Model for MZ variance/covariance
             A+C   | A+C+E) /
 Option RSidual
 End

 G3:  DZ  twin pairs
 Data NInput_vars=4 NObservations=80
 !Number of input variables=4; number of observations=80
 CMatrix Full
!Variance/covariance matrix of observed DZ twins' general language and general reading
scores
0.979458342431211        0.12772819546479908      0.4934221690754721       0.1669113860389371
```

```
0.12772819546479908     1.0386683373815007      0.011467910829128347    0.49053420007542525
0.4934221690754721      0.011467910829128347    1.0352198509841708      0.19235454793730472
0.1669113860389371      0.49053420007542525     0.19235454793730472     0.9906115409790638
Labels GL.a GR.a GL.b GR.b
 Matrices= Group 1
 H Full 1 1       !Fixed scalar in model specification =.5 (half) for DZ's genetic relatedness
 Covariances (A+C+E   | H@A+C _  !Model for DZ variance/covariance
             H@A+C | A+C+E) /
 Matrix H .5
 Start .1 All
 Option RSidual
 Options NDecimals=3
 End

 G4:  Calculation of standardized solution
 Calculation
 Matrices= Group 1
 I Iden nvar nvar
 Begin Algebra;
 V= A+C+E;
 G=  \sqrt(I.V)~*X;     !Standardized additive genetic path coefficients
 K=  \sqrt(I.V)~*Y;     !Standardized shared environmental path coefficients
 L=  \sqrt(I.V)~*Z;     !Standardized nonshared environmental path coefficients
 D= \SQRT(\V2D(\D2V(A+C+E)));              !Standardized total variances
 M=( D~*A*D~);          !Standardized additive genetic variances
 N= (D~*C*D~);          !Standardized shared environmental variances
 O= (D~*E*D~);          !Standardized nonshared environmental variances
 Q = \sqrt( I . A )~ * A * \sqrt( I . A )~;        !Additive genetic correlation
 R = \sqrt( I .C )~ * C * \sqrt( I . C )~;         !Shared environment correlation
 S = \sqrt( I .E )~ * E * \sqrt( I . E )~;         !Nonshared environment correlation
End Algebra;
Option SErrors !Standard errors for X Y and Z

intervals Q 4 2 1        !Confidence intervals for additive genetic correlation
intervals R 4 2 1        !Confidence intervals for shared env correlation
intervals S 4 2 1        !Confidence intervals for nonshared env correlation

 Options Multiple ISSAT
 End
```

Appendix 3C. Univariate ACE model OpenMx script.

```
require(OpenMx)
source("GenEpiHelperFunctions.R")

#Prepare Data
# -----------------------------------------------------------------------
mydatafile='CCC.dat' #Finds the .dat file in directory
alldat=read.table(mydatafile, header = TRUE)
mycols=colnames(alldat)
validmz=alldat$zygo_SNP==1 #MZ takes a value of 1 on zygo_SNP
validdz=alldat$zygo_SNP==2  #DZ takes a value of 2 on zygo_SNP
col1=11;col2=12 #Specify column numbers
myMZdata=alldat[validmz,col1:col2]
myDZdata=alldat[validdz,col1:col2]

datasetname=c(mydatafile,mycols[col1]," and ",mycols[col2]) #Identify the variables used
colnames(myMZdata)=c("twin1","twin2")
colnames(myDZdata)=c("twin1","twin2")
nucolnames=colnames(myMZdata)
datasetname
summary(myMZdata)
summary(myDZdata)

cov(myMZdata,use="complete")
cov(myDZdata,use="complete")

#Count number of data observations (discarding missing); will be used for DF calculation
MZlist=as.vector(as.matrix(myMZdata))
MZfinite=MZlist[is.finite(MZlist)]
DZlist=as.vector(as.matrix(myDZdata))
DZfinite=DZlist[is.finite(DZlist)]
Nobs=length(MZfinite)+length(DZfinite)

# -----------------------------------------------------------------------
#Fit Saturated Model with RawData and Matrices Input
# -----------------------------------------------------------------------

# Model specification starts here
mytwinSatModel <- mxModel("twinSat",
        mxModel("MZ",
                mxMatrix(type="Full", nrow=1,ncol= 2,
free=TRUE,values=c(0,0),name="expMeanMZ"),
                mxMatrix("Lower",nrow= 2,ncol=2,free=TRUE,values=.5,name="CholMZ"),
```

```
        mxAlgebra(CholMZ %*% t(CholMZ), name="expCovMZ"),
        mxData(myMZdata, type="raw"),
        mxFIMLObjective("expCovMZ", "expMeanMZ", nucolnames)
        ),

    mxModel("DZ",
        mxMatrix(type="Full", nrow=1,ncol= 2, free=TRUE, values=c(0,0), name="expMeanDZ"),
        mxMatrix(type="Lower", nrow=2, ncol=2, free=TRUE, values=.5, name="CholDZ"),
        mxAlgebra(CholDZ %*% t(CholDZ), name="expCovDZ"),
        mxData(myDZdata, type="raw"),
        mxFIMLObjective("expCovDZ", "expMeanDZ", nucolnames)
        ),

    mxAlgebra(MZ.objective + DZ.objective, name="twin"), #Adds together likelihoods for MZ and DZ groups
    mxAlgebraObjective("twin"))  #Evaluate expression from mxAlgebra, i.e. both submodels together
#-------------------------------------------------------------------------
mytwinSatFit <- mxRun(mytwinSatModel) #The mxRun command evaluates the model.

LL_Sat <- mxEval(objective, mytwinSatFit)
summary(mxRun(mytwinSatModel))
#-------------------------------------------------------------------------
DF_Sat=Nobs-nrow(mytwinSatFit@output$standardErrors) #Compute DF

# -----------------------------------------------------------------------
# Fit ACE Model with RawData and Matrices Input
# -----------------------------------------------------------------------
twinACEModel <- mxModel("twinACE",
        # Matrices X, Y, and Z to store additive genetic, shared env, and nonshared env path coefficients
        mxMatrix(type="Full", nrow=1,ncol=1,free=TRUE, values=.6,label="a",name="X",lbound=0),
        mxMatrix(type="Full", nrow=1, ncol=1, free=TRUE,values=.6,label="c",name="Y",lbound=0),
        mxMatrix(type="Full", nrow=1, ncol=1, free=TRUE,values=.6,label="e",name="Z",lbound=0),
    # Matrices A, C, and E compute variance components
        mxAlgebra(expression=X %*% t(X), name="A"),
        mxAlgebra(expression=Y %*% t(Y), name="C"),
        mxAlgebra(expression=Z %*% t(Z), name="E"),
       mxMatrix(type="Full", nrow=1, ncol=2, free=TRUE, values= 20,label="mean", name="expMean"),
    # Algebra for expected variance/covariance matrix in MZ
```

