Published in final edited form as:

Learn Individ Differ. 2013 April 1; 24: 160–167. doi:10.1016/j.lindif.2012.12.018.

A multivariate twin study of early literacy in Japanese Kana

Keiko K. Fujisawa¹, Sally J. Wadsworth², Shinichiro Kakihana³, Richard K. Olson^{2,5}, John C. DeFries², Brian Byrne^{4,5}, and Juko Ando¹

¹Department of Education, Faculty of Letters, Keio University

²University of Colorado, 1480 30th St Boulder, CO 80303, USA

³Koriyama Women's University, Kaisei 3-25-2, Koriyama city, Fukushima, 963-8503

⁴University of New England, Armidale NSW 2351, Australia

⁵Linköping University, SE-581 83 LINKÖPING, Sweden

Abstract

This first Japanese twin study of early literacy development investigated the extent to which genetic and environmental factors influence individual differences in prereading skills in 238 pairs of twins at 42 months of age. Twin pairs were individually tested on measures of phonological awareness, *kana* letter name/sound knowledge, receptive vocabulary, visual perception, nonword repetition, and digit span. Results obtained from univariate behavioral-genetic analyses yielded little evidence for genetic influences, but substantial shared-environmental influences, for all measures. Phenotypic confirmatory factor analysis suggested three correlated factors: phonological awareness, letter name/sound knowledge, and general prereading skills. Multivariate behavioral genetic analyses confirmed relatively small genetic and substantial shared environmental influences on the factors. The correlations among the three factors were mostly attributable to shared environment. Thus, shared environmental influences play an important role in the early reading development of Japanese children.

Keywords

Japanese Kana syllabary; early literacy; behavioral genetics

1. Introduction

1.1. Prereading skills and later reading abilities

Prereading skills predict later reading abilities (Scaborough, 2001), although some are stronger predictors than others. According to Scaborough (2001), prereading skills that are closely related to word recognition, especially letter-sound knowledge and phonological awareness, are among the best predictors. Although vocabulary, sentence/story recall, rapid naming, and verbal memory measures such as digit recall and nonword repetition are also reliable predictors (e.g., Wagner et al., 1997; de Jong & Olson, 2004; Swanson & Siegel, 2001), nonverbal abilities, such as visual discrimination and motor skills, are not (Scaborough, 2001).

Correlations among prereading skills and their longitudinal associations with later reading abilities have been also reported in non-alphabetical languages. For example, early literacy

skills, including phonological awareness (i.e., mora awareness, described later), vocabulary, visual perceptual skill, and verbal short term memory, are correlated in Japanese children (Inagaki, Hatano, & Otake, 2000; Kakihana, Ando, Koyama, Iitaka, & Sugawara, 2009). Moreover, phonological awareness and rapid number naming are significantly related to word recognition (Cho, McBride-Chang, & Park, 2008), and letter-name knowledge during early literacy development contributes to later word reading in Korean children (Kim & Petscher, 2011). These prereading phonological skills also significantly predict later reading performance in Chinese children (Ho & Bryant, 1997; Tong et al., 2011).

1.2. Genetic and environmental influences on early reading for different orthographies

Several previous studies have assessed genetic and environmental influences on early reading skills, including phonological awareness, verbal memory, vocabulary, and letter knowledge (e.g., Bishop et al., 1999; Dionne, Dale, Boivin, & Plomin, 2003; Hart, et al., 2009; Hayiou-Thomas, et al., 2006; Hohnen & Stevenson, 1999; Kovas, et al., 2005; Petrill, Deater-Deckard, Thompson, DeThorne, & Schatschneider, 2006; Samuelsson et al., 2005, 2007). However, these studies were based on twins who were learning alphabetic languages, mostly English. Thus, relatively little is currently known about genetic and environmental influences on reading development in different orthographies. We are aware of only one behavioral genetic study which focused on a non-alphabetic language. Chow, Ho, Wong, Waye, and Bishop (2011) investigated genetic and environmental influences on reading skills in Chinese-learning children and reported results similar to those found in English acquisition; moderate to substantial genetic influences on word reading, phonological memory, and rapid automatized naming, and moderate to substantial shared environmental influences on receptive vocabulary, syllable and rhyme awareness, and orthographic skills. However, their sample included a wide age range (3 to 11 years old) and did not focus on prereading skills. Thus we have very limited knowledge concerning the genetic and environmental etiologies of individual differences in early reading for different orthographies.

In this study, we focused on Japanese, especially on syllabic *kana* letters. To our knowledge this is the first Japanese twin study of genetic and environmental influences on prereading skills. Japanese is a non-alphabetic writing system originally derived from Chinese, but it is very different from Chinese in various aspects. In the next section, we provide a brief description of the Japanese writing systems and syllabic *kana* letters.

1.3. Japanese writing systems and orthographies

The Japanese writing system uses both logographic *kanji* and syllabic *kana*. In standard Japanese orthography, nouns and stems of verbs and adjectives are usually written in *kanji*, and function words and inflectional affixes are written in *kana* in most cases.

Though its roles and usage in the standard orthography are limited, *kana* is a full-fledged phonetic writing system which can represent any item of Japanese vocabulary. In fact, while mastering the enormous numbers of *kanji* characters, children's early literacy solely depends on the *kana*; children's books and text books are first written only in *kana* letters. The rate of *kanji* usage gradually increases as children advance in grade levels. Thus, mastering *kana* literacy is an independent literacy development.

Most Japanese syllables have a consonant-vowel (CV) or a single vowel (V) structure, which has neither a consonant cluster in the onset position nor a coda consonant. In Japanese phonology, just two types of consonants are allowed in the coda position; one is a nasal consonant [N] (e.g., /hoNda/), and the other is a geminate stop consonant [Q] (e.g., /niQpoN/).

Strictly speaking, each *kana* letter does not represent a syllable but a mora—a syllable-like phonological unit. The mora is a unit with which Japanese speakers segment speech streams (Otake, Hatano, Cutler, & Mehler, 1993). The mora is a syllable nucleus, preceded by a syllable onset, a syllable coda, or an extended portion of the vowel. When a syllable has a [V] or [CV] structure, the syllable has just one mora. However, when a syllable has a nasal coda [N], a geminate stop [Q], and an extended portion of the vowel, the syllable has two (e.g., CVN) or three morae (e.g., CV:N).

