Early Child Language Mediates the Relation Between Home Environment and School Readiness

Nadine Forget-Dubois and Ginette Dionne
Université Laval

Jean-Pascal Lemelin
Université de Sherbrooke

Daniel Pérusse and Richard E. Tremblay
Université de Montréal

Home environment quality is a well-known predictor of school readiness (SR), although the underlying processes are little known. This study tested two hypotheses: (a) child language mediates the association between home characteristics (socioeconomic status and exposure to reading) and SR, and (b) genetic factors partly explain the association between language and SR. Data were collected between 6 and 63 months in a large sample of twins. Results showed that home characteristics had direct effects on SR and indirect effects through child language. No genetic correlation was found between language and SR. These results suggest that home characteristics affect SR in part through their effect on early language skills, and show that this process is mainly environmental rather than genetic in nature.

School readiness (SR) is a multidimensional construct that includes behavioral, emotional, cognitive, and knowledge components that make the child “ready to learn” at school entry (Blair, 2002; Chew, 1981). As SR assessments are quite heterogeneous in content, their predictive validity regarding future school achievement tends to vary greatly (for a review, see LaParo & Pianta, 2000). Some have argued that fluid cognitive skills, such as executive functions and memory, are better predictors of future school achievement (Blair, 2006). However, more crystallized preacademic knowledge components typically assessed in SR batteries, such as number, letter, and color knowledge, have been shown to predict early school achievement over and above general cognitive ability (Forget-Dubois et al., 2007; Hess, Holloway, Dickson, & Price, 1984; Lemelin et al., 2007). Children who lack the underlying basic knowledge of the early curriculum may experience difficulties in keeping up. Moreover, preacademic knowledge likely reflects both the underlying fluid cognitive skills needed for early learning, as well as the exposure to basic knowledge (Blair, 2006). Therefore, focusing on preacademic knowledge is a valid way of assessing SR.

It is generally assumed that SR can be traced back to prior influences in the home environment (Hess et al., 1984; Melhuish et al., 2008; NICHD Early Child Care Research Network, 2000). Indeed, many characteristics of the home environment, including attachment security and continuing sensitive care, verbal stimulation, access to educational material in the home, and specific parental practices such as reading with the child, have been linked to SR (Belsky & Fearon, 2002; Bradley & Caldwell, 1984; Britto, Brooks-Gunn, & Griffin, 2006; Connell & Prinz, 2002; Mcloyd, 1998; Reese, Cox, Harte, & McAnally, 2003; Stipek & Ryan, 1997). In addition, the gains in SR by children of head start were shown to be greater when parents were more involved in the program (Parker, Boak, Griffin, Ripple, & Peay, 2009).
1999; Ramey, 1999). Therefore, the association between characteristics of the home environment and SR is empirically documented.

In contrast, relatively few studies have tested specific hypotheses regarding the processes leading to the observed association (NICHD Early Child Care Research Network, 2003). The process model that has been tested most often is a mediation model (Foster, Lambert, Abbott-Shim, McCarty, & Franze, 2005; NICHD Early Child Care Research Network, 2003). Mediation models can be thought of as directional models specifying direct and indirect effects of predictors on an outcome (Baron & Kenny, 1986; Cole & Maxwell, 2003). Regarding the effects of the home environment on SR, two types of mediation models have been considered: A first type of model posits that a distal measure of home environment influences the outcome through an ecologically proximal characteristic of the home environment. The underlying principle of this model is that the environment has specific rather than global influences over developmental outcome (Hoff, 2003). For example, socioeconomic status (SES), a distal characteristic of the environment, was shown to have an indirect effect on SR through parental practices regarding home learning, a proximal measure of the home environment quality (Foster et al., 2005). A second type of mediation model suggests that the environment could affect the child outcome of interest through its effect on another child outcome; for example, the influence of home environment quality on SR was shown to be mediated by sustained attention, a child characteristic (NICHD Early Child Care Research Network, 2003). These two types of processes do not have to be mutually exclusive; both allow the testing of explanatory models involving global effects of distal factors and specific effects of proximal factors, and the relation between these two types of effects.

The first goal of this study was to test whether the effect of home environment quality on SR can be explained in part by its influence on early language. We considered SES as an indicator of the general home environment quality and exposure to reading as a specific feature of the family environment that is likely to promote SR both directly and indirectly through its influence on early language.

### Language as a Mediator Between Home Environment Influence and SR

Before testing a mediation model, there must be evidence that the putative predictor, outcome, and mediator variables are associated (Baron & Kenny, 1986). Indeed, previous studies offer solid empirical reasons to hypothesize that children’s language skills may play a mediating role in the association between the quality of the environment and SR.