```
                mxAlgebra(expression= rbind  (cbind(A+C+E , A+C),
                                                        cbind(A+C   , A+C+E)),
                name="expCovMZ"),
        # Algebra for expected variance/covariance matrix in DZ
                        mxAlgebra(expression= rbind  (cbind(A+C+E     , 0.5%x%A+C),
                                                       cbind(0.5%x%A+C , A+C+E)),
                name="expCovDZ"),

        mxModel("MZ",
           mxData(observed=myMZdata, type="raw"),
           mxFIMLObjective(
                        covariance="twinACE.expCovMZ",
                        means="twinACE.expMean",
                        dimnames=nucolnames)),
        mxModel("DZ",
           mxData(observed=myDZdata, type="raw"),
           mxFIMLObjective(
                        covariance="twinACE.expCovDZ",
                        means="twinACE.expMean",
                        dimnames=nucolnames)),
        mxAlgebra(expression=MZ.objective + DZ.objective, name="twin"),
        mxAlgebraObjective("twin")
)

# ----------------------------------------------------------------------
#Run ACE model
# ----------------------------------------------------------------------
twinACEFit <- mxRun(twinACEModel)

DF_ACE=Nobs-nrow(twinACEFit@output$standardErrors)
LL_ACE <- mxEval(objective, twinACEFit)
mychi_ACE= LL_ACE - LL_Sat  #Subtract LL for Saturated model from LL for ACE
mychi_DF_ACE=DF_ACE-DF_Sat  #Subtract DF for Saturated model from DF for ACE
mychi_p_ACE=1-pchisq(mychi_ACE,mychi_DF_ACE) #Compute chi square probability

expMZcov_ACE <- mxEval(expCovMZ, twinACEFit)  #MZ's expected covariance matrix
expDZcov_ACE <- mxEval(expCovDZ, twinACEFit)   #DZ's expected covariance matrix
expMeans_ACE <- mxEval(expMean, twinACEFit)      #Expected mean
A_ACE <- mxEval(a*a, twinACEFit)  #Additive genetic variance
C_ACE <- mxEval(c*c, twinACEFit)  #Shared environmental variance
E_ACE <- mxEval(e*e, twinACEFit)  #Nonshared environmental variance
V <- (A_ACE+C_ACE+E_ACE)        #Total variance
a2_ACE <- A_ACE/V                #Standardized additive genetic variance
c2_ACE <- C_ACE/V                #Standardized shared environmental variance
e2_ACE <- E_ACE/V                #Standardized nonshared environmental variance
```

```
ACE_mySE=round(twinACEFit@output$standardErrors,3)
ACE_myest=round(twinACEFit@output$estimate,3)
ACE_mylower=round(ACE_myest-1.96*ACE_mySE,3) #Confidence intervals lower bound
ACE_myupper=round(ACE_myest+1.96*ACE_mySE,3) #Confidence intervals upper bound
```

Appendix 3D. Univariate ACE mediation model OpenMx script.

```
require(OpenMx)
source("GenEpiHelperFunctions.R")

# ----------------------------------------------------------------------
#Prepare Data
# ----------------------------------------------------------------------
mydatafile='HLEnvironment_ACEF.dat' #Finds the .dat file in directory
alldat=read.csv(mydatafile, header = TRUE)
mycols=colnames(alldat)
validmz=alldat$zygo_SNP==1  #MZ takes a value of 1 on zygo_SNP
 validdz=alldat$zygo_SNP==2  #DZ takes a value of 2 on zygo_SNP

col1=25;col2=26 #Specify column numbers
mycolnums=c(col1,col2) #Numbers for the columns with data for twin1 and twin2
col3=51  #Specify column number for identified environment, i.e. SES
mycolnums=c(mycolnums,col3)
myses=alldat[,col3]

alldat[,col3]=(myses-mean(myses,na.rm=TRUE))/sqrt(var(myses,na.rm=TRUE))
#Standardize identified environment, i.e. SES

#----------------------------------------------------------------------------------------
#  Select subset of data for analysis
#----------------------------------------------------------------------------------------

myMZdata=alldat[validmz,mycolnums]
myDZdata=alldat[validdz,mycolnums]

datasetname=c(mydatafile,mycols[col1]," and ",mycols[col2]) #Identify variables used
colnames(myMZdata)=c("twin1","twin2","iden_env")
colnames(myDZdata)=c("twin1","twin2","iden_env")
nucolnames=colnames(myMZdata)
datasetname
summary(myMZdata)
summary(myDZdata)

cov(myMZdata,use="complete")
cov(myDZdata,use="complete")

# Count number of data observations (discarding missing); will be used for DF calculation
MZlist=as.vector(as.matrix(myMZdata[,1:2]))
MZfinite=MZlist[is.finite(MZlist)]
```

```
DZlist=as.vector(as.matrix(myDZdata[,1:2]))
DZfinite=DZlist[is.finite(DZlist)]
Nobs=length(MZfinite)+length(DZfinite)

# -----------------------------------------------------------------------
#Fit Saturated Model with RawData and Matrices Input
# -----------------------------------------------------------------------

# Model specification starts here
mytwinSatModel <- mxModel("twinSat",
        mxModel("MZ",
                mxMatrix(type="Full", nrow=1,ncol= 3,
free=TRUE,values=c(0,0),name="expMeanMZ"),
                mxMatrix("Lower",nrow= 3,ncol=3,free=TRUE,values=.5,name="CholMZ"),
                mxAlgebra(CholMZ %*% t(CholMZ), name="expCovMZ"),
                mxData(myMZdata, type="raw"),
                mxFIMLObjective("expCovMZ", "expMeanMZ", nucolnames)
                ),

        mxModel("DZ",
                mxMatrix(type="Full", nrow=1,ncol= 3, free=TRUE, values=c(0,0),
name="expMeanDZ"),
                mxMatrix(type="Lower", nrow=3, ncol=3, free=TRUE, values=.5,
name="CholDZ"),
                mxAlgebra(CholDZ %*% t(CholDZ), name="expCovDZ"),
                mxData(myDZdata, type="raw"),
                mxFIMLObjective("expCovDZ", "expMeanDZ", nucolnames)
                ),

        mxAlgebra(MZ.objective + DZ.objective, name="twin"), #Adds together likelihoods
for MZ and DZ groups
        mxAlgebraObjective("twin"))  #Evaluate expression from mxAlgebra, i.e. both
submodels together
#-----------------------------------------------------------------------
mytwinSatFit <- mxRun(mytwinSatModel) #The mxRun command evaluates the model.