The *kana* writing system represents morae in Japanese. There are 46 basic *kana* letters, consisting of 45 letters representing CV and V morae, and one letter representing a mora of the nasal coda [N]. A mora of geminate stop consonant [Q] and an extended portion of the vowel are also given a letter in words, by using a certain basic letter.

There are 103 distinct morae in the Japanese language¹, though there are 46 basic letters. In addition to the basic letters, there are 2 supplementary notational systems to represent remainders of morae: diacritics and combinations. Twenty syllables with voiced stop and fricative are represented by a letter for its unvoiced counterpart with a *daku-ten* diacritic, the two little strokes on the right shoulder of the letter (e.g., か = ka, が = ga). Five syllables with a /p-/ are represented by /h-/ letter with another diacritic, a small circle (*handaku-ten* diacritic) on the right shoulder of the letter (e.g., は = ha, ぱ = pa). Thirty-three CjVs, which exceptionally have a consonant cluster on their onset (like the /kjo/ sound in /kjoto/), are represented by two-letter combinations. For example, /kja/ is written as “きゃ,” which is き (/ki/) with the subscription of ゃ (/ja/). To sum up, *kana* has 46 letters and 2 supplementary notational systems which represent 103 Japanese morae, all of the morae in the Japanese language.

As with children in western countries, Japanese children start their literacy development by memorizing letter names. However, importantly, the roles of letter name knowledge in *kana* literacy development are quite different from those in alphabetic languages. In alphabetic systems, letter names (e.g., /bi/ for B) or “letter sounds” (/bu/ for B) may be quite different from the exact phonetic value (e.g., /b/ for B) of the letters. In contrast, in a syllabic system such as *kana*, each of the letter name/sounds² (e.g., /ne/ for ね) are almost equal to the exact phonetic value of the letters. Thus, all that is needed to read a *kana* word is to sound out each letter name/sound in turn. For example, for ねこ (cat), giving each letter name ね (/ne/) and こ (/ko/) produces the word it represents (/neko/). Consequently, once someone has memorized the 46 letter names and understands the usage of the *daku-ten* diacritic, *handaku-ten* diacritic, and the combinations, they can read all Japanese words. Because of these characteristics of *kana*, letter name/sound knowledge is regarded as a direct measure of children’s reading ability in *kana* reading research (e.g., National Institute for Japanese Language, 1972).

1.4. Present study

In the present study, we investigated the extent to which genetic and environmental influences contribute to prereading skills and to correlations among them in Japanese-learning twins at 42 months of age. Although the children were younger than those in previous behavioral genetic studies of prereading skills (e.g., 4.5-year-olds, Kovas et al., 2005; Hayiou-Thomas et al., 2006; 5-year-olds, Samuelson et al., 2005; 6-year-olds, Petrill

¹One hundred and three morae consist of 69 CV morae, a nasal coda /N/, and 33 yo-on morae (CjV). Recently, however, at least 33 morae seemed to be added to the Japanese language, most of which came from Western languages (see Tamaoka & Makioka, 2004 for details). Most of these are written in *katakana* as CV letters with the subscription of V (e.g., “ティ (/te/i/)” for /ti/), a usage that is analogous to the conventional *yo-on* combination.

²There is no conceptual distinction between “letter name”, “letter sound” in Japanese.

et al., 2006), choosing 42 months of age was appropriate for studying prereading skills in Japanese *kana* because most Japanese children master reading *kana* before they start primary school education (Shimamura & Mikami, 1994). Also, choosing the younger age was expected to increase our knowledge concerning the etiology of early reading for both non-alphabetic languages and alphabetic languages because reading outcomes may be predicted by different sets of language variables at different ages (Scaborough, 2001; e.g., Siok & Fletcher, 2001). We included measures which were as similar as possible to those used in previous studies of alphabetic languages. In addition, we included a measure of visual perceptual skills, though little evidence has been found concerning the relation between visual perceptual skills and later reading in English learning (Scaborough, 2001). Visual perceptual skills are related to other literacy skills in the early phase of reading development and may be significant predictors of word reading in non-alphabetic languages (Chinese: Ho & Bryant, 1997; Siok & Fletcher, 2001; Tong et al., 2011; Korean: Cho et al., 2008; Japanese: Kakahana et al., 2009;).

2. Method

2.1. Participants

Participants were Japanese twins (mean = 3.57 years, SD = 0.14) tested in the Tokyo Twin Cohort Project (ToTCoP; Ando et al., 2006). ToTCoP is a twin registry based on data from Basic Resident Registrations in the Tokyo area (see Ando et al., 2006). The ToTCoP study was approved by the Ethics Committee at the Faculty of Letters, Keio University. Written informed consent was obtained from twins' parents.

Zygosity was determined using a questionnaire that has 95% accuracy based upon physical similarities of twin siblings at around one year of age (Ooki & Asaka, 2004). The sample in the present study included 55 male monozygotic twin pairs (MZm), 70 female MZ pairs (MZf), 60 male-male dizygotic twin pairs (DZm), and 53 female-female DZ pairs (DZf).

2.2. Tasks and procedures

Different trained testers tested the two twins individually and simultaneously in each pair. Kakahana et al. (2009) confirmed that all of the measures used in this study are related to *kana* reading performance of typically developing Japanese children at three and four years of age.

2.2.1. Kana letter name/sound knowledge—In order to shorten testing time, we chose (a) 11 letters from the 46 basic *kana* letters, (b) 3 letters with a *daku-ten* diacritic (voiced letters), and (c) 1 letter from 5 letters with a *handaku-ten* diacritic (/p-/ sound letters). The choice of letters was based on accuracy data for each letter from a large-scale survey (National Language Research Institute, 1972) in order to maximize variance and avoid ceiling and floor effects. Children were asked to read each letter to which a tester pointed. The order of pointing letters was basic letters, voiced letters, and /p-/ sound letters ($\alpha=.96$).

