

Why do children read more? The influence of reading ability on voluntary reading practices

Elsje van Bergen,^{1,2}  Margaret J. Snowling,^{2,3} Eveline L. de Zeeuw,¹ Catharina E.M. van Beijsterveldt,¹ Conor V. Dolan,¹ and Dorret I. Boomsma¹

¹Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands; ²Experimental Psychology, University of Oxford, Oxford; ³St. John's College, Oxford, UK

Background: This study investigates the causal relationships between reading and print exposure and investigates whether the amount children read outside school determines how well they read, or vice versa. Previous findings from behavioural studies suggest that reading predicts print exposure. Here, we use twin-data and apply the behaviour-genetic approach of direction of causality modelling, suggested by Heath et al. (1993), to investigate the causal relationships between these two traits. **Method:** Partial data were available for a large sample of twin children ($N = 11,559$) and 262 siblings, all enrolled in the Netherlands Twin Register. Children were assessed around 7.5 years of age. Mothers completed questionnaires reporting children's time spent on reading activities and reading ability. Additional information on reading ability was available through teacher ratings and performance on national reading tests. For siblings reading test, results were available. **Results:** The reading ability of the twins was comparable to that of the siblings and national norms, showing that twin findings can be generalized to the population. A measurement model was specified with two latent variables, Reading Ability and Print Exposure, which correlated .41. Heritability analyses showed that Reading Ability was highly heritable, while genetic and environmental influences were equally important for Print Exposure. We exploited the fact that the two constructs differ in genetic architecture and fitted direction of causality models. The results supported a causal relationship running from Reading Ability to Print Exposure. **Conclusions:** How much and how well children read are moderately correlated. Individual differences in print exposure are less heritable than individual differences in reading ability. Importantly, the present results suggest that it is the children's reading ability that determines how much they choose to read, rather than vice versa. **Keywords:** Direction of causality models; reading skills; print exposure; twin studies; causal modelling; behaviour-genetics.

Introduction

Learning to read builds on language skills, it requires instruction and it also requires practice. Cunningham and Stanovich (1997) were the first to formally propose that practice, or 'print exposure' is a vital ingredient in the development of fluent reading. However, there are vast individual differences in children's reading habits. It has been estimated that, whereas avid readers read as many as 1.8 million words per year, reluctant readers read only about 8,000 words for their own enjoyment (Anderson, Wilson, & Fielding, 1988; table 3). Measured longitudinally, the link between how much and how well a child reads holds over a 10-year time period (Cunningham & Stanovich, 1997). Measured concurrently, the link is consistently present from preschool, when the frequency of shared-reading correlates with emergent literacy skills, to university, when the amount of independent reading correlates with word-level reading skills, reading comprehension and vocabulary size (Mol & Bus, 2011). For decoding or word-level reading, the focus of the current paper, the concurrent correlation during the school years is estimated at .38 (Mol & Bus, 2011). The amount of time children read out-of-school

hours has been variously termed reading amount, reading frequency, reading for pleasure, independent reading and print exposure. Measured here is the quantity of reading that parents state their children do of their own volition and not as prescribed by school. We use the term 'print exposure' here but, by adopting this term, we do not assume that a child's print exposure is the outcome of a passive process. Indeed, a key issue is the causal direction of the link (or links) between reading skill and print exposure: do children who read more become better readers (print exposure \rightarrow reading), do poorer readers avoid reading (reading \rightarrow print exposure) or is there a reciprocal relationship between reading and print exposure?

To date, three studies have used a longitudinal design to investigate the relationships between reading and print exposure. Aarnoutse and van Leeuwe (1998) tracked the development of children's print exposure and reading comprehension from the second to sixth grades. Print exposure developed largely independent of reading comprehension, with marginal influences of reading comprehension on print exposure. Over a shorter time scale, Leppänen, Aunola, and Nurmi (2005) assessed print exposure and reading ability (accuracy, fluency and comprehension) between Grades 1 and 2 in a cross-lagged design. Causal relationships were primarily from

Conflict of interest statement: No conflicts declared.

reading skills to print exposure, though there was a small effect of print exposure on reading accuracy. Finally, Harlaar, Deater-Deckard, Thompson, DeThorne, and Petrill (2011) measured print exposure and reading skill (i.e. a composite of accuracy and comprehension) between the ages of 10 and 11 years. The effect again ran from reading skill to print exposure. However, interpretation of the findings from these studies is hampered by the strong stability of reading ability over time, for example, Harlaar et al. (2011) report an autoregressive effect of .90. Given the large autoregressive effect of earlier reading on later reading ability, there is little variance remaining for any other variable to explain. It follows that understanding any putative impact of print exposure on reading (or a reciprocal relationship) calls for an alternative design.

An important hypothesis regarding the relationship between reading ability and print exposure is that it reflects shared genetic influences. Further, on the grounds of temporal precedence, it might be assumed that reading mediates genetic influences on print exposure. While it is well established that differences among children in reading skills are largely due to genetic factors (Olson, Keenan, Byrne, & Samuelsson, 2014), with heritability across studies reported to be .73 (de Zeeuw, de Geus, & Boomsma, 2015), few studies have investigated the aetiology of individual differences in print exposure. Those that have report heritability estimates ranging from 0.10 (Harlaar, Dale, & Plomin, 2007) through 0.39 (Harlaar, Trzaskowski, Dale, & Plomin, 2014) to 0.65–0.67 (Harlaar et al., 2011; Martin et al., 2009), all with large E-components, suggestive of substantial measurement error.