LL_Sat <- mxEval(objective, mytwinSatFit)
summary(mxRun(mytwinSatModel))
#-----------------------------------------------------------------------
DF_Sat=Nobs-nrow(mytwinSatFit@output$standardErrors) #Compute DF

# -----------------------------------------------------------------------
# Fit ACEF Model with RawData and Matrices Input
# -----------------------------------------------------------------------
twinACEFModel <- mxModel("twinACEF",
#Matrices X, Y, Z to store additive genetic, share env and nonshared env path coefficients
```

```
# Matrix S to store identified env path coefficient
# Matrix V contains variance of identified env, fixed at one
                mxMatrix(type="Full", nrow=1,ncol=1,free=TRUE,
values=.6,label="a",name="X",lbound=0),
                mxMatrix(type="Full", nrow=1, ncol=1,
free=TRUE,values=.6,label="c",name="Y",lbound=0),
                mxMatrix(type="Full", nrow=1, ncol=1,
free=TRUE,values=.6,label="e",name="Z",lbound=0),
        mxMatrix(type="Full", nrow=1, ncol=1, free=TRUE, values=.6, label="f",name="F"),
        mxMatrix(type="Full", nrow=1, ncol=1, free=FALSE, values=1, label="v",name="V"),
        # Matrices A, C, and E and T compute variance components
                mxAlgebra(expression=X %*% t(X), name="A"),
                mxAlgebra(expression=Y %*% t(Y), name="C"),
                mxAlgebra(expression=Z %*% t(Z), name="E"),
            mxAlgebra(expression=F %*% t(F), name="T"),
                mxMatrix(type="Full", nrow=1, ncol=3, free=TRUE, values= 0,label="mean",
name="expMean"),
        # Algebra for expected variance/covariance matrix in MZ
                mxAlgebra(expression= rbind  (cbind(A+C+E+T , A+C+T, T),
                                                         cbind(A+C+T   , A+C+E+T,T),
                            cbind(T  , T,V)),
                name="expCovMZ"),
        # Algebra for expected variance/covariance matrix in DZ
                mxAlgebra(expression= rbind  (cbind(A+C+E+T     , 0.5%x%A+C+T,T),
                                                        cbind(0.5%x%A+C+T , A+C+E+T, T),
                            cbind(T, T, V)),
                name="expCovDZ"),

        mxModel("MZ",
           mxData(observed=myMZdata, type="raw"),
           mxFIMLObjective(
                        covariance="twinACEF.expCovMZ",
                        means="twinACEF.expMean",
                        dimnames=nucolnames)),
        mxModel("DZ",
           mxData(observed=myDZdata, type="raw"),
           mxFIMLObjective(
                        covariance="twinACEF.expCovDZ",
                        means="twinACEF.expMean",
                        dimnames=nucolnames)),
        mxAlgebra(expression=MZ.objective + DZ.objective, name="twin"),
        mxAlgebraObjective("twin")
)

# ---------------------------------------------------------------------
#Run ACEF model
```

```
# ----------------------------------------------------------------------
twinACEF_Fit <- mxRun(twinACEFModel)

DF_ACEF=Nobs-nrow(twinACEF_Fit@output$standardErrors)
LL_ACEF <- mxEval(objective, twinACEF_Fit)
mychi_ACEF= LL_ACEF - LL_Sat  #Subtract LL for Saturated model from LL for ACEF
mychi_DF_ACEF=DF_ACEF-DF_Sat  #Subtract DF for Saturated model from DF for ACEF
mychi_p_ACEF=1-pchisq(mychi_ACEF,mychi_DF_ACEF) #Compute chi square probability

expMZcov_ACEF <- mxEval(expCovMZ, twinACEF_Fit) #MZ's expected covariance matrix
expDZcov_ACEF <- mxEval(expCovDZ, twinACEF_Fit)  #DZ's expected covariance matrix
expMeans_ACEF <- mxEval(expMean, twinACEF_Fit)     #Expected mean
A_ACEF <- mxEval(a*a, twinACEF_Fit)    #Additive genetic variance
C_ACEF <- mxEval(c*c, twinACEF_Fit)    #Shared environmental variance
E_ACEF <- mxEval(e*e, twinACEF_Fit)    #Nonshared environmental variance
F_ACEF <- mxEval(f*f,twinACEF_Fit)      #Identified environmental variance
VV <- (A_ACEF+C_ACEF+E_ACEF+F_ACEF)    #Total variance
a2_ACEF <- A_ACEF/VV                   #Standardized additive genetic variance
c2_ACEF <- C_ACEF/VV                   #Standardized shared environmental variance
e2_ACEF <- E_ACEF/VV                   #Standardized nonshared environmental variance
f2_ACEF <- F_ACEF/VV                   #Standardized identified environmental variance

ACEF_mySE=round(twinACEF_Fit@output$standardErrors,3)
ACEF_myest=round(twinACEF_Fit@output$estimate,3)
ACEF_mylower=round(ACEF_myest-1.96*ACEF_mySE,3)
ACEF_myupper=round(ACEF_myest+1.96*ACEF_mySE,3)
```

Appendix 3E. Univariate ACE homogenous and heterogeneous model comparison Open Mx script.