First, many features of the home environment are significantly predictive of later language skills (Hart & Risley, 1992; Hoff & Tian, 2005; Peterson, Jesso, & McCabe, 1999; Snow, Tabors, & Dickinson, 2001; Tabors, Roach, & Snow, 2001), especially of skills associated with written language (Bus, van IJzendoorn, & Pellegrini, 1995). Early vocabulary has been associated with SES and the quality of maternal speech; in fact, the effect of SES on early vocabulary has been shown to be entirely mediated by maternal speech in a sample of middle and high SES children, a result consistent with the principle of a specific (rather than global) effect of the family environment on child language (Hoff, 2003). The quality of the home literacy environment was associated with vocabulary in first grade (Van Steensel, 2006). Sensitive parenting, maternal responsiveness, and the feedback children receive in their interactions with adults have also been shown to predict early language skills (Hirsh-Pasek & Burchinal, 2006; Tamis-LeMonda & Bornstein, 2002; Tomasello, 1992; Tomasello & Farrar, 1986). Finally, regular reading in the first 3 years of life in low-income families was shown to predict later vocabulary and general cognitive skills (Raikes et al., 2006).

Second, the association between language skills and SR is also well documented. In children from low SES background, higher language competence was correlated with SR ratings (Fiorentino & Howe, 2004). Narrative abilities were related to children’s functioning in the classroom environment (Peterson, 1994). Similarly, children’s mean length of utterances at age three was correlated with emerging literacy in kindergarten (Tabors et al., 2001). Language skills at school entry were also found to be a protective factor moderating the association between social risk and school achievement (Burchinal, Roberts, Zeisel, Hennon, & Hooper, 2006). However, the relation between language and SR remains to be clarified because language has often been considered not only as a predictor of school achievement (see, e.g., Agostin & Bain, 1997; Snow et al., 2001), but also as a SR measure in itself (Dunn & Dunn, 1981; Majsterek & Ellenwood, 1995; Mantzicopoulos, 1999; Williams, Voelker, & Ricciardi, 1995). Given the documented sequence of predictive associations between home environment quality, early language skills, and early school achievement, there is ground
to hypothesize that early language skills mediate the association between home environment and SR.

Correlation Between Environmental and Genetic Factors

Even though previous studies have established the importance of home environment quality for the development of SR, the possibility that genetic factors could play a role in this process needs to be considered for two main reasons.

First, parents create the family environment partly on the basis of their own genetic characteristics. Thus, the family environment is probably associated with parental genes and, by way of consequence, to their children’s genes. Given the likelihood of such a genotype–environment correlation \( r_{ge} \) (Plomin, DeFries, McClearn, & McGuffin, 2001), any model proposing a mediation of distal environmental characteristics by a proximal environmental characteristic (Foster et al., 2005) could in fact reflect an association between two environmental characteristics via the parental genotype. This type of \( r_{ge} \) is rarely taken into account because measures of the environment are often not genetically informative (Turkheimer, D’Onofrio, Maes, & Eaves, 2005). It is to be noted that in the context of genetic studies, any source of variance other than genes is considered as “the environment,” so the definition of environment is much less specific than in studies based on models of environmental influences.

Second, the association between two child characteristics may stem in part from a shared genetic basis, as most human traits and behaviors are partly heritable (Turkheimer, 2000). This is especially true of cognitive development where many cognitive skills are genetically correlated (see Plomin & Spinath, 2004). The importance of these genetic correlations led Plomin and Kovas (2005) to conclude that the same “generalist genes” influence many aspects of cognition. Therefore, a mediation model positing child characteristics as mediator and outcome, like the model tested by NICHD Early Child Care Research Network (2003), could partly reflect the effect of generalist genes.

Unfortunately, there is no simple solution for including measures of the environment in genetically informative studies. The conclusions that can be drawn remain limited by the difficulty to control for \( r_{ge} \) (Purcell & Koenen, 2005; Turkheimer et al., 2005), and the present study is no exception. However, it is possible to address the problem of potential genetic confounds in the association between two child outcomes in this study by testing for a genetic correlation between the mediator and the outcome of our model, language and SR. Their association could reflect a common genetic basis rather than result from the indirect influence of home environment quality and, at the very least, this should be considered in the interpretation of the mediation model.

There is indeed evidence that many language, learning, and general cognitive abilities share a common genetic basis, and that these may be more important than environmental factors in explaining the association and stability of these abilities (Kovas, Haworth, Dale, & Plomin, 2007; Plomin & Kovas, 2005). Yet, although heritability estimates of fluid skills (e.g., IQ) range between 40% and 60%, varying with age (McGue, Bouchard, Iacono, & Lykken, 1993; Plomin et al., 2001) and show little influence of the environment shared by children of the same family, a different picture emerges when SR is considered. Genetic studies of specific aspects of SR show only modest heritability, ranging from 30% to 45% (Lemelin et al., 2007; Oliver, Dale, & Plomin, 2005); a greater proportion of their variance seems environmentally driven, more specifically through factors shared by children in a family. Similarly, genetic studies of early language have shown that the environment shared by children of a family is crucial to vocabulary development and, to a lesser extent, to grammar skills, although early language skills are also moderately heritable (between 25% and 50%; Dale, Dionne, Eley, & Plomin, 2000; Dionne, Dale, Boivin, & Plomin, 2003; Dionne, Tremblay, Boivin, Laplante, & Pérusse, 2003; Hohnen & Stevenson, 1999). To sum up, environmental processes are important factors in the etiology of both SR and early language skills, but both are also heritable; whether their association is partly accounted for by common genetic factors, as the generalist genes model would suggest, remains to be tested.