Arguably, the typically lower heritability estimates for print exposure than reading are to be expected given that print exposure depends on the presence of printed material in the child's environment. Turning to estimates of the genetic correlation between reading and print exposure; these are about .60 (Harlaar et al., 2011, 2014; Martin et al., 2009). However, since a genetic correlation may be due to genetic effects on both traits (pleiotropy), such estimates cannot speak to the causal relationship between two traits (here reading and print exposure). Alternatively, in a causal model, genetic effects on the causal phenotype are transmitted through a direct relationship with the outcome phenotype, hence giving rise to a genetic correlation between the phenotypes (e.g. genes → reading → print exposure).

A design that can speak to the direction of causality is the behaviour-genetic 'direction of causality' model (Duffy & Martin, 1994; Heath et al., 1993). This design requires cross-sectional data on family members, such as twin pairs, and, therefore, does not depend on longitudinal data so that the stability of traits is not an issue. Denoting the phenotypes X and Y, competing models are tested which explain the X-Y correlation as resulting from either (a) a common genetic factor, (b) a

common underlying environmental cause, (c) both 1 and 2, (d) X influencing Y at the phenotypic (i.e. behavioural) level, (e) Y influencing X, or (f) both 4 and 5 (i.e. reciprocal phenotypic influences). Twin data allow the investigation of causality because the different models give rise to different expectations for the cross-trait cross-twin correlation (i.e. the correlation between X in twins with Y in the co-twin). Competing models can be distinguished best if the correlation between traits is reasonably large (i.e. >.25, Duffy & Martin, 1994), the traits differ in the relative importance of genetic and environmental influences, and measurement error is accounted for by using multiple indicators of the phenotypes (Heath et al., 1993). This model has been applied successfully to address causality (e.g. Ebejer et al., 2010; Gillespie, Zhu, Neale, Heath, & Martin, 2003; Thomsen et al., 2009; Touloupoulou et al., 2015).

In the current study, we apply direction of causality models to infer the causal relations (if any) between reading ability and print exposure in a large sample ($N > 11,000$) of 7-year-olds. Given the robust association between our traits of interest, the larger impact of environmental differences on print exposure than on reading ability, and the use of latent variables to account for measurement error, direction of causality modelling should work well.

Methods

Participants

The present project was approved by the medical ethical committee of the Vrije Universiteit Amsterdam (NTR/25-05-2007). Analyses were based on 11,559 twins, born between 1994 and 2004, and 262 siblings. The sample was obtained from the Netherlands Twin Register, a nationwide database of multiple births and their family members, including data from birth onwards (van Beijsterveldt et al., 2013). For present purposes, we employed data from school achievement records and from questionnaires, which mothers and teachers completed when the children were 7½ years old. Having obtained parental consent, we approached teachers to complete the questionnaire and provide test scores from the achievement records (national pupil monitoring system (Cito, 2014)).

In total, 368 (3.1%) twin children were excluded, and 11,559 were retained for analyses (Appendix S1). Each of the 11,559 children provided at least one data point. These 11,559 twins (of whom 5,723 were boys) came from 6,072 twin pairs: 2,175 MZ twin pairs (1,034 male; 1,141 female) and 3,897 DZ twin pairs (1,021 male; 939 female; 1,937 opposite sex). Zygosity of same sex twin pairs was determined using DNA polymorphisms (in 10.0%) or using a parent-report zygosity questionnaire comprising 10 items on twin similarity, with an accuracy of 93% (Rietveld et al., 2000). Mothers and teachers completed the questionnaire including items on reading when the children were on average 7.50 (SD 0.33) and 7.44 (SD 0.36) years old, respectively. For neither informant were the item scores systematically related to age ($-.05 < r_s < .05$). Teachers completed the questionnaires at the end of the school year: most children attended Grade 1 (69%) or 2 (26%). The proportion of twins who were in the same classroom (and rated by the same teacher) and the proportion of twins who were in different classrooms (rated by different teachers) did not differ significantly by zygosity (see Appendix S1).

Teachers also provided test scores for reading ability of 262 siblings of the twins from achievement records. These additional data enabled us to examine how representative the twin sample was of nontwin children. When there were multiple siblings, we selected the data of the oldest sibling ($n = 262$; 127 boys).

Measures

The current study employed school achievement records and data from questionnaires that were mailed or offered online to mothers and teachers. Print exposure was based on maternal ratings and reading ability was based on maternal and teacher ratings and achievement records.

Reading ability. We used five indicators of reading ability: one item from the mother questionnaire, two from the teacher questionnaire and two test scores. Mothers were asked to report the current school grade for Dutch language [scale points: 1 (*fail*), 2 (*poor*), 3 (*satisfactory*), 4 (*good*) and 5 (*excellent*)]. Teachers were asked the child's usual grades for reading [scale points: 1 (*fail*), 2 (*poor*), 3 (*satisfactory*), 4 (*above average*) and 5 (*good/excellent*)] and from the Connors' Teacher Rating Scales (Connors, Sitarenios, Parker, & Epstein, 1998) whether the child lagged behind in reading [scale points: 1 (*not true at all*), 2 (*just a little true*), 3 (*pretty much true*), to 4 (*very much true*)].