2.2.2. Phonological awareness—Because each *kana* letter represents mora, our phonological awareness task was to measure mora awareness rather than phonemic awareness. We used the mora awareness task in Kakahana et al. (2009), and a mora segmentation task based on Amano (1970) and Inagaki et al. (2000). In the task, children were presented with drawings one at a time and asked to make a doll jump on a series of circles on paper while they articulated the drawing's name. First, a tester introduced a small Snoopy figure, showing how to do the task. The tester said "Let's play a word game today. I will show you how to play it. This is *kamera* [camera], right? I will make this Snoopy doll jump and take steps ahead on these circles". Then the tester moved the Snoopy doll three

circles ahead with voicing “ka/me/ra”. The experimenter then asked the child to do the game in the same way. Nine test items included (a) 3 CV type words (e.g., /kuruma/ [car]), (b) 3 words including a syllable with long vowel (CV:) which has two morae in one syllable (/suka:to/ [skirt]), and (c) 3 words including a geminate stop Q which also has two morae in one syllable (e.g., /roboQto/ [robot]). The child’s segmentation accuracy by mora was used as the test score ($\alpha=.75$).

2.2.3. Nonword repetition—We used the Japanese version of a nonword repetition (NWR) task (Kakihana et al., 2009), following Gathercole et al. (1994), who included nonwords with 2 to 5 syllables. Considering the importance of mora in the Japanese language, our NWR task was created with regard to mora length rather than syllable length. Children were asked to repeat nonwords presented from a CD player. Forty nonwords were used in the task, consisting of 10 nonwords for each of four mora lengths -- 4 morae, 5 morae, 6 morae and 7 morae ($\alpha=.89$).

2.2.4. Receptive vocabulary—We used the subtest of Auditory Reception in the Japanese version of the Illinois Test of Psycholinguistic Abilities (ITPA; Kirk, McCarthy, & Kirk, 1968). A tester said a noun or a verb and children were asked to choose one of four pictures which corresponded to the word. Forty items were the maximum ($\alpha=.78$).

2.2.5. Visual perception skill—We used the subtest Position in Space from the Frostig Developmental Test of Visual Perception (Frostig, 1961). There were two parts. In Part A, children were asked to choose a line drawing which was reversed up and down or right and left from five same shaped line drawings. In Part B, children were asked to choose a line drawing which was the same shape and in the same orientation as the sample line drawing from four drawings. Eight test items were used ($\alpha=.41$).

2.2.6. Digit span—We used the digit span task from the Kaufman Assessment for Children (K-ABC, Kaufman & Kaufman, 1983). This task included two to four forward digits ($\alpha=.77$).

2.3. Analyses

The following analyses were conducted.

2.3.1. Descriptive statistics—Means and standard deviations were calculated for each measure. We performed *t*-tests for each measure to test whether the means differed by gender and by zygosity.

2.3.2. Exploratory and confirmatory factor analyses—To assess the structure of prereading skills, we first performed an exploratory factor analysis with oblique rotation. Based on the results of this exploratory factor analysis, we performed a confirmatory factor analysis to assess the adequacy of its fit to data. Because data from twin pairs are not statistically independent, we performed a confirmatory factor analysis using structural equation modeling based on dyad-level data (Olsen & Kenny, 2006) We used EQS software (Bentler, 2006) for the analysis.

2.3.3. Phenotypic correlations among measures—To assess the phenotypic relationships among the individual and composite measures, we calculated correlations among measures and factors suggested by factor analyses, using the pairwise approach suggested by Griffin and Gonzalez (1995). The correlations were then transformed into *z*-values with adjusted sample size and tested for significance using *Z* tests (Griffin & Gonzalez, 1995).

2.3.4. Twin intraclass and cross-trait/cross-twin correlations and univariate genetic analysis—Behavioral genetic methods assess the extent to which individual differences are due to genetic and environmental influences (Plomin, DeFries, McClearn, & McGuffin, 2008). Environmental influences may be shared, generating similarities among children growing up in the same family, or nonshared, in which case their experiences are statistically independent. Because members of MZ twin pairs are genetically identical, the correlation between additive genetic values is 1.0. Members of DZ twin pairs share half of their segregating genes on average, so the correlation between their additive genetic values is 0.5. In contrast, the correlation between shared environmental influences is 1.0 for both MZ and DZ twin pairs who are reared in the same family; and the correlation between nonshared environmental influences of a twin pair is 0 for both MZ and DZ twin pairs because their nonshared environmental influences are independent, regardless of zygosity. Thus, the intraclass correlation (i.e., a correlation between Twin 1's trait and Twin 2's trait) should be higher for MZ twins than for DZ twins if observed individual differences are due at least in part to genetic influences. The same holds true for a cross-trait/cross-twin correlation; if a cross-trait/cross-twin correlation between two traits (i.e., a correlation between one trait of Twin 1 and another trait of Twin 2) is higher for MZ twins than for DZ twins, it indicates the existence of a genetic contribution to the phenotypic correlation between the two traits.

In order to test for the presence of genetic and environmental influences on each measure and the bivariate relationships among them, we calculated intraclass and cross-twin/cross-trait correlations by zygosity. Following Griffin & Gonzalez (1995), the correlations were transformed into z-values with adjusted sample size and then tested for significance using Z tests. In addition, we performed univariate genetic analyses for each measure using Mx software (Neale, Boker, Xie, & Maes, 2002). For this analysis, the variance of each measure was decomposed into genetic and environmental components using structural equation modeling (Neale & Cardon, 1992).

2.3.5. Multivariate genetic analyses using common pathway model—In order to assess genetic and environmental influences on the correlations among the phenotypic factors (i.e., factors which were obtained in the confirmatory factor analyses that preceded the genetic modeling), the extent to which each measure loaded on factors, and the genetic and environmental influences unique to each measure, a common pathway model with correlated factors was fitted to the data. This model decomposes the variance of each latent factor into genetic and environmental components, and the correlations among the latent factors and the residual variance of each measure are partitioned into parts due to genetic influences and to environmental influences. Also, the relative importance of the latent factors for each measure is evaluated by each measure's loading on factors. Because measurement errors are included in these residual nonshared environment components, nonshared environmental components for latent common factors are error-free. Raw data were modeled using the Mx statistical modeling package (Neale, Boker, Xie & Maes, 2002), which facilitates the use of all available data as well as maximum likelihood estimation of parameters. When analyzing data from twin pairs, twice the negative log-likelihood (−2LL) is calculated for each pair, and then summed across all pairs. For evaluation of alternative models, the change in chi-square is calculated as the difference between −2LL for the two models, with degrees of freedom equal to the difference in the number of free parameters in each model.