Objectives

Although child development research has moved away from the Nature versus Nurture dichotomy (Collins, Maccoby, Steinberg, Hetherington, & Bornstein, 2000; Maccoby, 2000), in practice, studies of genetic factors and studies of environmental processes generally have different goals and are conducted independently of each other. The main goal of this study was to test the hypothesis that the contribution of home environment quality (SES and exposure to reading) to SR is partly mediated by the child’s language skills. Specifically, we
hypothesized that a home environment offering a variety of stimulating experiences and learning opportunities during infancy would contribute to SR, assessed just before school entry, partly through its effect on early child language. We also tested the hypothesis that the influence of SES, a global, distal measure of the family environment, is mediated by a more proximal measure, exposure to reading. As language may share a common genetic basis with SR, we also tested the hypothesis that language skills and SR could be associated through shared genes. This genetically informed analysis provided the basis for the interpretation of the mediation model: If language and SR were associated because they share a genetic basis, then the process through which language mediates the relation between home characteristics and SR could not be seen as strictly environmental. However, the absence of a genetic correlation linking early language and later SR would be consistent with a genuine environmental process, or at least with the absence of a direct genetic influence on the mediation process.

Method

Participants

The Quebec Newborn Twin Study (QNTS) is based on a representative sample of twins born between April 1995 and December 1998 in the Greater Montreal Area, Canada. Names, addresses, and phone numbers of all mothers of newborn twins were collected from the birth records of the Québec Bureau of Statistics. The participating families have been assessed yearly, starting when the children were 5 months old (corrected for gestational age: they would have been 5 months old if gestational duration had been 40 weeks, which is rarely the case with twins. The uncorrected mean age was 6.28 months ($SD = .82$). The participating families were predominantly of European descent (84%), whereas the remainder were of African (3%), Asian (2%), and other (2%) descent. The remaining 9% did not provide information on ethnicity. The maternal language of the parents was French for 71.9% of mothers and 72.3% of fathers, English for 9.9% of mothers and 8.8% of fathers, and another language for 19.2% of mothers, and 18.9% of fathers. The average household income was around CAN $54,000, which is slightly above average for similar households in this geographic area at that time. Regarding education, 14% of the mothers did not complete high school, 28% were high school graduates, 48% had some college education or graduated from college (9% did not provide any information regarding education). For fathers, 11% did not complete high school, 34% were high school graduates, and 39% had some college education or graduated from college (information was unavailable for 16% of fathers). The analyses presented here are based on data gathered when the twins were 6, 19, 32, and 63 months. At the first assessment, 662 twin pairs were assessed. At the 63-month assessment, 446 twin pairs were still enrolled in the study. Full data were available for 693 individual children in the present study (see below for a detailed description of attrition and treatment of missing data).

Predictors of SR

The predictors of SR were SES and a measure of exposure to reading material. In an ecological perspective, SES represented a distal measure of home environment quality, and exposure to reading a proximal measure of stimulation in the home setting. SES was computed from data on parental education and household income obtained during a home interview with the mothers when the twins were 6 and 19 months. Education and income were aggregated into 5-point scales (from 0 to 4): from high school not completed to university diploma obtained for the education scale, and from income less than $20,000 to over $80,000 for the income scale. We averaged 6- and 19-month family income and paternal and maternal highest education level to obtain a measure of family SES ($\alpha = .80$, $M = 2.18$, $SE = .04$, $SD = 1.07$).

The measure of exposure to reading material was also based on data reported by the mother during the home interviews. The items were initially developed for Canada’s National Longitudinal Study of Children and Youth and adapted for the QNTS. We retained three items rated on a 8-point Likert scale regarding exposure to reading at 19 months: The mothers reported on how often they looked at books with the children, how often they read to their children, and how often the children looked at books by themselves at the time of the interview. We computed a mean score from these three items ($\alpha = .87$) to reflect exposure to reading. This score was reflected and log-transformed to reduce negative skewness ($M = .66$, $SE = .01$, $SD = .32$). The Exposure to Reading score was available for 553 pairs of twins. The mothers did not report on reading for each child individually, so exposure to reading is the same for both
twins of a pair. The measure did not include the mother’s own literacy practices.

Child Language

Child language was measured using the expressive vocabulary short form checklist of the MacArthur–Bates Communicative Development Inventory (MCDI; Fenson et al., 1994) when the twins were 19 and 32 months. The parents were asked to indicate from a list of root words (77 words at 19 months and 100 words at 32 months) those they had heard each of their twins say. To reduce contamination of a twin’s assessment on the cotwin, the parents completed the assessment for one twin during the home visit and were asked to complete and send in the assessment for the cotwin 2 weeks later. Adapted and translated versions of the MCDI were used for French-speaking families as described in Dionne, Tremblay, et al. (2003). At the time of the measurement, no French norms existed for the MCDI. For this reason, raw scores were retained, regressed for age (taking into account gestational duration) and standardized. The correlation between the 19- and 32-month scores is \( r = .44, p < .001 \).