Reading ability was tested at school with a word-reading fluency list (Verhoeven, 1995; List 3) given by the teacher to children individually. The list, which is part of the Dutch pupil monitoring system, comprises 120 polysyllabic words varying in orthographic complexity (Verhoeven & van Leeuwe, 2009). Children were asked to read aloud as many words as possible within one minute in mid-Grade 2 and mid-Grade 3. These test scores correlated .86 (Table 1).

Print exposure. The two indicators of print exposure came from the mothers' questionnaire. Mothers were asked 'How many books (no comics) does the child read per week?' [scale points: 1 (*none*), 2 (*one or two*), 3 (*three or four*) and 4 (*more than four*)], and 'How much time does the twin spend on the following activities?' for which one of the listed activities was 'reading books' [scale points: 1 (*every day*), 2 (*almost every day*), 3 (*a couple of days per week*), 4 (*once a week*), 5 (*less than once a week*), 6 (*so far once*) and 7 (*never*)]. As an index of reliability we also calculated how much time children spend on reading books compared with the other activities that were listed (e.g. gaming, watching TV, playing with friends). This relative measure correlated .92 with the raw item.

Results

Sample representativeness

We assessed two indicators of how representative our sample is based on how well the twins in the

sample read. First, we compared the reading test scores of the twins to those of their siblings. As shown online in Table S1, the means did not differ significantly (Cohen's d s -0.02). Second, the test manual reports the following mean scores (i.e. cut-offs between below and above average, or C and B levels) for second to fifth grades: 39.5, 59.5, 71.5 and 81.5. Our twin sample means are ~ 0.18 *SD* above these population means. So, good reading families are somewhat overrepresented, but within families, twins are representative of singletons. Further analyses are based on the twin data.

Descriptive statistics

Some items were reverse scored so higher scores reflect better and more reading. Data from some questionnaire items relating to reading ability were negatively skewed (Table 2), that is, less sensitive in discriminating among good readers. Nevertheless, the correlations between the questionnaire items and the test scores were between .54 and .74 (Table 1). The distributions of the two test scores were approximately normal. Intercorrelations among reading ability indicators were between .54 and .86. The two print exposure indicators correlated .51. Correlations among ability and exposure items were low (.15–.31). The association between reading ability and print exposure is visualized in Figure 1.

Missing data. Out of seven possible data points, children had on average 3.33 data points ($SD = 1.43$). See Appendix S1 for details. We undertook missing value analyses to examine whether missingness of data was related to mother's educational level or children's reading ability (Tables S2 and S3). We did not run these analyses for print exposure, because we only had data from one informant. These analyses found missingness to be unrelated to the dependent variable reading ability; coupled with the fact that the data set is very large, the missing data do not, therefore, pose major issues. The structural equation modelling described below was performed in Mplus (Muthén & Muthén, 1998–2015). Mplus handles missing data by fitting the model using robust raw-data maximum-likelihood

Table 1 Correlations among items

Variable	Source	1.	2.	3.	4.	5.	6.
Reading ability							
1. School grade for language	Mother report	–					
2. Reading difficulties	Teacher report	.60	–				
3. School grade for reading	Teacher report	.68	.79	–			
4. Reading test Grade 2 ^a	School records	.62	.57	.70	–		
5. Reading test Grade 3 ^a	School records	.56	.54	.66	.86	–	
Print exposure							
6. Number of books per week	Mother report	.28	.21	.26	.31	.31	–
7. Time spent reading	Mother report	.21	.15	.20	.21	.20	.51

^aThe test measured word-reading fluency. All p 's < .001.

Table 2 Descriptive twin data statistics

Variable	Source	Range ^a	Twins		
			<i>N</i>	<i>M</i>	<i>SD</i>
Reading ability					
1. School grade for language	Mother report	1–5	7,496	3.69	0.93
2. Reading difficulties	Teacher report	1–4	6,587	3.47	0.93
3. School grade for reading	Teacher report	1–5	5,826	3.60	1.24
4. Reading test Grade 2 ^b	School records	0–120	1,702	42.25	18.78
5. Reading test Grade 3 ^b	School records	0–120	1,774	63.43	18.24
Print exposure					
6. Number of books per week	Mother report	1–4	7,594	2.09	0.71
7. Time spent reading	Mother report	1–7	7,564	5.65	1.39

Items were mirrored when necessary so that higher scores reflect better and more reading.

^aRange of possible values.

^bThe test measured word-reading fluency.

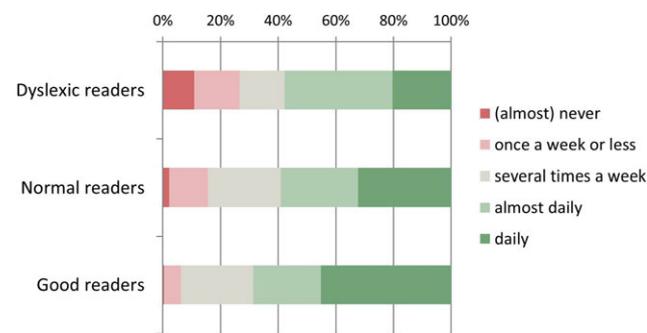


Figure 1 ‘How much time does the child spend reading books?’ per reading group. Only children with data on the word-reading fluency test were included in this figure. If both Grades 2 and 3 scores were available, the first was taken. Some answer categories were collapsed to simplify the figure. Dyslexic is defined as scoring <10th percentile, normal as 10th–75th percentile, and good as >75th, according to national norms

estimation on the assumption that data are missing at random.