```
require(OpenMx)
source("GenEpiHelperFunctions.R")

#------------------------------------------------------------------------------
# Specify name of data file, columns with data for twin 1 and twin 2, hetergroup cutoff
#------------------------------------------------------------------------------
col1=25;col2=26 #Specify column numbers
hetergroupcutoff=0; #Split sample at SES=0 (mean)
runtype=1;
#Set to 1 for analysis of original data; Set to 2 for analysis that equates variances across
groups

mydatafile='HLEnvironment.dat' #Finds the .dat file in directory 'myData'

#---------------------------------------------------------
#Prepare Data
#---------------------------------------------------------
alldat=read.table(mydatafile, header = TRUE)
mycols=colnames(alldat)
attach(alldat)

#---------------------------------------------------------
#  Select to give SES and zygosity groups
#---------------------------------------------------------
myMZdata_1=alldat[zygo_SNP==1 & SES<hetergroupcutoff,col1:col2];   #LowSES MZ
myMZdata_2=alldat[zygo_SNP==1 & SES>=hetergroupcutoff,col1:col2]; #HighSES MZ
myDZdata_1=alldat[zygo_SNP==2 & SES<hetergroupcutoff,col1:col2];   #LowSES DZ
myDZdata_2=alldat[zygo_SNP==2 & SES>=hetergroupcutoff,col1:col2]; #HighSES DZ

detach(alldat);

datasetname=c(mydatafile,mycols[col1]," and ",mycols[col2]) #Identify the variables used in
analysis
colnames(myMZdata_1)=c("twin1","twin2")
colnames(myDZdata_1)=c("twin1","twin2")
colnames(myMZdata_2)=c("twin1","twin2")
colnames(myDZdata_2)=c("twin1","twin2")

nucolnames=colnames(myMZdata_2)
```

```
 if (runtype==2) # with runtype 2, recompute values after equating variances
  {
  all1=rbind(myMZdata_1,myDZdata_1)
  all2=rbind(myMZdata_2,myDZdata_2)
  myMZdata_1=(myMZdata_1-mean(all1)/sd(all1))
  myDZdata_1=(myDZdata_1-mean(all1)/sd(all1))
  myMZdata_2=(myMZdata_2-mean(all2)/sd(all2))
  myDZdata_2=(myDZdata_2-mean(all2)/sd(all2))
   }

#-------------------------------------------------------------------------------------
# Count number of data observations (discarding missing); will be used for DF calculation
#-------------------------------------------------------------------------------------


MZlist_1=as.vector(as.matrix(myMZdata_1))
MZfinite_1=MZlist_1[is.finite(MZlist_1)]
DZlist_1=as.vector(as.matrix(myDZdata_1))
DZfinite_1=DZlist_1[is.finite(DZlist_1)]
MZlist_2=as.vector(as.matrix(myMZdata_2))
MZfinite_2=MZlist_2[is.finite(MZlist_2)]
DZlist_2=as.vector(as.matrix(myDZdata_2))
DZfinite_2=DZlist_2[is.finite(DZlist_2)]
Nobs=length(MZfinite_1)+length(DZfinite_1)+length(MZfinite_2)+length(DZfinite_2)

#-------------------------------------------------------
# Print out summaries for each group
#-------------------------------------------------------
datasetname
myprefix="N = ";
print(summary(myMZdata_1))
print(myprefix)
print(length(MZfinite_1)/2)
print(summary(myDZdata_1))
print(myprefix)
print(length(DZfinite_1)/2)
print(summary(myMZdata_2))
print(myprefix)
print(length(MZfinite_2)/2)
print(summary(myDZdata_2))
print(myprefix)
print(length(DZfinite_2)/2)


cov(myMZdata_1,use="complete")
cov(myDZdata_1,use="complete")
```

```
cov(myMZdata_2,use="complete")
cov(myDZdata_2,use="complete")

startmean=mean(myMZdata_1[1]) #Use values from MZ1 group as start values for mean
startcov=mean(cov(myMZdata_1)) #Use covs from MZ1 group as start values for covs

# -----------------------------------------------------------------------------
#Fit Saturated Model with RawData and Matrices Input; 4 groups MZ1 MZ2 DZ1 DZ2
# NB assumes equal means
# -----------------------------------------------------------------------------


# Model specification starts here
mytwinSatModel <- mxModel("twinSat",

        mxModel("MZ1",
                mxMatrix(type="Full", nrow=1,ncol= 2, free=TRUE, values=startmean,
name="expMean"),
                mxMatrix("Lower",nrow=
2,ncol=2,free=TRUE,values=startcov,name="CholMZ1"),
                mxAlgebra(CholMZ1 %*% t(CholMZ1), name="expCovMZ1"),
                mxData(myMZdata_1, type="raw"),
                mxFIMLObjective("expCovMZ1", "expMean", nucolnames)
                ),

        mxModel("DZ1",
                mxMatrix(type="Full", nrow=1,ncol= 2, free=TRUE, values=startmean,
name="expMean"),
                mxMatrix(type="Lower", nrow=2, ncol=2, free=TRUE, values=startcov,
name="CholDZ1"),
                mxAlgebra(CholDZ1 %*% t(CholDZ1), name="expCovDZ1"),
                mxData(myDZdata_1, type="raw"),
                mxFIMLObjective("expCovDZ1", "expMean", nucolnames)
                ),
        mxModel("MZ2",
                mxMatrix(type="Full", nrow=1,ncol= 2,
free=TRUE,values=startmean,name="expMean"),
                mxMatrix("Lower",nrow=
2,ncol=2,free=TRUE,values=startcov,name="CholMZ2"),
                mxAlgebra(CholMZ2 %*% t(CholMZ2), name="expCovMZ2"),
                mxData(myMZdata_2, type="raw"),
                mxFIMLObjective("expCovMZ2", "expMean", nucolnames)
                ),

        mxModel("DZ2",
```

```
        mxMatrix(type="Full", nrow=1,ncol= 2, free=TRUE, values=startmean,
name="expMean"),
        mxMatrix(type="Lower", nrow=2, ncol=2, free=TRUE, values=startcov,
name="CholDZ2"),
        mxAlgebra(CholDZ2 %*% t(CholDZ2), name="expCovDZ2"),
        mxData(myDZdata_2, type="raw"),
        mxFIMLObjective("expCovDZ2", "expMean", nucolnames)
        ),
    mxAlgebra(MZ2.objective + DZ2.objective+MZ1.objective+DZ1.objective,
name="twin"),
    mxAlgebraObjective("twin"))  #Evaluate expression from mxAlgebra, i.e. both
submodels together
#-----------------------------------------------------------------------------------
mytwinSatFit <- mxRun(mytwinSatModel) #The mxRun command evaluates the model.