At 19 months, 83.1% of children were assessed in French and 16.9% in English. There was no difference between the means and variances of the French and English subsamples, \( t(511) = 1.11, p = .27 \). The 32-month figures are very similar. A mean score was computed from the 19- and 32-month scores allowing for one missing value to maximize sample size. This measure of 19- and 32-month child expressive vocabulary was used in the analysis as an index of early child language skills (\( M = .01, SE = .04, SD = .91 \)).

SR and General Cognitive Ability

SR was assessed during a laboratory visit when the twins were 63 months old on average, in the summer prior to kindergarten entry. We used the Lollipop test, a validated instrument assessing knowledge of colors and shapes, spatial recognition, numbers, and letters (Chew, 1981; Chew & Morris, 1984; Lemelin et al., 2007). A total score (with a maximum of 69) was calculated from the sum of the items (\( \alpha = .89, M = 42.16, SE = .59, SD = 12.89 \)).

General cognitive ability was assessed during the same visit using the Block Design subtest of the Wechsler Preschool and Primary Scale of Intelligence–Revised (WPPSI-R; Wechsler, 1989). The scores were adjusted for age as instructed in the test manual (\( M = 9.95, SE = .13, SD = 2.85 \)). We deemed it necessary to control for general cognitive ability to capture the crystallized aspects of SR independently of more fluid cognitive skills.

Analyses

The analytic strategy combined a classic path model and a genetically informative model. The path model required independence of measures, but the genetically informative model required twin pairs, who cannot be considered independent. Thus, we analyzed two data sets: Data Set 1 comprising one twin per pair selected randomly, and Data Set 2 including all available twins of all pairs.

We first tested the hypothesis regarding the prediction of SR using a path analysis model. Two indirect paths were examined alongside the direct paths: the mediation of contribution of SES and exposure to reading to SR through child expressive vocabulary and the mediation of SES by exposure to reading. Full data were available for 351 children in Data Set 1; the actual number of participants for each variable varied from \( n = 450 \) (general cognitive ability) to \( n = 658 \) (sex). Missing data were taken into account using the full information maximum likelihood (FIML) estimation procedure available in the statistical package Amos 5 (Arbuckle, 2003).

We then examined the common genetic and environmental etiology of language and SR using classic twin modeling. This design is based on the comparison of the covariance between monozygotic (MZ) and dizygotic (DZ) twins for a given measure, knowing that the genetic correlation is 1 between MZ twins and .5 on average between DZ twins. Formal analysis of twin data was done using the Mx package (Neale, Boker, Xie, & Maes, 1999) to partition the variance of measures between genetic and environmental sources (Plomin et al., 2001). The environmental variance was further decomposed to estimate the proportion of variance attributable to the common environment, which augments the similarity between twins regardless of zygosity, and to the unique environment, which makes twins of a pair more different (Turkheimer & Waldron, 2000). Moreover, twin studies allow estimating the magnitude of genetic and environmental correlations between traits or behaviors, indicating to what extent they share the same genetic and environmental sources of variance. Specifically, the genetic correlation refers to the correlation between the genetic components of variance associated with each measured behavior. Full data for expressive vocabulary, SR, sex, age, general
cognitive ability, and zygosity were available for 344 (148 MZ and 198 DZ) complete pairs in Data Set 2. Missing data in covariables (definition variables) in the statistical package Mx results in the loss of the whole case (Mx Script Library, n.d.), so we conducted the analyses on the twins for which sex, age, and general cognitive ability were available and allowed for missing data on SR and expressive vocabulary using the FIML estimation procedure available in Mx (Neale et al., 1999). The model included a total of 428 twin pairs with at least one twin assessed on at least one of the two outcomes.

The correlation between language and other cognitive abilities could be explained by general intelligence $g$ and its underlying genetic factors (Plomin & Kovas, 2005; Plomin & Spinath, 2004). Accordingly, we first conducted the genetic analysis using the Block Design scores, a putative proxy of general intelligence, as a control variable, to assess the crystallized aspects of SR rather than the fluid skills. However, putative “generalist genes” underlying the variance of $g$ may also influence other aspects of language and learning (Plomin & Kovas, 2005), and partialing out the variance associated with the general cognitive ability score (i.e. the Block Design score) may underestimate the genetic correlation between early language and SR. Therefore, the genetic analyses were performed twice, with and without controlling for the contribution of general cognitive ability to SR.

**Attrition and Missing Data Treatment**

A total of 989 families were asked to join the QNTS; of these, 662 were assessed during the first wave of data collection. The average attrition rate from 6 to 63 months was 9.3% per year. The remaining 441 families assessed at 63 months differed from those who left the study on socioeconomic variables: At enrollment, those who stayed until 63 months were more educated and had a larger family income; they were also more likely to have an intact biological family and to have French or English as their first language, and the mothers were also slightly older (see Lemelin et al., 2007, for details). We further analyzed the pattern of missing data for the variables included in the study with the MVA module in SPSS. The data were not missing completely at random (MCAR) according to Little’s MCAR test ($\chi^2 = 154.48$, $df = 73$, $p < .000$). Separate variance $t$ tests obtained with the MVA module showed that children who were evaluated at 63 months for SR had higher mean 6/19 months SES, $t(727.6) = 4.1$, $p < .001$, and were more exposed to reading at 19 months, $t(464.7) = 2.4$, $p = .02$). Only 1% of SR assessments were rejected as invalid so most of the missing assessments resulted from participants lost to attrition.