Analytical approach. Three sets of models were fitted, each set of models building on the previous set. The first set is based on a two-factor phenotypic measurement model (Figure 2). The model formed the basis for the behavioural-genetic model (Figures 3 and 4), which in turn formed the starting point for the causal models (Figure 5).

First, we fitted phenotypic two-factor (measurement) models to the data set. To account for dependency among the observations (twins clustered in pairs), we corrected standard errors and model fit statistics as proposed by Rebollo, de Moor, Dolan, and Boomsma (2006). The final measurement model formed the basis of subsequent behavioural-genetic modelling.

After establishing the measurement model, we investigated the relative contribution of genes and environment on the phenotypic variances and covariance of reading ability and print exposure. The twin method exploits data on MZ and DZ twin pairs who are raised together. MZ twins are

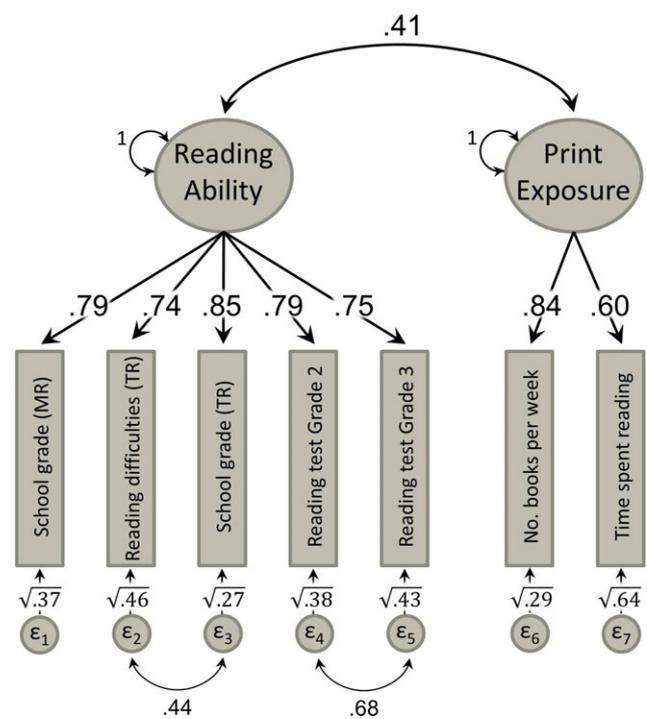


Figure 2 Measurement model. Note. MR = mother report; TR = teacher report. Model fit: $\chi^2(11, N = 11,559) = 38.83$, $p < .001$; RMSEA = 0.015, 90% CI (0.010–0.020); CFI = 0.997

genetically identical, DZ twin share on average 50% of their segregating genes. We decomposed the total trait variance into an additive-genetic variance component (A), a common environmental variance component, reflecting the effect of environmental influences that the twins share (shared or common environment: C), and an unshared-environmental variance component, which reflects unshared environmental influences (nonshared environment: E). The phenotypic twin correlations are diagnostic of the underlying model. Genetic influences are implicated if the MZ twin correlation is larger than the DZ correlation.

We specified for Twin 1 and Twin 2 members the previously established two-factor (measurement)

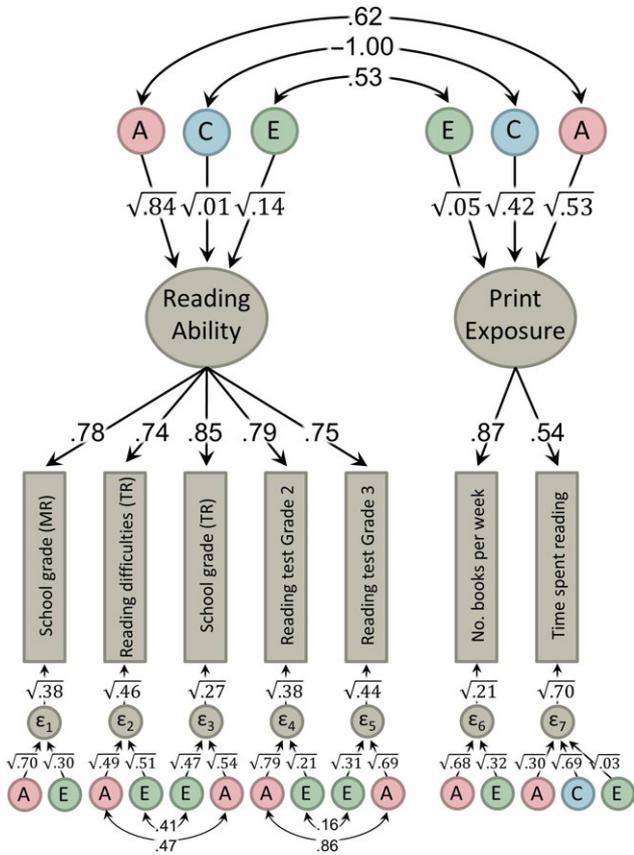


Figure 3 Full correlational model. The contribution of C to the variance of Reading Ability was estimated at only 1% and not significant ($p = .120$)

model. Next, we tested the basic twin-model assumption of equal means and variances over twin members and over zygosity. This more restricted model, labelled the phenotypic twin model, formed the basis of the ACE models, in which the variances of the latent factors were decomposed into A, C and E.