3. Results

3.1. Descriptive statistics

Descriptive statistics for raw scores of each task are shown in Table 1. Means did not differ by zygosity or by gender, except for the phonological awareness task including a long sound vowel (Boys: mean = 1.05, SD = 1.08; Girls: mean = 1.28, SD = 1.09, $t = -2.01$, $p < .05$) and receptive vocabulary (Boys: mean = 17.95, SD = 5.13; Girls: mean = 19.14, SD = 5.01, $t = -2.42$, $p < .05$). In order to preclude the possible influence of differences in age, gender and zygosity, each measure was adjusted by age and gender, and then standardized within zygosity. These standardized scores were used in the following analyses.

3.2. Exploratory factor analysis

Two factors were obtained with eigenvalues greater than 1.0. Factor loadings for each measure are shown in Table 2. The first factor is labeled “phonological awareness” because it includes the three types of words for the phonological awareness task (eigenvalue = 2.65, accounting for 33.17% of the variance). The second factor is labeled “general prereading skills”, because it includes all measures except for those of phonological awareness and letter name/sound knowledge (eigenvalue = 1.25, accounting for 15.62% of the variance). The correlation between the two factors was .52.

3.3. Confirmatory factor analysis

As shown in Table 2, the *kana* letter name/sound knowledge measure loaded on both factors, although the factor loadings were not large. Thus, a three correlated factors model may fit the data better than the two-factor model, in which the three factors are phonological awareness, general prereading skills, and letter name/sound knowledge. The three types of words for phonological awareness would load on the first factor, other measures of prereading skills would load on the second factor, and *kana* letter name/sound knowledge would load on the third factor. This model is plausible given the distinct role of letter name/sound knowledge in Japanese *kana* reading. Thus, we performed a confirmatory factor analysis to compare the fits of the two possible models.

The result of this confirmatory factor analysis showed that both models fit the data well but that the three correlated factors model fit the data significantly better (two factors model: $\chi^2 = 80.41$, adjusted df = 43, $p = .99$, AIC = -146.70; three factors model: $\chi^2 = 62.54$, adjusted df = 39, $p = .99$, AIC = -159.47). Moderate correlations exist among the three factors: .49 between phonological awareness and general prereading skills; .42 between phonological awareness and letter name/sound knowledge; and .39 between general prereading skills and letter name/sound knowledge. Based on these results, subsequent analyses used the three correlated factors model.

We created a composite score for the phonological awareness factor using a weighted average based on the factor loadings of the indicators for the three types of words adjusted by age and gender, and then standardized within zygosity. Similarly, we created a composite score for the general prereading skills factor by calculating the weighted average of visual perceptual skills, receptive vocabulary, digit span and nonword repetition adjusted by age and gender and then standardizing them within zygosity. The single measure of letter name/sound knowledge defined the third correlated factor. These scores were then used to compute phenotypic correlations, intraclass and cross-trait/cross-twin correlations, and variance components among the three factors and indicators.

3.4. Phenotypic correlations among variables

Phenotypic correlations among the factors and prereading measures are shown in Table 3. Phonological awareness, general prereading skills, and letter name/sound knowledge factors, and each individual task, were all significantly correlated, except for that between phonological awareness of words including a geminate stop and visual perception skill.

3.5. Intraclass and cross-trait/cross-twin correlations and variance component estimates

Intraclass and cross-trait/cross-twin correlations are shown in Table 4. Most MZ and DZ correlations did not differ substantially, except for the intraclass correlations for letter name/sound knowledge and for phonological awareness of words including long vowel, which were significantly higher for MZ than DZ twin pairs for the former ($z = 2.20, p < .05$) and significantly higher for DZ than MZ twin pairs for the latter ($z = -2.27, p < .05$). These correlations suggest that genetic influences on the individual differences in these measures are relatively small, whereas shared environmental influences are substantial. In addition to the significant contrast for letter name/sound knowledge, the intraclass correlations for receptive vocabulary, nonword repetition and digit were slightly higher for MZ twin pairs than for DZ twin pairs. These correlations suggest the presence of at least some genetic influences on the individual differences in these measures.

These results are consistent with those of the univariate genetic analyses for each measure. As shown in Table 5, the estimates of the proportion of variance due to genetic influences on the individual differences (a^2) were zero for phonological awareness, general prereading skill, and visual perception skill, and were small for letter name/sound knowledge, receptive vocabulary, nonword repetition and digit span. Only a^2 for letter name/sound knowledge was significant. In contrast, estimates for shared environmental influences (c^2) were moderate to substantial.

3.6. Common pathway model with correlated factors

We used the scores of the measures adjusted by age and gender and then standardized within zygosity as input for the common pathway model with correlated factors.

Results of fitting the common pathway model with correlated factors to the data are shown in Figure 1. Variance component estimates for the correlated factors and factor loadings of indicators are shown in Table 6. The model fit the data well ($\chi^2 = 272.96, df = 223, p = .99, AIC = -173.05$). Genetic influences on the three factors are relatively small, but the a^2 estimate for letter name/sound knowledge is significant. In contrast, shared environmental influences on the three factors are all substantial. Nonshared environmental influences (e^2) varied among the factors. That for phonological awareness is substantial, whereas e^2 estimates for letter name/sound knowledge and general prereading skills are marginal, although the latter is significant. Moreover, as shown in Table 7 (below the diagonal) and Figure 1, the phenotypic correlations among the three factors were due primarily to the substantial and significant correlations of the shared environmental influences on the factors, plus the significant correlation of the nonshared environmental influences on phonological awareness and general prereading skills. As also reported in Table 7 (above the diagonal), the bivariate shared environmentalities are substantially larger than the corresponding bivariate heritabilities.

As shown in Table 6, none of the genetic and shared environmental influences which were specific to each indicator were significant. In contrast, the nonshared environmental influences which were specific to each indicator were all significant, except for the single *kana* measure. Note that these nonshared influences include measurement error. This suggests that individual differences for each measurement are due primarily to the three

correlated factors (i.e., phonological awareness, general prereading skills, and letter name/sound knowledge).