These analyses revealed a pattern commonly found in longitudinal studies, that attrition did not happen entirely at random; low SES families were more likely to dropout of the study before completing the 63 months assessment. As missing status in participants was associated with measured sociodemographic characteristics included in the model tested, we deemed it reasonable to consider them missing at random (McKnight, McKnight, Sidani, & Figueredo, 2007, p. 53) and treat them accordingly. The FIML estimation procedure, contrary to multiple imputation, is implemented directly in the process of fitting a model; it treats missing data by fitting the model to all nonmissing data for each observation (McCartney, Burchinal, & Bub, 2006). We chose FIML over multiple imputation because there is evidence that FIML has more power unless a great number of imputations are done (Graham, Olchowski, & Gilreath, 2007). The downside of the FIML procedure is that a single $N$ is not representative of the study sample and thus cannot be reported. All statistics reported in this article, including the means and variances reported earlier, were estimated in models fitted using FIML. It is to be noted that the parameters estimated with the FIML procedure did not differ from the parameters estimated with a listwise deletion of missing cases in any way likely to change the general conclusions.

**Results**

**Correlations**

Table 1 shows the correlations between SR, SES, exposure to reading, and expressive vocabulary as well as with the covariables. The correlation model was fitted in Amos using Data Set 1 (one twin selected per pair). Correlations were modest but significant between the main variables, and the values of $r$ were in the same range as those reported in other studies that assessed the strength of the relation between home environment and SR (e.g., NICHD Early Child Care Research Network, 2003).

**Does Expressive Vocabulary Mediate the Relation Between Home Environment and SR?**

The path model including all the hypothesized direct and indirect effects on SR was tested using Amos 5. Covariables (sex, age when SR was measured, and general cognitive ability) were also...
The overall model explained 33% of the variance in SR; unsurprisingly, the most important predictor was general cognitive ability, measured simultaneously with SR and entered as a covariable. The hypothesized direct and indirect effects were significant but small as indicated by the small values of the standardized path coefficients (which can be interpreted like the Bs and bs in regression models, as an indication of how the outcome varies as a function of the predictors). Overall, the model was consistent with the double mediation hypothesis: Exposure to reading, a proximal measure of environment, mediated the effect of SES, the more distal measure, on SR but also on expressive vocabulary; expressive vocabulary partially mediated the effect of both SES and exposure to reading on SR. The significance of the simple partial mediation can be further investigated with the Sobel test (Preacher & Leonardelli, 2001; Sobel, 1982). The partial mediation by expressive vocabulary of the effect of SES on SR was near significance (z = 1.92, p = .05). The direct path from SES to expressive vocabulary was itself barely significant (p = .04), suggesting that the most important indirect effect of SES was through the more complex path of its effect on exposure to reading and expressive vocabulary. The significance of this indirect effect through two mediators cannot be demonstrated with the Sobel test; however, it is accepted that the effect is significant if all paths are significant for a given α level (Kline, 2005, p. 162). All the paths involved in this indirect effect are significant at the p < .001 level. The simple mediation of the effect of SES on expressive vocabulary by exposure to reading was significant (z = 3.82, p < .001), as was the mediation of the effect of exposure to reading on SR by expressive vocabulary (z = 3.42, p < .001).

The total indirect effects of SES and exposure to reading on SR were small (as reported in the lower part of Table 2), but were considered significant because all the path coefficients representing the indirect effects were significant. The total effects include the sum of direct and indirect effects of SES and exposure to reading on SR. SES was the environmental predictor that explained the biggest proportion of variance in SR.

Is There a Genetic Correlation Between Expressive Vocabulary and SR?

This analysis was based on Data Set 2 (all twins). As indicated in Table 1, the phenotypic correlation between early expressive vocabulary and 63-month SR was r = .34, p < .001. The correlation in Data Set 2 was very similar (r = .33, p < .001). Descriptive statistics and intraclass correlations for MZ and DZ twins are reported in Table 3. Means and variances of MZ and DZ twins are expected to be equal, which was the case with these data. The correlation differences for MZ and DZ twin pairs suggested a modest heritability and a substantial effect of common environment for both measures.

A first bivariate correlated factors model (Loehlin, 1996) was fitted to the data to estimate formally the genetic and environmental components of the variance for SR and expressive vocabulary, as well as the genetic and environmental correlations included. The full model with standardized estimates of path coefficients is illustrated in Figure 1 and the unstandardized path coefficients with their associated p values are reported in Table 2.
across measures. In this first model, we controlled for the effect of sex on expressive vocabulary and SR, and the effect of general cognitive ability and age at testing on SR. This model showed a good fit to the data compared with a phenotypic saturated model ($-2LL = 7786.68, df = 1569, p = .66, Akaike's information criterion (AIC) = 46487.68$).