Modelling results

Measurement model. A model with two common factors representing Reading Ability and Print Exposure fitted the data reasonably well: $\chi^2(df = 13, N = 11,559) = 743.91, p < .001$; RMSEA = 0.070 (0.066–0.074); CFI = 0.931. Based on the inspection of modification indices (which indicate possible sources of misspecification), we allowed the residuals of the two teacher items to covary, and the residuals of the two reading tests to covary. This revised model fitted significantly better than the previous one: $\Delta\chi^2(df = 2) = 705.08, p < .001$. It showed excellent overall fit: $\chi^2(df = 11, N = 11,559) = 38.83, p < .001$; RMSEA = 0.015, (0.010–0.020); CFI = 0.997. We used this revised model as our measurement model (Figure 2).

Behavioural-genetic models. Mean differences in item scores between boys and girls were accounted

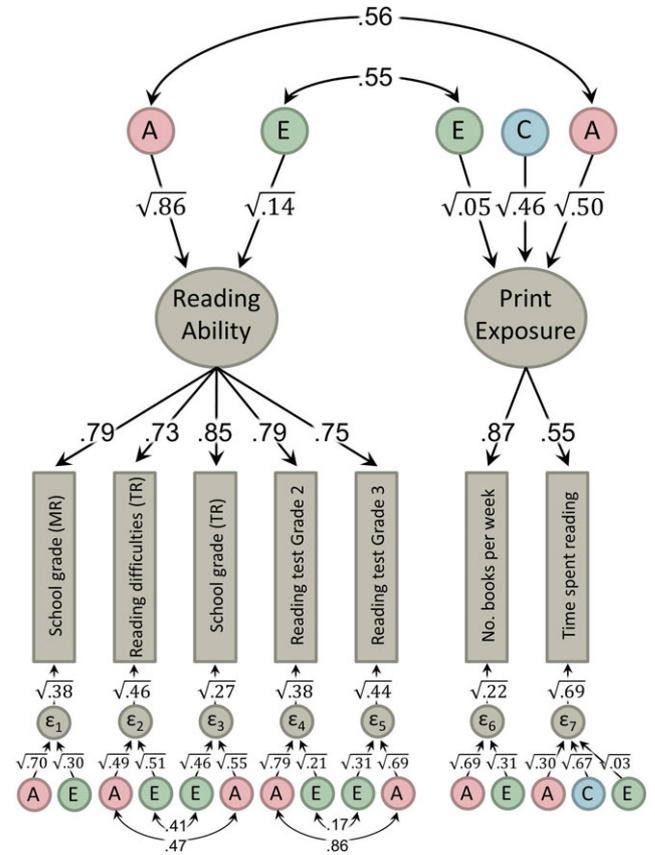


Figure 4 Final correlational model

for by regressing all items on sex, coded 0 (boys) or 1 (girls). On average, girls were rated as reading more ($\beta_s .12-.15, ps < .001$) and slightly better than boys ($\beta_s .07-.09, ps < .001$). However, on average, boys and girls scored similarly on the reading tests ($\beta_s .02, ps > .33$). We subsequently applied an omnibus test of basic assumptions of the twin model concerning equal means and variances. This yielded $\Delta\chi^2$ statistic of 171.6 ($df = 46; p < .01$) and an increase in AIC of 79.64. As inspection of the results revealed no evident misfit, we attribute the significance and the increase in AIC to the power afforded by the large sample size (i.e. the number of twin pairs is 6,072). We note that the BIC, which compared to AIC favours more parsimonious models, decreased by 229.08. Twin correlations are reported in Table 3.

Subsequently, we carried out the decomposition of the phenotypic variances into the A, C and E variance components. We did not consider sex differences in heritability. Previous work on reading ability (de Zeeuw, van Beijsterveldt, Glasner, de Geus, & Boomsma, 2016) and the current twin correlations suggest that the A, C and E variance components do not differ between boys and girls. Residual variances of the items were also decomposed. As, the C components of the residual variances (except the one from the last indicator) were all close to zero, these were set to zero. To reflect the correlation between the residuals of the two teacher