4. Discussion

4.1. Relatively small genetic influences and substantial shared environmental influences in early *kana* reading

In this first Japanese twin study of early reading development, we investigated to what extent genetic and environmental influences contribute to individual differences in prereading skills and to concurrent correlations among them in Japanese-learning twins at 42 months of age. Our sample was unique in terms of the age of the twins and the language: The twins in the present study were younger than those in previous behavioral genetic studies on prereading skills, and they were learning a non-alphabetic language, Japanese.

Before we conducted behavioral genetic analyses, phenotypic factor analyses were performed. A three correlated-factors model, which included distinct factors of general prereading abilities, phonological awareness and letter name/sound knowledge, fit the data well. Of course, these results cannot be directly compared to previous studies because their measures and languages differed from ours. However, the factors for phonological awareness and letter name/sound knowledge distinct from general prereading skills were generally consistent with previous studies conducted with children who were learning English (e.g., Samuelsson et al., 2005; Lonigan, Burgess, & Anthony, 2000).

Results of our behavioral genetic analyses showed that genetic influences on the three factors of prereading skills (i.e., phonological awareness, general prereading skills, and letter name/sound knowledge) were relatively small, although the a^2 estimate of .14 for letter name/sound knowledge was significant. Moreover, there were no significant genetic correlations among them. Results of the multivariate behavioral genetic analysis also yielded no significant genetic or shared environmental influences which were specific to each prereading skill. Only nonshared environmental influences, which included measurement errors, were significant.

Our findings of the relatively small and non-significant genetic influences on prereading skills, except for letter name/sound knowledge, differ from most previous behavioral genetic studies, which have reported significant genetic influences on prereading skills and significant genetic correlations among them. However, our findings that there were modest genetic influences on letter name/sound knowledge and significant shared environmental influences on all of the factors for prereading skills are consistent with some of the findings from Samuelsson et al. (2005), Spinath, Price, Dale, & Plomin, 2004, and Hayiou-Thomas et al. (2006). For example, Samuelsson et al. (2005) reported the similar findings to ours in modest genetic influences and substantial shared environmental influences on some of their measures: they reported genetic influence on preschool print knowledge, a test which shares features with letter name/sound knowledge used here, of just .23, compared to .68 for shared environmental influence, and genetic influence on vocabulary of .32, compared to .60 for shared environmental influence (From Table 7 in Samuelsson et al. 2005).

Why are genetic influences on prereading skills in the present study relatively small? There are two possible ways to account for this discrepancy between the present study and previous studies. One possibility is the difference between the languages. Given the substantial differences between Japanese *kana* syllabary and alphabetic languages, including the fact that Japanese children learn *kana* quickly and most of them can read before formal education, most of the observed variance in *kana* may be due to differences in parental instruction which are shared by members of both MZ and DZ twin pairs. Results obtained in

the International Longitudinal Twin Study (Samuelsson et al., 2007; 2008) have also suggested that genetic influences on very early reading skills are relatively small when the environmental range is substantial.

A second possible explanation for the discrepancy between the present study and previous studies may be the difference in ages of the children. Children in the present study were younger than those included in most previous studies on prereading skills. Because genetic influences on various domains, including language, tend to become larger with age, whereas shared environmental influences tend to decrease (Hayiou-Thomas, Dale & Plomin, 2012; Plomin et al., 2008; Samuelsson et al., 2008), it is possible that our twins were in a developmental phase in which substantial genetic influences on prereading/reading skills have not been manifested. To test these possibilities, follow-up data at 60 months of age for the children in the present study are currently being collected.

4.2. Limitations

Our twin study has several limitations. First, the sample size is relatively small. Consequently, we could not examine whether the etiology of prereading skills differed by gender, which should be a topic for future studies. Second, the present study did not include direct measures of the environments. It has been previously reported that a wide range of measured environments, including home literacy environments, are associated with children's reading outcomes, independent of genetic effects (Petrill, et al., 2005; Hart, et al., 2009). Therefore, it would be interesting to examine how and to what extent the measured environments account for the individual differences in Japanese prereading skills in future studies. Finally, the findings of the present study should be interpreted with caution because the sample was based on twins, who generally lag slightly in language development (e.g., Rutter, Thorpe, Greenwood, Northstone, & Golding, 2003). Thus, it is possible that a mild delay in language development of twins may have reduced the sample means. However, the various language abilities did not differ appreciably by zygosity, and, if environmental hardships affect twins regardless of zygosity, the logic of the twin design is unaffected and the estimates of genetic and environmental influences should not be affected (Stromswold, 2001).

5. Conclusion

This first Japanese twin study of early literacy development tested young children who were learning a non-alphabetic language. Thus, it makes an important contribution to the early reading-development literature. Results suggest that individual differences in early kana reading development are due substantially to environmental influences that twin children share, such as family literacy processes, and less to genetic influences.

Acknowledgments

We are grateful to the Tokyo Twin Cohort Project, our testers, the twins and their parents. M. Koyama, S., Yamagata, K. Ozaki and K. Fukunaka provided us with valuable comments on tasks and analyses in this study. This study was funded by a grant from the Japan Science and Technology (JST), Research Institute of Science and Technology for Society (RISTEX), and the Grant-in-Aid for Scientific Research (KAKENHI). S. J. Wadsworth, R. K. Olson and J. C. DeFries were supported in part by the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) grant HD027802.

References

- Amano K. Formation of the act of analyzing phonemic structure of words and its relation to learning Japanese syllabic characters (kanamaji). *Japanese Journal of Educational Psychology*. 1970; 18:76–89. [in Japanese].