LL_Sat <- mxEval(objective, mytwinSatFit)
summary(mxRun(mytwinSatModel))
#-----------------------------------------------------------------------------------
# compute DF for this model # N observations (all rows and variables, minus N estimated
parameters)
DF_Sat=Nobs-nrow(mytwinSatFit@output$standardErrors)



# -----------------------------------------------------------------------------------
# Fit ACE Model with RawData and Matrices Input, different a c e for low and high SES
(Heterogeneous model)
# -----------------------------------------------------------------------------------
twinACEModel_het <- mxModel("twinACE_het", #heterogeneous
        mxModel("group1",
        #Matrices X1, Y1, and Z1 to store a, c, and e path coefficients for low SES
group
            mxMatrix(type="Full", nrow=1,ncol=1,free=TRUE,
values=startcov,label="a1",name="X1",lbound=0),
            mxMatrix(type="Full", nrow=1, ncol=1,
free=TRUE,values=startcov,label="c1",name="Y1",lbound=0),
            mxMatrix(type="Full", nrow=1, ncol=1,
free=TRUE,values=startcov,label="e1",name="Z1",lbound=0),
        #Matrices A, C, and E compute variance components
            mxAlgebra(expression=X1 %*% t(X1), name="A1"),
            mxAlgebra(expression=Y1 %*% t(Y1), name="C1"),
            mxAlgebra(expression=Z1 %*% t(Z1), name="E1"),
            mxMatrix(type="Full", nrow=1, ncol=2, free=TRUE, values=
startmean,label="mean", name="expMean"),

        #Algebra for expected variance/covariance matrix in MZ1
```

```
                        mxAlgebra(expression= rbind  (cbind(A1+C1+E1 , A1+C1),
                                                  cbind(A1+C1   , A1+C1+E1)),
                        name="expCovMZ1"),
                #Algebra for expected variance/covariance matrix in DZ1
                mxAlgebra(expression= rbind  (cbind(A1+C1+E1     , 0.5%x%A1+C1),
                                            cbind(0.5%x%A1+C1 , A1+C1+E1)),
                        name="expCovDZ1"),
                        mxModel("MZ1",
                                mxData(observed=myMZdata_1, type="raw"),
                                mxFIMLObjective(
                                covariance="group1.expCovMZ1",
                                means="group1.expMean",
                                dimnames=nucolnames)),
                        mxModel("DZ1",
                            mxData(observed=myDZdata_1, type="raw"),
                                mxFIMLObjective(
                                covariance="group1.expCovDZ1",
                            means="group1.expMean",
                                dimnames=nucolnames)),
        mxAlgebra(expression=MZ1.objective+DZ1.objective,name="twin_group1"),
        mxAlgebraObjective("twin_group1")),

                mxModel("group2",
                #Matrices X1, Y1, and Z1 to store a, c, and e path coefficients for high SES
group
                        mxMatrix(type="Full", nrow=1,ncol=1,free=TRUE,
values=startcov,label="a2",name="X2",lbound=0),
                        mxMatrix(type="Full", nrow=1, ncol=1,
free=TRUE,values=startcov,label="c2",name="Y2",lbound=0),
                        mxMatrix(type="Full", nrow=1, ncol=1,
free=TRUE,values=startcov,label="e2",name="Z2",lbound=0),
                #Matrices A, C, and E compute variance components
                        mxAlgebra(expression=X2 %*% t(X2), name="A2"),
                        mxAlgebra(expression=Y2 %*% t(Y2), name="C2"),
                        mxAlgebra(expression=Z2 %*% t(Z2), name="E2"),
            mxMatrix(type="Full", nrow=1, ncol=2, free=TRUE, values=
startmean,label="mean", name="expMean"),
                        #Algebra for expected variance/covariance matrix in MZ2
                        mxAlgebra(expression= rbind  (cbind(A2+C2+E2 , A2+C2),
                                                  cbind(A2+C2   , A2+C2+E2)),
                        name="expCovMZ2"),
                #Algebra for expected variance/covariance matrix in DZ2
                mxAlgebra(expression= rbind  (cbind(A2+C2+E2     , 0.5%x%A2+C2),
                                            cbind(0.5%x%A2+C2 , A2+C2+E2)),
                        name="expCovDZ2"),
                        mxModel("MZ2",
```

```
                                mxData(observed=myMZdata_2, type="raw"),
                                mxFIMLObjective(
                                covariance="group2.expCovMZ2",
                                means="group2.expMean",
                                dimnames=nucolnames)),
                        mxModel("DZ2",
                            mxData(observed=myDZdata_2, type="raw"),
                                mxFIMLObjective(
                                covariance="group2.expCovDZ2",
                                means="group2.expMean",
                                dimnames=nucolnames)),
        mxAlgebra(expression=MZ2.objective+DZ2.objective,name="twin_group2"),
        mxAlgebraObjective("twin_group2")),

        mxAlgebra(expression=MZ1.objective +
DZ1.objective+MZ2.objective+DZ2.objective, name="twin_all"),
        mxAlgebraObjective("twin_all"))

# -----------------------------------------------------------------------
#Run ACE model for two groups
# -----------------------------------------------------------------------
twinACEFit_het <- mxRun(twinACEModel_het)

DF_ACE_het=Nobs-nrow(twinACEFit_het@output$standardErrors)
LL_ACE_het <- mxEval(objective, twinACEFit_het)
mychi_ACE_het= LL_ACE_het - LL_Sat  #Subtract LL for Saturated model from LL for
ACE
mychi_DF_ACE_het=DF_ACE_het-DF_Sat  #Subtract DF for Saturated model from DF for
ACE
mychi_p_ACE_het=1-pchisq(mychi_ACE_het,mychi_DF_ACE_het) #Compute chi square
probability

expMZcov_ACE_1 <- mxEval(group1.expCovMZ1, twinACEFit_het)
#MZ's expected covariance matrix (low SES group)
expDZcov_ACE_1 <- mxEval(group1.expCovDZ1, twinACEFit_het)
#DZ's expected covariance matrix (low SES group)
A_ACE_1 <- mxEval(a1*a1, twinACEFit_het)
#Additive genetic variance genetic variance (low SES group)
C_ACE_1 <- mxEval(c1*c1, twinACEFit_het)
#Shared environmental variance (low SES group)
E_ACE_1 <- mxEval(e1*e1, twinACEFit_het)
#Nonshared environmental variance (low SES group)
V <- (A_ACE_1+C_ACE_1+E_ACE_1)      #Total variance (low SES group)
a2_ACE_1 <- A_ACE_1/V   #Standardized additive genetic variance (low SES group)
c2_ACE_1 <- C_ACE_1/V   #Standardized shared environmental variance (low SES group)
```