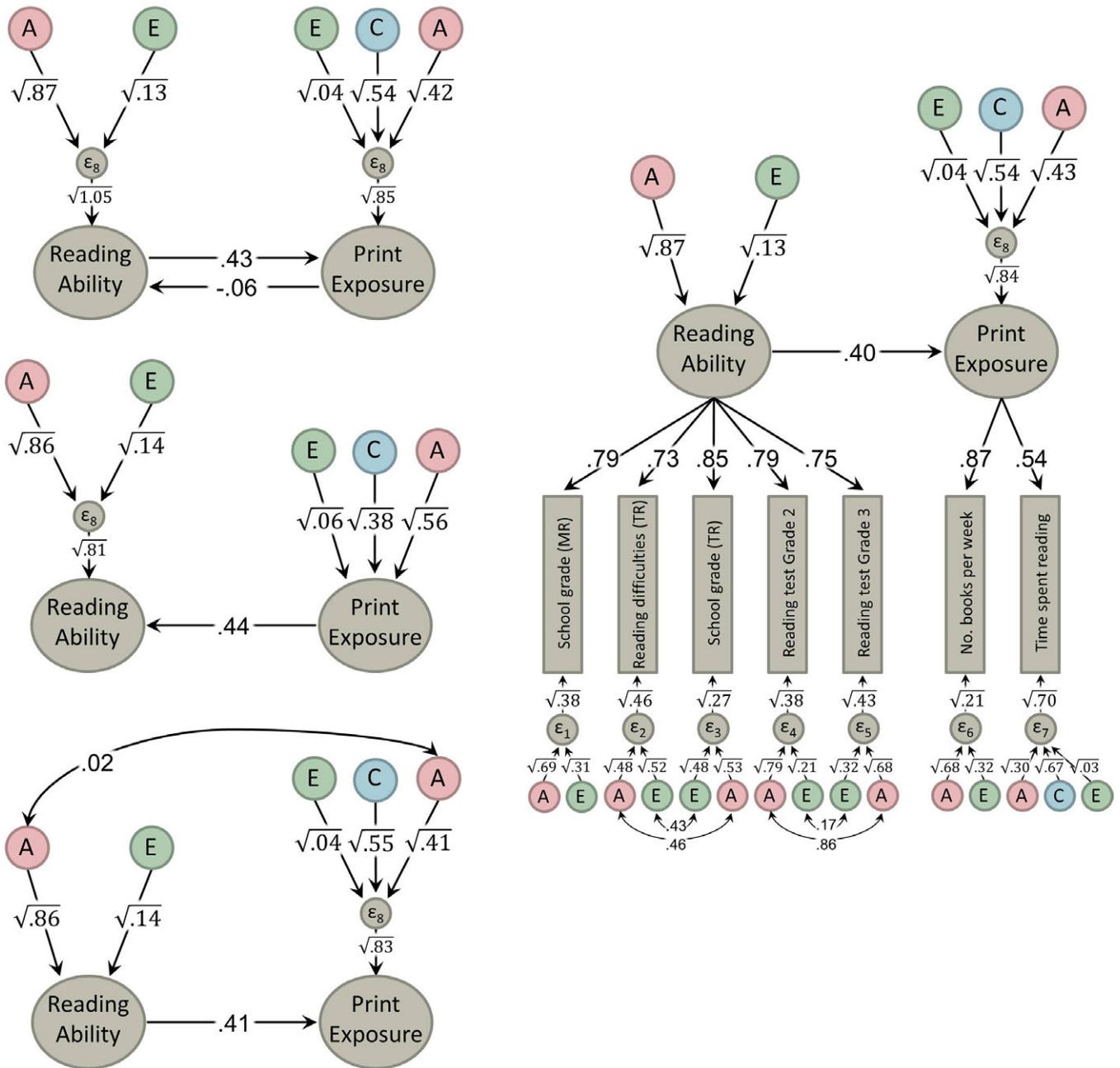


Figure 5 Direction of causality models. In the interest of space, only the top part of the three models on the left-hand side is shown. In the bidirectional model (top left), the $-.06$ path is not significant ($p = .254$). The middle left one has a very poor fit. In the bottom left one, the rA-path of $.02$ has a p -value of $.031$. The one on the right-hand side is the one that is parsimonious and supported by the data

Table 3 Within and cross-trait twin correlations

	MZ	DZ
Univariate		
Reading ability	.86	.43
Print exposure	.95	.70
Multivariate		
Reading ability – Print exposure	.33	.12

items and the correlation between the residuals of the two test items, correlation between their respective As and Es were added. This bivariate ACE model

is depicted in Figure 3. The contribution of C to the variance in Reading Ability was estimated at only 1% and was not significant ($p = .120$). As a consequence, the correlation between the C factors of reading ability and print exposure could not be estimated reliably. Therefore, we dropped both paths ($\Delta\chi^2$ -test: $\Delta\chi^2(df = 2, N = 6,072 \text{ pairs}) = 9.37, p = .001$), resulting in the model depicted in Figure 4.

We then tested the four causal models depicted in Figure 5. To get from the correlational model in Figure 4 to the bidirectional model in Figure 5 (top left), two correlations were dropped and two causal paths were added, resulting in a model with the

same number of degrees of freedom. The χ^2 , AIC and BIC went up by 3.88, indicating only slightly poorer fit than the correlational model. The path from Reading Ability to Print Exposure was substantial and dropping this path (Figure 5, centre left) resulted in a large deterioration in fit: $\Delta\chi^2(df = 1, N = 6,072 \text{ pairs}) = 152.02, p < .001, \Delta AIC = 150.02, \Delta BIC = 143.31$. However, the path from Print Exposure to Reading Ability in the bidirectional model was not significant ($p = .254$) and could be dropped (Figure 5 right): $\Delta\chi^2(df = 1, N = 6,072 \text{ pairs}) = 1.35, p = .246, \Delta AIC = -0.65, \Delta BIC = -7.37$. Finally, we tested the model in which the association between Reading Ability and Print Exposure was modelled as both a direct effect (flowing from Reading Ability to Print Exposure) and common genetic influences. We opted for adding a shared genetic effect (i.e. rA), rather than a shared nonshared environmental effect (i.e. rE) because rE hardly contributes to the phenotypic association (estimated at only 4%). The fit of this final model (Figure 5, bottom left) was exactly the same as the correlational model in Figure 4. The rA was just .02, so could be left out. To conclude, the unidirectional causal model 'reading ability \rightarrow print exposure' is both parsimonious and supported by the data.