- Ando J, Nonaka K, Ozaki K, Sato N, Fujisawa KK, Suzuki K, et al. The Tokyo Twin Cohort Project; Overview and initial findings. *Twin Research and Human Genetics*. 2006; 9:817–826. [PubMed: 17254415]
- Bentler, PM. EQS 6 Structural Equations Program Manual. Encino, CA: Multivariate Software, Inc; 2006.
- Bishop DVM, Bishop SJ, Bright P, James C, Delaney T, Tallal P. Different origin of auditory and phonological processing problems in children with language impairment: Evidence from a twin study. *Journal of Speech Language and Hearing Research*. 1999; 42(1):155–168.
- Cho JR, McBride-Chang C, Park SG. Phonological awareness and morphological awareness: differential associations to regular and irregular word recognition in early Korean Hangul readers. *Reading and Writing*. 2008; 21(3):255–274.
- Chow BWY, Ho CSH, Wong SWL, Waye MMY, Bishop DVM. Genetic and Environmental Influences on Chinese Language and Reading Abilities. *Plos One*. 2011; 6(2)
- Dionne G, Dale PS, Boivin M, Plomin R. Genetic evidence for bidirectional effects of early lexical and grammatical development. *Child Development*. 2003; 74(2):394–412. [PubMed: 12705562]
- Frostig, M. Frostig Developmental Test of Visual Perception. Palo Alto, CA: Consulting Psychologists Press; 1961.
- Gathercole SE, Willis C, Baddeley AD, Emslie H. The Children's Test of Nonword Repetition: A test of phonological working memory. *Memory*. 1994; 2:103–127. [PubMed: 7584287]
- Griffin D, Gonzalez R. Correlational analysis of dyad-level data in the exchangeable case. *Psychological Bulletin*. 1995; 118:430–439.
- Hart SA, Petrill SA, DeThorne LS, Deater-Deckard K, Thompson LA, Schatschneider C, et al. 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*. 2009; 50(8):911–919. [PubMed: 19298476]
- Hayiou-Thomas ME, Kovas Y, Harlaar N, Plomin R, Bishop DVM, Dale PS. Common aetiology for diverse language skills in 41/2-year-old twins. *Journal of Child Language*. 2006; 33(2):339–368. [PubMed: 16826830]
- Hayiou-Thomas ME, Dale PS, Plomin R. The etiology of variation in language skills changes with development: a longitudinal twin study of language from 2 to 12 years. *Developmental Science*. 2012; 15:233–249. [PubMed: 22356179]
- Ho CSH, Bryant P. Phonological skills are important in learning to read Chinese. *Developmental Psychology*. 1997; 33(6):946–951. [PubMed: 9383617]
- Hohnen B, Stevenson J. The structure of genetic influences on general cognitive, language, phonological, and reading abilities. *Developmental Psychology*. 1999; 35(2):590–603. [PubMed: 10082029]
- Inagaki K, Hatano G, Otake T. The effect of kana literacy acquisition on the speech segmentation unit used by Japanese young children. *Journal of Experimental Child Psychology*. 2000; 75(1):70–91. [PubMed: 10660904]
- Kakihana S, Ando J, Koyama M, Iitaka S, Sugawara I. Cognitive factors relating to the development early literacy in the Kana syllabary. *Japanese Journal of Educational Psychology*. 2009; 57:295–308. [in Japanese].
- Kaufman, AS.; Kaufman, NL. Kaufmann Assessment Battery for Children Interpretive manual. Circle Pines, MN: American Guidance Service; 1983.
- Kim YS, Petscher Y. Relations of emergent literacy skill development with conventional literacy skill development in Korean. *Reading and Writing*. 2011; 24(6):635–656.
- Kirk, SA.; McCarthy, JJ.; Kirk, W. Illinois Test of Psycholinguistic Abilities. Champaign, IL: University of Illinois Press; 1968.
- Kovas Y, Hayiou-Thomas ME, Oliver B, Dale PS, Bishop DVM, Plomin R. Genetic influences in different aspects of language development: The etiology of language skills in 4.5-year-old twins. *Child Development*. 2005; 76(3):632–651. [PubMed: 15892783]
- Kubozono, H.; Honma, T. Syllable and mora [In Japanese]. Tokyo: Kenkyusha; 2002.

- Lonigan CJ, Burgess SR, Anthony JL. Development of emergent literacy and early reading skills in preschool children: evidence from a latent-variable longitudinal study. *Developmental Psychology*. 2000; 36:596–613. [PubMed: 10976600]
- National Language Research Institute. Reading and writing ability in pre-school children. Tokyo: Tokyo Shoseki; 1972. [In Japanese]
- Neale, MC.; Cardon, LR. Methodology for genetic studies of twins and families. Dordrecht, the Netherlands: Kluwer; 1992.
- Neale, MC.; Boker, SM.; Xie, G.; Maes, HH. Mx: Statistical modeling. 6. Richmond, VA: Virginia Commonwealth University, Department of Psychiatry; 2002.
- Olsen JA, Kenny DA. Structural equation modeling with interchangeable dyads. *Psychological Methods*. 2006; 11:127–141. [PubMed: 16784334]
- Ooki S, Yamada K, Asaka A, Hayakawa K. Zygosity Diagnosis of Twins by Questionnaire. *Acta Geneticae Medicae et Gemellologiae*. 1990; 39:109–115. [PubMed: 2392887]
- Otake T, Hatano G, Cutler A, Mehler J. Mora or syllable-speech segmentation in Japanese. *Journal of Memory and Language*. 1993; 32(2):258–278.
- Petrill SA, Deater-Deckard K, Schatschneider C, Davis C. Measured environmental influences on early reading: Evidence from an adoption study. *Scientific Studies of Reading*. 2005; 9(3):237–259.
- Petrill SA, Deater-Deckard K, Thompson LA, DeThorne LS, Schatschneider C. Genetic and environmental effects of serial naming and phonological awareness on early reading outcomes. *Journal of Educational Psychology*. 2006a; 98(1):112–121. [PubMed: 19444324]
- Petrill SA, Deater-Deckard K, Thompson LA, DeThorne LS, Schatschneider C. Reading skills in early readers: Genetic and shared environmental influences. *Journal of Learning Disabilities*. 2006b; 39(1):48–55. [PubMed: 16512082]
- Plomin, R.; DeFries, JC.; McClearn, GE.; McGuffin, ZP. Behavioral Genetics. 5. New York: Worth Publishers; 2008.
- Rutter M, Thorpe K, Greenwood R, Northstone K, Golding J. 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*. 2003; 43:326–341. [PubMed: 12635964]
- Samuelsson S, Byrne B, Olson RK, Hulslander J, Wadsworth S, Corley R, et al. Response to early literacy instruction in the United States, Australia, and Scandinavia: A behavioral-genetic analysis. *Learning and Individual Differences*. 2008; 18(3):289–295. [PubMed: 19122888]
- Samuelsson S, Byrne B, Quain P, Wadsworth S, Corley R, DeFries JC, et al. Environmental and genetic influences on prereading skills in Australia, Scandinavia, and the United States. *Journal of Educational Psychology*. 2005; 97(4):705–722.
- Samuelsson S, Olson R, Wadsworth S, Corley R, DeFries JC, Willcutt E, et al. Genetic and environmental influences on prereading skills and early reading and spelling development in the United States, Australia, and Scandinavia. *Reading and Writing*. 2007; 20(1–2):51–75.
- Scarborough, HS. Connecting early language and literacy to later reading (dis)abilities: Evidence, theory and practice. In: Neuman, SB.; Dickinson, DK., editors. *Handbook of early literacy research*. New York: Guilford Press; 2001. p. 97–110.
- Shimamura N, Mikami H. Acquisition of hiragana letters by pre-school children: In comparison with the 1967 investigation of the national language research institute. *Japanese Journal of Educational Psychology*. 1994; 42:70–76. [In Japanese].
- Siok WT, Fletcher P. The role of phonological awareness and visual-orthographic skills in Chinese reading acquisition. *Developmental Psychology*. 2001; 37:886–899. [PubMed: 11699761]
- Spinath FM, Price TS, Dale PS, Plomin R. The genetic and environmental origins of language disability and ability. *Child Development*. 2004; 75:445–454. [PubMed: 15056198]
- Stromswold K. The heritability of language: A review and metaanalysis of twin, adoption, and linkage studies. *Language*. 2001; 77:647–723.
- Swanson HL, Siegel L. Learning disabilities as a working memory deficit. *Issues in Education: Contribution of Educational Psychology*. 2001; 7:1–48.