The best-fitting model is depicted in Figure 2 (the covariables are not represented for clarity purposes). Both child measures showed a small but significant heritability, but the genetic correlation between the two variables could be dropped from the model without significantly affecting the fit, $\chi^2$ difference = 2.10(1), $p = .15$. Common environment was the most important source of variance, and the common environmental correlation between vari-
ables was .55 (95% confidence interval (CI): .38–.76). The unique environment factors, which include measurement error, were also correlated at $r = .21$ (95% CI: .08–.34). Overall, the environmental correlations suggest that expressive vocabulary and SR were correlated because they shared the same underlying environmental influences, not because they shared the same underlying genetic influences.

We tested a second model without controlling for the effect of general cognitive ability on SR. Removing this covariable resulted in a significant decrease of fit, $-2LL = 7908.59$, $\chi^2$ difference = 119.81(1 df), $p < .000$. However, the estimates of A, C, E, and the correlations between these factors were very similar compared with the first model.

Table 2
Unstandardized Path Models and Effect Sizes

<table>
<thead>
<tr>
<th>Path</th>
<th>Estimate</th>
<th>SE</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exposure to reading ← SES</td>
<td>0.10</td>
<td>0.01</td>
<td>.000</td>
</tr>
<tr>
<td>Exposure to reading ← sex</td>
<td>0.05</td>
<td>0.03</td>
<td>.06</td>
</tr>
<tr>
<td>18/30-month expressive vocabulary ← exposure to reading</td>
<td>0.59</td>
<td>0.14</td>
<td>.000</td>
</tr>
<tr>
<td>18/30-month expressive vocabulary ← SES</td>
<td>0.08</td>
<td>0.04</td>
<td>.04</td>
</tr>
<tr>
<td>18/30-month expressive vocabulary ← sex</td>
<td>0.10</td>
<td>0.08</td>
<td>.21</td>
</tr>
<tr>
<td>School readiness ← 18/30-month expressive vocabulary</td>
<td>3.37</td>
<td>0.61</td>
<td>.000</td>
</tr>
<tr>
<td>School readiness ← exposure to reading</td>
<td>4.10</td>
<td>1.78</td>
<td>.02</td>
</tr>
<tr>
<td>School readiness ← SES</td>
<td>2.09</td>
<td>0.53</td>
<td>.000</td>
</tr>
<tr>
<td>School readiness ← general intelligence</td>
<td>1.58</td>
<td>0.19</td>
<td>.000</td>
</tr>
<tr>
<td>School readiness ← sex</td>
<td>1.235</td>
<td>1.01</td>
<td>.22</td>
</tr>
<tr>
<td>School readiness ← age</td>
<td>10.03</td>
<td>1.97</td>
<td>.000</td>
</tr>
</tbody>
</table>

Effects on school readiness

<table>
<thead>
<tr>
<th>Path</th>
<th>Estimate</th>
<th>SE</th>
<th>$p$ value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SES</td>
<td>0.86</td>
<td>0.07</td>
<td>2.95</td>
</tr>
<tr>
<td>Exposure to reading</td>
<td>2.00</td>
<td>0.05</td>
<td>6.10</td>
</tr>
</tbody>
</table>

Table 3
Comparison of MZ and DZ Twins Means, Variances and Intra-Class Correlations (ICC) for Expressive Vocabulary and School Readiness

<table>
<thead>
<tr>
<th></th>
<th>School readiness</th>
<th>Expressive vocabulary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MZ</td>
<td>DZ</td>
</tr>
<tr>
<td>Mean</td>
<td>42.66</td>
<td>43.51</td>
</tr>
<tr>
<td>Variance</td>
<td>152.56</td>
<td>175.02</td>
</tr>
<tr>
<td>ICC</td>
<td>.79</td>
<td>.65</td>
</tr>
</tbody>
</table>

Note. The differences in means and variances between MZ and DZ twins for school readiness and expressive vocabulary are not significant. MZ = monozygotic; DZ = dizygotic; ICC = intraclass correlation coefficient.
and the genetic correlation was still nonsignificant. These results suggest that the covariance of expressive vocabulary and SR reflected shared common and unique environmental influences but did not reflect shared genetic influences; therefore, the contribution of early child language to SR is not likely to reflect an underlying genetic liability or ability.

**Discussion**

The main objective of this article was to demonstrate that the predictive association between characteristics of the home environment and SR is partly mediated by early child language skills. In addition, the genetic–environmental etiology of the association between early language and SR was investigated to assess the extent to which this association stems from shared genetic and environmental factors. The results indicated that early child language skills partly mediated the contribution of SES and exposure to reading on SR. SES also showed an indirect contribution to child expressive vocabulary through its contribution to exposure to reading. Second, although child expressive vocabulary and SR were both moderately heritable, their association was entirely explained by shared environmental factors. The results of the mediation model and the absence of a genetic correlation between child language and SR are consistent with the view that characteristics of the home environment, such as exposure to reading and SES, contribute to early language skills and that these, in turn, contribute to SR. The specific conclusions of the mediation model and genetic analysis are discussed in detail in the following section.