Discussion

The present study applied direction of causation modelling (Duffy & Martin, 1994; Heath et al., 1993) to a large twin dataset to assess the relationships between reading ability and print exposure. We found evidence for a causal influence of reading ability on print exposure, consistent with previous findings from behavioural studies (Aarnoutse & van Leeuwe, 1998; Harlaar et al., 2011; Leppänen et al., 2005). Our findings refute the common belief that there is an influence of print exposure on reading ability, or that there are reciprocal influences between them. Interestingly, according to a Twitter poll, only 6% of people responding thought that reading ability \rightarrow print exposure (van Bergen, 2017).

The finding that reading ability is the driver of print exposure does not, of course, imply that exposure to print and thus exposure to orthographic forms is irrelevant to learning to read. To become a skilled reader, it is undoubtedly important to develop detailed lexical representations of words (Nation, 2017; Perfetti, 2007). However, while this may take as little as a single exposure in some readers (Tamura, Castles, & Nation, 2017), in poor readers, it takes much longer to consolidate new learning (Share & Shalev, 2004). Moreover, although a fair assumption is that schools provide the necessary practice, measures of print exposure on which good and poor readers differ, tap reading outside of school hours. We demonstrate here that whether children choose to read for themselves depends, in part, on their reading ability, underlining the fact that poor

readers choose to read less. In fact, we found that reading ability accounted for 16% of the variance in print exposure. In addition, other influences, both genetic and shared environmental, are also at play.

According to the unexplained variance in the model, additional genetic factors that influence the amount children read are independent of those influencing reading skills. We speculate that they may include inherited factors associated with ADHD symptomatology, such that more restless children are less likely to sit down and read than those with good attentional control. Likewise, we believe that the shared-environmental component may partly reflect the values of parents, the supply of books at home and the importance school places on book reading outside of the classroom curriculum. Our findings also highlight that print exposure is not something imposed upon the child. Rather the fact that it depends on environmental stimulation (or the absence thereof) and on innate child factors suggests that gene-environment correlation is important, both passive, evocative and active *gene-environment* (*g-e*) correlation (Plomin, DeFries, & Loehlin, 1977; van Bergen, van der Leij, & de Jong, 2014).

An objection to the present conclusion is that we only measured reading at one point in time, during the early stages of reading acquisition (1 year after the commencement of reading instruction in the Netherlands). Moreover, the data cannot speak to how the home literacy environment might influence, not only children's reading development (Hamilton, Hayiou-Thomas, Hulme, & Snowling, 2016; Sénéchal & LeFevre, 2002), but also their motivation to read. However, we argue that, given the stability of reading over time, it is unlikely that later levels of print exposure could account for future growth in reading, or override the powerful *g-e* correlation that manifest itself in dyslexic readers choosing literary activities less. We contend, nonetheless, that longitudinal data are required to validate these assumptions. We also acknowledge weaknesses in the measure of print exposure used. In the light of the increased use of screen time, e-readers and other technologies, the use of parent's ratings of how many books children read outside of school may be questioned as a way of measuring print exposure. On the other hand, studies have shown that print exposure from digital sources (like email, Wikipedia, Facebook) is unrelated or only weakly related to reading ability (McGeown, Osborne, Warhurst, Norgate, & Duncan, 2016; Pfof, Dörfler, & Artelt, 2013).

In studying causality, both observational and intervention studies are important, but they address different questions. We observed children's reading skills and reading habits and investigated their origins in natural settings. Observational research like this addresses the 'what is' question. In contrast, intervention studies address the 'what could be' question (Plomin, Shakeshaft, McMillan, & Trzaskowski, 2014). By studying causality in the natural situation, we demonstrated that reading ability

drives print exposure. Ultimately, it is perhaps not surprising that in natural settings poorer readers choose to read less in their spare time. Intervention studies cannot demonstrate causal processes in the natural situation, and conversely, an observational study like the current one cannot say what effects could be achieved by intervention. For instance, our study does not rule out potential benefits of getting children to read more than they would normally choose to. Likewise, a successful reading intervention may well result in poor readers wanting to read more, and gains in reading skill may mediate an increase in print exposure. To our knowledge, no such controlled trial has been conducted. More generally, it is known that successful reading interventions use evidenced-based programmes with individualized instruction and may need to be offered for up to several years to produce lasting effects (McDonald Connor et al., 2013; Regtvoort, Zijlstra, & van der Leij, 2013). There is also evidence that interventions that focus on phonological awareness and alphabetic skills do not improve reading outcomes unless emergent phonological skills are practised in the context of reading texts or books (Hatcher, Hulme, & Ellis, 1994; Wise, Ring, & Olson, 2000). If true, even though print exposure does not causally influence reading as measured in this study, moving children beyond their 'natural' amount of reading may be a sensible target of intervention, alongside improving decoding skills.

By using direction of causality modelling, we extend what is known about the context in which children's reading skills develop. We show that it is a useful technique for understanding individual differences in reading attainment and the factors which determine the enjoyment (or dislike) of reading. We endorse previous findings of a genetic influence on word-level reading and extend this to show that the same genetic factors influence print exposure causatively and this, in turn, depends on additional genetic and environmental factors.