```
e2_ACE_1 <- E_ACE_1/V    #Standardized nonshared environmental variance (low SES
group)

expMZcov_ACE_2 <- mxEval(group2.expCovMZ2, twinACEFit_het)
#MZ's expected covariance matrix (high SES group)
expDZcov_ACE_2 <- mxEval(group2.expCovDZ2, twinACEFit_het)
#DZ's expected covariance matrix (high SES group)
A_ACE_2 <- mxEval(a2*a2, twinACEFit_het)
#Additive genetic variance genetic variance (high SES group)
C_ACE_2 <- mxEval(c2*c2, twinACEFit_het)
#Shared environmental variance (high SES group)
E_ACE_2 <- mxEval(e2*e2, twinACEFit_het)
#Nonshared environmental variance (high SES group)
V <- (A_ACE_2+C_ACE_2+E_ACE_2)      #Total variance (high SES group)
a2_ACE_2 <- A_ACE_2/V    #Standardized additive genetic variance (high SES group)
c2_ACE_2 <- C_ACE_2/V    #Standardized shared environmental variance (high SES group)
e2_ACE_2 <- E_ACE_2/V    #Standardized nonshared environmental variance (high SES
group)

ACE_mySE_het=round(twinACEFit_het@output$standardErrors,3)
ACE_myest_het=round(twinACEFit_het@output$estimate,3)
ACE_mylower_het=round(ACE_myest_het-1.96*ACE_mySE_het,3)
ACE_myupper_het=round(ACE_myest_het+1.96*ACE_mySE_het,3)

# -----------------------------------------------------------------------------------
# Fit ACE Model with RawData and Matrices Input, same a c e for low and high SES
(Homogeneous model)
# ------------------------------------------------------------------------------------
twinACEModel_hom <- twinACEModel_het

  twinACEModel_hom[['group2.X2']]@labels=twinACEModel_hom[['group1.X1']]@labels
  twinACEModel_hom[['group2.Y2']]@labels=twinACEModel_hom[['group1.Y1']]@labels
  twinACEModel_hom[['group2.Z2']]@labels=twinACEModel_hom[['group1.Z1']]@labels
#
twinACEModel_hom[['group2.expMean2']]@labels=twinACEModel_hom[['group1.expMean
1']]@labels

#preceding lines equate paths a, c, e and means by equating labels of estimated parameters

twinACEFit <- mxRun(twinACEModel_hom)

LL_ACE <- mxEval(objective, twinACEFit)
DF_ACE=Nobs-nrow(twinACEFit@output$standardErrors)
mychi_ACE= LL_ACE - LL_ACE_het  #Subtract LL for Het model from LL for ACE

mychi_DF_ACE=DF_ACE-DF_ACE_het  #Subtract DF for Het model from DF for ACE
```

```
mychi_p_ACE=1-pchisq(mychi_ACE,mychi_DF_ACE) #Compute chi square probability

expMZcov_ACE <- mxEval(group1.expCovMZ1, twinACEFit)
#MZ's expected covariance matrix
expDZcov_ACE <- mxEval(group1.expCovDZ1, twinACEFit)
#DZ's expected covariance matrix
A_ACE <- mxEval(a1*a1, twinACEFit)
#Additive genetic variance genetic variance
C_ACE <- mxEval(c1*c1, twinACEFit)
#Shared environmental variance
E_ACE <- mxEval(e1*e1, twinACEFit)
#Nonshared environmental variance
V <- (A_ACE_1+C_ACE_1+E_ACE_1)     #Total variance

a2_ACE <- A_ACE_1/V     #Standardized additive genetic variance
c2_ACE <- C_ACE_1/V     #Standardized shared environmental variance
e2_ACE <- E_ACE_1/V     #Standardized nonshared environmental variance

ACE_mySE=round(twinACEFit@output$standardErrors,3)
ACE_myest=round(twinACEFit@output$estimate,3)
ACE_mylower=round(ACE_myest-1.96*ACE_mySE,3)
ACE_myupper=round(ACE_myest+1.96*ACE_mySE,3)
```

Appendix 3F. Cholesky decomposition model OpenMx script.

```
require(OpenMx)
source("GenEpiHelperFunctions.R")
# -----------------------------------------------------------------------
 myfileout="longcwr.xls"  #name for xls file #Specify name for a .xls file to hold output

# -----------------------------------------------------------------------
# Prepare Data
# -----------------------------------------------------------------------
mydatafile='bothwavesIQZ.dat' #Finds the .dat file in directory

alldat=read.table(mydatafile, header = FALSE)
colnames(alldat)=c("family","zygoSNP","CWR1_a","CWR1_b","CVocab1_a","CVocab1_b",
"CPM1_a","CPM1_b","CTA1_a","CTA1_b","CPA1_a","CPA1_b","RAN1_a","RAN1_b","C
MA1_a","CMA1_b","COrtho1_a","COrtho1_b","CWR2_a","CWR2_b","CVocab2_a","CVoc
ab2_b","CPM2_a","CPM2_b","CTA2_a","CTA2_b","CPA2_a","CPA2_b","RAN2_a","RAN
2_b","CMA2_a","CMA2_b","COrtho2_a","COrtho2_b")
mycols=colnames(alldat)

 validmz=alldat$zygoSNP==1  #MZ takes a value of 1 on zygo_SNP
 validdz=alldat$zygoSNP==2   #DZ takes a value of 2 on zygo_SNP
mycolnums=c(3,19,4,20)   #Specify column numbers

mzData=alldat[validmz,mycolnums]
dzData=alldat[validdz,mycolnums]

nv<-2  #Number of variables per twin

selVars = c("CWR1a","CWR2a","CWR1b","CWR2b")       #Select variables
ntv = nv*2 #Number of columns
datasetname= paste(mydatafile,", columns: ",paste(mycols[mycolnums],collapse=","))
originalname=colnames(mzData)
colnames(mzData)= c("CWR1a","CWR2a","CWR1b","CWR2b")
colnames(dzData)=colnames(mzData)

meanstartvalue=0
pathstartvalue=.6

# -----------------------------------------------------------------------
# Fit Multivariate Saturated Model
# -----------------------------------------------------------------------
multivTwinSatModel <- mxModel("multivTwinSat",
   mxModel("MZ",
```