- Tamaoka K, Makioka S. Frequency of occurrence for units of phonemes, morae, and syllables appearing in a lexical corpus of a Japanese newspaper. *Behavior Research Methods, Instruments, & Computers*. 2004; 36:531–547.
- Tong XL, McBride-Chang C, Wong AMY, Shu H, Reitsma P, Rispens J. Longitudinal predictors of very early Chinese literacy acquisition. *Journal of Research in Reading*. 2011; 34(3):315–332.
- Wagner RK, Torgesen JK, Rashotte CA, Hecht SA, Barker TA, Burgess SR, et al. 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*. 1997; 33:468–479. [PubMed: 9149925]

Figure 1.

Results of common pathway model with correlated factors. Path estimates shown in bold type and underlined did not include 0 in their 95% confidence intervals. Estimates for the genetic and environmental correlations among factors were omitted from the figure. However, the estimates of the correlations shown by bold curved arrows did not include 0 in their 95% confidence intervals, whereas the estimates of the correlations shown by thin curved arrows did. A: Genetic influence. C: Shared environmental influence. E: Nonshared environmental influence. a, c, e: genetic, shared, and nonshared environmental influences which are specific to each measure.

Table 1

Descriptive statistics for raw scores of each prereading skill.

	Mean	SD	Range of performance	
			min	max
Phonological awareness				
Ordinary CV words	1.14	1.13	0	3
Words including a long sound vowel	1.17	1.09	0	3
Words including a geminate consonant	0.69	0.86	0	3
Letter name/sound knowledge	2.92	4.61	0	15
Receptive vocabulary	18.56	5.10	0	34
Visual perceptual skill	2.81	1.57	0	8
Nonword repetition	17.74	9.94	0	39
Digit span	4.84	2.38	0	14

Table 2

Results of the exploratory factor analysis.

	Phonological awareness (PA)	General prereading skills (General)
Phonological awareness		
Words including a long sound vowel	.79	.00
Ordinary CV words	.68	.06
Words including a geminate consonant	.57	-.10
Letter name/sound knowledge	.32	.24
Digit span	-.05	.66
Nonword repetition	-.01	.58
Visual perceptual skill	.01	.39
Receptive vocabulary	.03	.34

Table 3

Phenotypic correlations among factors and prereading measures.

	1	2	3	4	5	6	7	8	9	10
1 PA	1									
2 General	.32	1								
3 LNK (Kana)	.36	.30	1							
4 PA: Normal	.82	.27	.36	1						
5 PA: Long vowel	.86	.30	.31	.55	1					
6 PA: Geminate	.68	.16	.17	.36	.42	1				
7 RV	.18	.50	.10 ⁺	.12 [*]	.17	.15	1			
8 VS	.17	.56	.18	.17	.16	.04 ^{ns}	.25	1		
9 NWR	.24	.76	.17	.23	.24	.11 [*]	.21	.21	1	
10 Digit	.24	.76	.29	.24	.23	.11 [*]	.20	.21	.39	1

Note. All correlations are significant at $p < .01$ unless otherwise indicated.

* : $p < .05$.

ns; *n.s.*

PA: composite score for phonological awareness, General: composite score for general prereading skill, LNK (Kana): letter name/sound knowledge, PA: Normal: phonological awareness of CV type words, PA: Long vowel: phonological awareness of words including a syllable with long vowel, PA: Geminate: phonological awareness of words including a geminate stop Q, RV: receptive vocabulary, VS: visual perception skill, NWR: nonword repetition, Digit: digit span.

Table 4

Intraclass and cross-trait/cross-twin correlations among factors and prereading measures.

	MZ	DZ	1	2	3	4	5	6	7	8	9	10
1 PA	.25**	.48**	---	.26**	.39**	.43**	.41**	.30**	.12	.20*	.14 ⁺	.30**
2 General	.50**	.49**	.22**	---	.28**	.25**	.25**	.11	.36**	.30**	.28**	.39**
3 LNK (Kana)	.91**	.84**	.23**	.31**	---	.42**	.32**	.18*	.05	.19*	.18*	.29**
4 PA: Normal	.27**	.46**	.27**	.18*	.20**	---	.33**	.24**	-.02	.23**	.16*	.29**
5 PA: Long vowel	.10	.40**	.18 ⁺	.22**	.18*	.21**	---	.23**	.14 ⁺	.16*	.17*	.27**
6 PA: Geminate	.13	.29**	.15 ⁺	.10	.16*	.13 ⁺	.11	---	.17*	.09	-.03	.14 ⁺
7 RV	.41**	.38**	.05	.24**	.11	.03	.04	.09	---	.14 ⁺	.26**	.15 ⁺
8 VS	.23*	.25*	.12*	.27**	.12	.12	.12	.01	.16*	---	.18*	.20*
9 NWR	.28**	.23*	.15*	.35**	.19*	.08	.18**	.09	.02	.15*	---	.24**
10 Digit	.47**	.33**	.18*	.43**	.32**	.22**	.16*	.04	.21**	.18*	.33**	---

Note..