**Mediation Analysis**

The pattern of indirect contributions of SES to SR was more complex than initially expected: The indirect contribution of SES to SR was accounted for by mediation through expressive vocabulary and also by an indirect effect via both exposure to reading and child expressive vocabulary. This result suggests two processes by which SES affects early language skills: a direct effect, and because higher SES is predictive of greater exposure to reading, which in turn has a positive effect on SR, an indirect effect. Exposure to reading made a direct contribution to SR and an indirect contribution through child expressive vocabulary.

These results are globally consistent with our proposed model of the processes leading to the association between home environment and SR: children who benefit from presumably more stimulating home environments, as indicated by SES and exposure to reading, acquire better language skills, and in turn acquire more of the competences and knowledge associated with SR. The mediation of the effect of SES (a distal measure of the quality of the environment) by exposure to reading (a proximal measure) is consistent with the results reported by Foster et al. (2005). These authors found a similar mediation process regarding SES, proximal measures of home environment, and emerging literacy, which showed how differences in SES may become associated with aspects of SR. Our results add another component to that explanatory model, one that takes into account early child language skills. In other words, high SES parents tend to read more to their toddlers which in turn contributes to their children having more developed language skills. In turn, better language skills help toddlers to profit from the stimulating home environment, which leads to better SR skills. These indirect effects are added to the direct effects of SES and exposure to reading on SR.

Prediction of the quality of early language was also central to our model, because language had to be predicted by SES and exposure to reading to mediate the effect of these two environmental characteristics on SR. Our results are consistent with previous findings showing that reading has a positive impact on children’s vocabulary (Raikes et al., 2006; Sénéchal, LeFevre, Hudson, & Lawson, 1996; Van Steensel, 2006). Our results are also consistent with Hoff’s studies of the association between SES and vocabulary (Hoff, 2003; Hoff & Tian, 2005). Hoff (2003) found a complete mediation of the effect of SES by maternal speech abilities in a middle- to high-SES sample that was interpreted as evidence for specific rather than global effects of the home environment on child language. We found a partial mediation of the effect of SES by exposure to reading, our proximal measure of home environment quality and thus found evidence for specific effects, but could not rule out a global effect of SES on language development. As Hoff noted, it is not known whether the environmental specificity principle applies across the whole range of SES. Moreover, the remaining direct effect of SES in our study could be mediated by unmeasured variables. Whether early language skills could act as a protective factor in preacademic and academic development, as Burchnal et al. (2006) showed, also remains to be tested.

Although significant, the direct and indirect effects of the environmental variables on SR were
small. This is not surprising in the context of a longitudinal study with predictive measures taken more than 30 months prior to the outcome measure. Yet, the initial correlations between the variables were similar in size to the correlations reported in NICHD’s mediation model of the influences on SR (NICHD Early Child Care Research Network, 2003), in which correlations between predictor, mediator, and outcome variables ranged from nonsignificant to .34, with predictor and outcome measures assessed between 6 and 54 months of age. These results may reflect in part the difficulty encountered in obtaining reliable measures of the relevant aspects of home environment associated with preschool knowledge acquisition. It may also indicate that the strongest environmental influences on SR are found later than 32 months.

Genetically Informative Analysis

Contrary to expectations derived from the theory of “generalist genes,” the association between child expressive vocabulary and SR was not explained by a common genetic basis; it follows that the mediation linking children’s language skills and preschool knowledge cannot be attributed to a common genetic basis either. Previous genetically informative studies (reviewed by Kovas et al., 2007; Plomin & Kovas, 2005) point to a common genetic basis for various cognitive abilities, language, and learning processes. Considering that our main goal was to test a model of environmental influences, the concurrent test for a possible genetic mediation is a significant contribution of this study.

The absence of common genetic substrate between language and SR is surprising. Plomin and Kovas (2005) demonstrated convincingly that finding a common genetic basis for different cognitive abilities and disabilities is the norm rather than the exception. They further propose that genetic factors are responsible for the association between language, general cognitive abilities, and learning abilities, whereas environmental influences are responsible for change in the same abilities over the span of childhood (Kovas et al., 2007). On the contrary, our findings showed that environment rather than genes is responsible for the predictive association between early language skills and SR. Furthermore, we found that common environment is the most important source of stability between the two child outcomes. We propose three possible explanations for this discrepancy. First, the children in this study were younger than those in the studies reviewed by Kovas et al. It has been suggested that common environment is a more important source of variance early in life (Plomin & Petrill, 1997). Second, the child outcomes considered in the present study required an important input from the environment as children, regardless of their potential cognitive abilities, must be taught preacademic notions. Children learn vocabulary and basic knowledge from competent adults in the context of specific cultural groups; although all normally developing children learn to talk, the specific familial, cultural, economical, etc. context in which they grow up induces variance in language skills (Hoff, 2006). Third, it is possible that the association between family environment, child language, and SR is influenced by a genotype–environment correlation that could not be detected in our model. For example, children who inherit genetic risks associated with poor SR skills could also be more likely to receive less stimulation, if poorer environment is correlated with the genes that they received from their parents. Obviously, these explanations are not mutually exclusive.