```
        mxMatrix( type="Lower", nrow=ntv, ncol=ntv, free=TRUE, values=pathstartvalue,
name="CholMZ" ),
        mxAlgebra( expression=CholMZ %*% t(CholMZ), name="ExpCovMZ" ),
        mxAlgebra( expression=diag2vec(ExpCovMZ), name="ExpVarMZ"),
        mxMatrix( type="Full", nrow=1, ncol=ntv, free=T, values=meanstartvalue,
name="ExpMeanMZ" ),
        mxData( observed=mzData, type="raw" ),
        mxFIMLObjective( covariance="ExpCovMZ", means="ExpMeanMZ",
dimnames=selVars)
    ),
    mxModel("DZ",
        mxMatrix( type="Lower", nrow=ntv, ncol=ntv, free=TRUE, values=pathstartvalue,
name="CholDZ" ),
        mxAlgebra( expression=CholDZ %*% t(CholDZ), name="ExpCovDZ" ),
        mxAlgebra( expression=diag2vec(ExpCovDZ), name="ExpVarDZ"),
        mxMatrix( type="Full", nrow=1, ncol=ntv, free=T, values=meanstartvalue,
name="ExpMeanDZ" ),
        mxData( observed=dzData, type="raw" ),
        mxFIMLObjective( covariance="ExpCovDZ", means="ExpMeanDZ",
dimnames=selVars)
    ),
    mxAlgebra( MZ.objective + DZ.objective, name="neg2sumLL" ), #optimizes function and
computes -2LL
    mxAlgebraObjective("neg2sumLL")
)

multivTwinSatFit <- mxRun(multivTwinSatModel)
multivTwinSatSumm <- summary(multivTwinSatFit)

# -----------------------------------------------------------------------
# Generate Saturated Output
# -----------------------------------------------------------------------
parameterSpecifications(multivTwinSatFit)
expectedMeansCovariances(multivTwinSatFit)
tableFitStatistics(multivTwinSatFit)

# -----------------------------------------------------------------------
# Fit Multivariate ACE Model with RawData and Matrices Input
# -----------------------------------------------------------------------
multiCholACEModel <- mxModel("multiCholACE",
            mxModel("ACE",
            #Matrices a, c, and e to store additive genetic, shared env and nonshared env path
coefficients
                        mxMatrix( type="Lower", nrow=nv, ncol=nv, free=TRUE,
values=pathstartvalue, name="a" ),
```

```
                mxMatrix( type="Lower", nrow=nv, ncol=nv, free=TRUE,
values=pathstartvalue, name="c" ),
                mxMatrix( type="Lower", nrow=nv, ncol=nv, free=TRUE,
values=pathstartvalue, name="e" ),
        #Matrices A, C, and E compute variance components
                mxAlgebra( expression=a %*% t(a), name="A" ),
                mxAlgebra( expression=c %*% t(c), name="C" ),
                mxAlgebra( expression=e %*% t(e), name="E" ),
        #Algebra to compute total variances and standard deviations (diagonal only)
                mxAlgebra( expression=A+C+E, name="V" ),
                mxMatrix( type="Iden", nrow=nv, ncol=nv, name="I"),
                mxAlgebra( expression=solve(sqrt(I*V)), name="sd"),
        #Matrix & Algebra for expected means vector
                mxMatrix( type="Full", nrow=1, ncol=nv, free=TRUE, values=meanstartvalue,
name="M" ),
                mxAlgebra( expression= cbind(M,M), name="expMean"),
   #Algebra for expected variance/covariance matrix in MZ
        mxAlgebra( expression= rbind  ( cbind(A+C+E , A+C),
                                                                      cbind(A+C   , A+C+E)),
name="expCovMZ" ),
   #Algebra for expected variance/covariance matrix in DZ
        mxAlgebra( expression= rbind  ( cbind(A+C+E     , 0.5%x%A+C),
                                                                      cbind(0.5%x%A+C ,
A+C+E)),  name="expCovDZ" )
                ),
        mxModel("MZ",
           mxData( observed=mzData, type="raw" ),
           mxFIMLObjective( covariance="ACE.expCovMZ", means="ACE.expMean",
dimnames=selVars )
           ),
        mxModel("DZ",
           mxData( observed=dzData, type="raw" ),
           mxFIMLObjective( covariance="ACE.expCovDZ", means="ACE.expMean",
dimnames=selVars )
           ),
    mxAlgebra( expression=MZ.objective + DZ.objective, name="neg2sumLL" ),
        mxAlgebraObjective("neg2sumLL")
)
multiCholACEFit <- mxRun(multiCholACEModel)
multiCholACESumm <- summary(multiCholACEFit)

# -----------------------------------------------------------------------
# Generate Multivariate Cholesky ACE Output
# -----------------------------------------------------------------------
expectedMeansCovariances(multiCholACEFit)
```

```
# -------------------------------------------------------------------------------------
# Print Descriptive Statistics and create formatted table for xls
# -------------------------------------------------------------------------------------
myNMZ=colSums(is.finite(as.matrix(mzData)))
myNDZ=colSums(is.finite(as.matrix(dzData)))
mymeanMZ=round(colMeans(mzData,na.rm=TRUE),3)
mysdMZ=round(sd(mzData,na.rm=TRUE),3)
mymeanDZ=round(colMeans(dzData,na.rm=TRUE),3)
mysdDZ=round(sd(dzData,na.rm=TRUE),3)
mysum=data.frame(cbind(originalname,myNMZ,mymeanMZ,mysdMZ,myNDZ,mymeanDZ,
mysdDZ))
mysum
write.table("Cholesky",myfileout,row.names=F)
myheader=matrix(c(".",colnames(mysum)),nrow=1)
write.table(myheader,myfileout,sep="\t",
append=TRUE,row.names=FALSE,col.names=FALSE)
write.table(mysum,myfileout,sep="\t", append=TRUE,col.names=FALSE)

# -------------------------------------------------------------------------------------
# Find values for expected covariances and write to xls file
# -------------------------------------------------------------------------------------

expcovM=multiCholACEFit@submodels[['ACE']]@algebras[['expCovMZ']]@result
blankbit=matrix(c(".",".",".","."),nrow=1)
blankbit2=matrix(c(".",".","."),ncol=1)
myhead=matrix(c("Expctd Cov" ,"MZ",".","."),nrow=1)
myhead=rbind(blankbit,myhead,colnames(expcovM))
myhead=cbind(blankbit2,myhead)
write.table(myhead,myfileout,sep="\t",append=TRUE ,col.names=FALSE,row.names=FALS
E)
write.table(round(expcovM,3), myfileout,sep="\t",col.names=FALSE,append=TRUE)

expcovD=multiCholACEFit@submodels[['ACE']]@algebras[['expCovDZ']]@result
myhead=matrix(c("Expctd Covs", "DZ",".","."),nrow=1)
myhead=rbind(blankbit,myhead,colnames(expcovD))
myhead=cbind(blankbit2,myhead)
write.table(myhead,myfileout,sep="\t",append=TRUE ,col.names=FALSE,row.names=FALS
E)
write.table(round(expcovD,3), myfileout,sep="\t",col.names=FALSE,append=TRUE)