** : $p < .01$.* : $p < .05$.+ $p < .10$.

PA: composite score for phonological awareness, General: composite score for general prereading skill, LNK (Kana): letter name/sound knowledge, PA: Normal: phonological awareness of CV type words, PA: Long vowel: phonological awareness of words including a syllable with long vowel, PA: Geminate: phonological awareness of words including a geminate stop Q, RV: receptive vocabulary, VS: visual perception skill, NWR: nonword repetition, Digit: digit span. Correlations for MZ twins are shown below the diagonal and those for DZ twins are shown above the diagonal.

Table 5

Variance component estimates in the univariate models.

Variable	a ²	c ²	e ²
PA	.00 (.00, .15)	.53 (.39, .63)	.47 (.37, .57)
General	.00 (.00, .25)	.66 (.43, .74)	.34 (.25, .43)
LNK (Kana)	.07 (.02, .14)	.88 (.82, .92)	.05 (.03, .07)
PA: Normal	.00 (.00, .15)	.53 (.38, .62)	.47 (.38, .57)
PA: Long vowel	.00 (.00, .13)	.39 (.24, .50)	.61 (.50, .74)
PA: Geminate	.00 (.00, .22)	.34 (.13, .46)	.66 (.54, .79)
RV	.07 (.00, .39)	.51 (.23, .65)	.41 (.32, .52)
VS	.00 (.00, .42)	.39 (.03, .49)	.61 (.48, .73)
NWR	.10 (.00, .54)	.34 (.00, .52)	.56 (.43, .71)
Digit	.19 (.00, .59)	.43 (.06, .65)	.38 (.28, .51)

Note. PA: composite score for phonological awareness, General: composite score for general prereading skill, LNK(Kana): letter name/sound knowledge, PA: Normal: phonological awareness of CV type words, PA: Long vowel: phonological awareness of words including a syllable with long vowel, PA: Geminate: phonological awareness of words including a geminate stop Q, RV: receptive vocabulary, VS: visual perception skill, NWR: nonword repetition, Digit: digit span. a², c², and e² are the estimates of proportion of variance due to genetic, shared and nonshared environmental influences, respectively. 95% confidence intervals are shown in parenthesis.

Table 6

Variance component estimates in the common pathway model with correlated factors.

Indicator	Factor					
	a^2_{total}	c^2_{total}	e^2_{total}	a^2_{specific}	c^2_{specific}	e^2_{specific}
PA: Normal	PA					
	.04 [.00, .23]	.32 [.14, .44]	.63 [.53, .75]	.07 [.00, .38]	.43 [.14, .61]	.49 [.35, .66]
	.04 [.00, .23]	.26 [.08, .38]	.69 [.59, .80]	.05 [.00, .48]	.76 [.37, .94]	.18 [.03, .35]
PA: Long vowel	General					
	.01 [.00, .22]	.21 [.05, .33]	.77 [.66, .89]	.14 [.02, .25]	.77 [.67, .94]	.08 [.00, .12]
PA: Geminate	LNK					
	.14 [.05, .25]	.77 [.67, .85]	.09 [.07, .12]			
Kana	Factor loading					
	.15 [.00, .46]	.30 [.06, .50]	.55 [.43, .69]	.00 [.00, .15]	.09 [.00, .18]	.37 [.27, .49]
RV	.05 [.00, .29]	.21 [.06, .36]	.74 [.59, .88]	.00 [.00, .19]	.00 [.00, .06]	.40 [.29, .51]
VS	.02 [.00, .24]	.31 [.13, .43]	.67 [.53, .79]	.00 [.00, .10]	.10 [.00, .20]	.64 [.52, .78]
NWR	.12 [.00, .33]	.34 [.16, .53]	.54 [.42, .70]	.00 [.00, .15]	.00 [.00, .15]	.00 [.00, .10]
Digit				.14 [.00, .45]	.18 [.00, .39]	.52 [.39, .67]
				.05 [.00, .28]	.08 [.00, .23]	.71 [.56, .85]
				.00 [.00, .18]	.03 [.00, .15]	.61 [.45, .74]
				.10 [.00, .25]	.00 [.00, .19]	.46 [.33, .62]

Note. 95% confidence intervals are shown in parenthesis. PA: phonological awareness factor, General: general prereading skill factor, LNK: letter name/sound knowledge factor, PA: Normal: phonological awareness of CV type words, PA: Long vowel: phonological awareness of words including a syllable with long vowel, PA: Geminate: phonological awareness of words including a geminate stop Q, RV: receptive vocabulary, VS: visual perception skill, NWR: nonword repetition, Digit: digit span. Note that letter name/sound knowledge is the single measure (shown as Kana). a^2 , c^2 , and e^2 are the estimates of proportion of variance due to genetic, shared and nonshared environmental influences, respectively. a^2_{total} , c^2_{total} , and e^2_{total} are the estimates of proportion of variance due to overall genetic and environmental influences to each indicator, respectively. a^2_{specific} , c^2_{specific} , and e^2_{specific} are the estimates of proportion of variance due to genetic and environmental influences which are specific to each indicator, respectively.

Table 7

Genetic and environmental correlations among factors (below the diagonal) and corresponding bivariate heritabilities and environmentalities (above the diagonal).

Genetic correlations			
	PA	General	LNK
PA	1.00	.02	.10
General	.37	1.00	-.09
LNK	1.00	-1.00	1.00
Shared environmental correlations			
	PA	General	LNK
PA	1.00	.40	.30
General	.69	1.00	.46
LNK	.52	.60	1.00
Nonshared environmental correlations			
	PA	General	LNK
PA	1.00	.13	.03
General	.42	1.00	.02
LNK	.16	.14	1.00

Note. PA: phonological awareness, General: general prereading skill, LNK: letter name/sound knowledge.