Implications of the Findings

Given that SR is one of the best predictors of school success and school completion (Battin-Pearson et al., 2000; LaParo & Pianta, 2000), having empirical evidence that the early family environment matters gives even more support to family-oriented prevention programs aimed at early development. Indeed, the child language skills assessed in this study were measured prior to 30 months. This indicates that environmental processes contributing to the acquisition of basic knowledge are not limited to the late preschool years, but begin much earlier in infancy. This is consistent with results from studies of early language acquisition showing that the verbal content of infant–mother interactions predicts the rate of language development as early as age 2 (Bornstein, Tamis-LeMonda, & Haynes, 1999). Learning language may be one of the first manifestations of the transactional processes involved between children and their environment in any form of learning. It follows that interventions aimed at raising the quality of child language at a very early age may have enduring results well into the late preschool years and beyond.

Successful intervention programs with a focus on the early years and improved child language before school entry have shown positive contributions to early school success (Landry, Swank,
Smith, Assel, & Gunnewig, 2006; Wasik, Bond, & Hindman, 2006). These studies have shown that the language development and literacy skills of children living in poverty can be substantially increased by interventions targeting early language skills. Successful programs have involved parental training (Sénéchal & LeFevre, 2002), showing that parents can be taught how to promote better learning. Such interventions would probably benefit more children living in impoverished environments, thus reducing the gap between low-SES and middle-class children at school entry.

Exposure of children to reading and reading material is one early parental behavior that has repeatedly been shown to contribute positively not only to early reading skills (e.g., Sénéchal, 2006; Sénéchal & LeFevre, 2002; Van Steensel, 2006) but also to early child language (e.g., DeBaryshe, 1993; Foy & Mann, 2003; Sénéchal et al., 1996) and SR (Britto et al., 2006; Reese et al., 2003). The child-directed joint-attention episodes that joint reading promotes have been repeatedly shown to provide the best environment for early learning (Tomasello & Farrrar, 1986). Finally, joint reading provides age-appropriate material in an attractive format and gives the child an introduction to the importance of words in our societies as vehicles of ideas and feelings. For all of these reasons, early exposure to reading may be one of the more potent learning experiences of early childhood.

**Limitations**

This study, as are most genetically informative studies, is limited by the difficulty to control for the effects of genotype–environment correlations. Turkheimer et al. (2005) observed that the promises to assess environmental influences while controlling genetic influences never really materialized. In this study, we chose to use a genetic analysis to provide a context for the interpretation of a nongenetic model. Even though our results strongly favor an environmental process linking child language and SR, we cannot claim that our analysis dismissed all genetic influences on the mediation process; measures of home environment and child outcomes may still be associated because of undetected genotype–environment correlation. The conclusion of the genetic analysis is limited to showing no evidence of a direct genetic effect at work in the tested mediation process and that, at the very least, environmental influences on this process far outweigh any genetic influences. If a genotype–environment correlation is present, children who inherited genes associated with better SR may have been more likely to inherit favorable environments and vice versa. This would not invalidate the idea that home environment has an important effect on SR; it would rather imply that children who are likely to inherit genetic risk factors are also likely to experience environmental risk, while children likely to inherit protective genetic factors would also be likely to grow up in a protective environment. In such a case, children at risk would be likely to benefit more from intervention than children not at risk.

Another limitation concerns the variables used in the longitudinal model. Although we chose variables measured prior to 30 months, we cannot assume that their effect is limited to that period. In fact, it is very likely that family SES assessed when the children were 5 and 18 months was strongly correlated with SES at 60 months. Thus, the model presented in this article represents a plausible explanation of the relation between various environmental influences and child outcomes, but does not demonstrate causal relations. A third limitation is the use of a single informant, the mother, for the measure of expressive vocabulary in both twins. Although there was a 2-week delay between the assessments, they cannot be considered independent. Furthermore, the mother was also the only informant for the measures of SES and exposure to reading, which cannot be considered independent either. However, the questions from which these measures were derived left little room for interpretation so it is unlikely that the results would have been very different had our sources for the environmental measures been independent. Finally, we measured the occurrence and frequency of exposure to reading, but the quality of interaction between parent and child during book reading may be important to consider (Bus, Belsky, van Ijzendoorn, & Crnic, 1997). All these limitations underline the difficulty of assessing the quality of the home setting and the near impossibility to assess characteristics of the home setting that may differ for each twin. This study needs to be replicated with better, independent measures of language and home environment.

**Conclusion**

The aim of this study was to assess the role of early language skills in the process through which the quality of home environment and academic SR become associated. Concurrently, we assessed the genetic and environmental etiology of the association between early language skills and SR.
We found that early language skills partly but significantly mediated the influence of SES and exposure to reading on SR. Moreover, the genetically informative analysis suggested that this process could represent a genuine influence of the home environment on SR as there was no evidence of a genetic correlation between early language skills and SR. If a genetic effect was involved in the process, it was necessarily indirect through undetected $r_{ge}$. We conclude that a stimulating family environment has a positive effect on SR both directly and through its effect on language development. These results support early intervention in the family environment, or through surrogate parenting, aiming at improving child language and SR through augmented exposure to reading.

References