# -------------------------------------------------------------------------------------
# Find values for expected means and write to xls file
# -------------------------------------------------------------------------------------
meanbit=multiCholACEFit@submodels[['ACE']]@algebras[['expMean']]@result
myhead=matrix(c("Expctd Means",".",".","."),nrow=1)
myhead=rbind(blankbit,myhead,colnames(meanbit))
```

```
write.table(myhead,myfileout,sep="\t",append=TRUE ,col.names=FALSE,row.names=FALSE)
write.table(round(meanbit,3),
myfileout,sep="\t",col.names=FALSE,append=TRUE,row.names=FALSE)

# --------------------------------------------------------------------------------
# Find values for model fit, and write to xls file
# --------------------------------------------------------------------------------
estparams=sum(multiCholACEFit@submodels[['ACE']]@matrices[['a']]@free)+
                    sum(multiCholACEFit@submodels[['ACE']]@matrices[['c']]@free)+
                    sum(multiCholACEFit@submodels[['ACE']]@matrices[['e']]@free)+
                    sum(multiCholACEFit@submodels[['ACE']]@matrices[['M']]@free)

NObs=sum(myNDZ)+sum(myNMZ)
myDF=NObs-estparams
myLL= round(multiCholACEFit@objective@result,2)
AIC=round(myLL-2*myDF,2)
mytablefit=matrix(c("multiCholACE",estparams,myLL,myDF,AIC))
mytablefit=t(mytablefit)
colnames(mytablefit)=c("model","N params","-2LL","DF","AIC")
myhead="Model Fit"
write.table(myhead,myfileout,quote=F,append=T,row.names=F)
write.table(mytablefit,myfileout,sep = "\t",quote=F,append=T,row.names=FALSE)

# --------------------------------------------------------------------------------
# Path estimates
# --------------------------------------------------------------------------------
mya=multiCholACEFit@submodels[['ACE']]@matrices[['a']]@values #Additive genetic path coefficients
myc=multiCholACEFit@submodels[['ACE']]@matrices[['c']]@values  #Shared environmental path coefficients
mye=multiCholACEFit@submodels[['ACE']]@matrices[['e']]@values  #Nonshared environmental path coefficients
myi=multiCholACEFit@submodels[['ACE']]@matrices$I@values
myA=mya%*%t(mya)  #Additive genetic variances
myC=myc%*%t(myc)  #Shared enviornmental variances
myE=mye%*%t(mye)  #Nonshared environmental variances
myv=myA+myC+myE #Sum of unstandardized variances
mystanda=(solve(sqrt(myi*myv))%*%mya) #Standardized unsquared additive genetic paths
mystandc=(solve(sqrt(myi*myv))%*%myc) #standardized unsquared share env paths
mystande=(solve(sqrt(myi*myv))%*%mye) #standardized unsquared nonshared env paths

# --------------------------------------------------------------------------------
# Compute genetic and environmental correlations
# --------------------------------------------------------------------------------
```

```
rg=solve(sqrt(myi*myA))%*%myA%*%solve(sqrt(myi*myA)) #Genetic correlation
rc=solve(sqrt(myi*myC))%*%myC%*%solve(sqrt(myi*myC)) #Shared env correlation
re=solve(sqrt(myi*myE))%*%myE%*%solve(sqrt(myi*myE)) #Nonshared env correlation

# --------------------------------------------------------------------------------
#  Compute standard errors
# --------------------------------------------------------------------------------
myest=multiCholACEFit@output$estimate
mySE=multiCholACEFit@output$standardErrors

mycounter=0
myestimate=matrix(c(1:(3*(nv^2))),nrow=3*nv)
dim(myestimate)=c(nv,nv,3)
myestimate2=myestimate
mystanderr=myestimate
mystanderr2=myestimate
for (mysource in 1:3)
 {
  for (thisrow in 1:nv)
  {
   for (thiscol in 1:nv)
    {
    if (thiscol<=thisrow)
      {mycounter=mycounter+1
      myestimate[thisrow,thiscol,mysource]=myest[mycounter]
      mystanderr[thisrow,thiscol,mysource]=mySE[mycounter]
      }
    else
      {myestimate[thisrow,thiscol,mysource]=0
       mystanderr[thisrow,thiscol,mysource]=NaN
      }
    }
   }
  }
allvar=matrix(c(0,0),nrow=1)
for (myn in 1:nv)
 {
 allvar[myn]=sum(myestimate[myn,,]^2)  #Total variance for each variable
 mystanderr2[myn,,]=mystanderr[myn,,]/allvar[myn]
 myestimate2[myn,,]=myestimate[myn,,]/allvar[myn]
 }

mylower=round(myestimate2-1.96*mystanderr2,3)
myupper=round(myestimate2+1.96*mystanderr2,3)
myestimate2=round(myestimate2,3)#round to 3 dec places
mystanderr2=round(mystanderr2,3)
```

```
myCI=paste(mylower,"to",myupper)

write.table("Standardized unsquared path estimates",myfileout,append=T,row.names=F)
header3=matrix(c(".","a1","a2","c1","c2","e1","e2"),nrow=1)
write.table(header3,myfileout,sep="\t",quote=F,append=T,row.names=FALSE,col.names=FA
LSE)
write.table(myestimate2,myfileout,sep="\t",quote=F,append=T,col.names=FALSE)
header4="SEs"
write.table(header4,myfileout,sep="\t",quote=F,append=T,row.names=FALSE,col.names=FA
LSE)
write.table(mystanderr2,myfileout,sep="\t",quote=F,append=T,col.names=FALSE)

# --------------------------------------------------------------------------------
# Write Genetic and env correlations to xls
# --------------------------------------------------------------------------------
myhead="Genetic correlations"
write.table(myhead,myfileout,quote=F,append=T,row.names=F)
write.table(round(rg,3),myfileout,sep =
"\t",quote=F,append=T,row.names=FALSE,col.names=FALSE)
myhead="Shared env. correlations"
write.table(myhead,myfileout,quote=F,append=T,row.names=F)
write.table(round(rc,3),myfileout,sep =
"\t",quote=F,append=T,row.names=FALSE,col.names=FALSE)
myhead="Nonshared env. correlations"
write.table(myhead,myfileout,quote=F,append=T,row.names=F)
write.table(round(re,3),myfileout,sep =
"\t",quote=F,append=T,row.names=FALSE,col.names=FALSE)

print(paste("Estimates and CIs saved in ",myfileout))
print("This file can be opened in xls")
```