Conclusions

We dissected the association between 7½-year-old children's reading ability and reading frequency and volume (called print exposure). We confirmed that individual differences in reading ability were mostly due to genetic differences, while print exposure was equally genetic and environmental in origin.

Importantly, we found evidence that children's reading level fuels how much they choose to read – it follows, as practitioners know, that children tend to avoid reading if they find it difficult. Interventions should focus not only on promoting reading skills but also motivation to read.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Appendix S1. Exclusion of participants, same versus different classrooms and missing data.

Table S1. Descriptive sibling data statistics and test statistics for the twin-sib comparison.

Table S2. Missing value analyses: results of *t*-tests comparing whether the groups with and without missing data differ on mother's educational level.

Table S3. Missing value analyses: results of *t*-tests comparing whether the groups with and without missing data differ on factor scores of reading ability.

Acknowledgements

The Netherlands Twin Register (NTR) is supported by grants from the Netherlands Organization for Scientific Research (NWO), including NWO VENI grant 451-15-017 awarded to EvB ('Decoding the gene-environment interplay of reading ability'), and NWO grant 480-15-001/674 ('Netherlands Twin Registry Repository: researching the interplay between genome and environment'), NWO Gravitation grant 024.001.003 ('the Consortium on Individual Development (CID)'). Furthermore, the NTR is supported by the Avera Institute for Human Genetics, the Royal Netherlands Academy of Science's Professor Award (PAH/6635) to DIB, and the European Union Seventh Framework Programme (FP7/2007-2013, 602768, project ACTION: 'Aggression in Children: Unravelling gene-environment interplay to inform Treatment and Intervention strategies'). The authors gratefully acknowledge the ongoing contribution of the participants in the NTR, including twins, their families and teachers. The authors have declared that they have no competing or potential conflicts of interest.

Correspondence

Elsje van Bergen, Vrije Universiteit Amsterdam - Biological Psychology, Van der Boeorchorststraat 1, Amsterdam, 1081 BT, The Netherlands; Email: e.van.bergen@vu.nl; Twitter: @drElsje

Key points

- Dyslexic readers typically have low levels of print exposure.
- It cannot be said that poor reading is due to limited reading exposure.
- Individual differences in reading ability are mostly due to genetic differences, whereas individual differences in print exposure have equal genetic and environmental origins.
- Individual differences in reading ability predict print exposure, rather than vice versa.

- Practitioners should not take this to mean that interventions to promote print exposure cannot improve reading skills.
- Encouraging children to read more is of itself a sensible target.

References

- Aarnoutse, C., & van Leeuwe, J. (1998). Relation between reading comprehension, vocabulary, reading pleasure, and reading frequency. *Educational Research and Evaluation, 4*, 143–166.
- Anderson, R.C., Wilson, P.T., & Fielding, L.G. (1988). Growth in reading and how children spend their time outside of school. *Reading Research Quarterly, 23*, 285–303.
- Cito. (2014). *Cito Leerling Volgstelsysteem*. Arnhem, the Netherlands: Cito.
- Conners, C.K., Sitarenios, G., Parker, J.D., & Epstein, J.N. (1998). Revision and restandardization of the Conners Teacher Rating Scale (CTRS-R): Factor structure, reliability, and criterion validity. *Journal of Abnormal Child Psychology, 26*, 279–291.
- Cunningham, A.E., & Stanovich, K.E. (1997). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology, 33*, 934–945.
- de Zeeuw, E.L., de Geus, E.J., & Boomsma, D.I. (2015). Meta-analysis of twin studies highlights the importance of genetic variation in primary school educational achievement. *Trends in Neuroscience and Education, 4*, 69–76.
- de Zeeuw, E.L., van Beijsterveldt, C.E., Glasner, T.J., de Geus, E.J., & Boomsma, D.I. (2016). Arithmetic, reading and writing performance has a strong genetic component: A study in primary school children. *Learning and Individual Differences, 47*, 156–166.
- Duffy, D.L., & Martin, N.G. (1994). Inferring the direction of causation in cross-sectional twin data: Theoretical and empirical considerations. *Genetic Epidemiology, 11*, 483–502.
- Ebejer, J.L., Coventry, W.L., Byrne, B., Willcutt, E.G., Olson, R.K., Corley, R., & Samuelsson, S. (2010). Genetic and environmental influences on inattention, hyperactivity-impulsivity, and reading: Kindergarten to grade 2. *Scientific Studies of Reading, 14*, 293–316.
- Gillespie, N.A., Zhu, G., Neale, M.C., Heath, A.C., & Martin, N.G. (2003). Direction of causation modeling between cross-sectional measures of parenting and psychological distress in female twins. *Behavior Genetics, 33*, 383–396.
- Hamilton, L.G., Hayiou-Thomas, M.E., Hulme, C., & Snowling, M.J. (2016). The home literacy environment as a predictor of the early literacy development of children at family-risk of dyslexia. *Scientific Studies of Reading, 20*, 401–419.
- Harlaar, N., Dale, P.S., & Plomin, R. (2007). Reading exposure: A (largely) environmental risk factor with environmentally-mediated effects on reading performance in the primary school years. *Journal of Child Psychology and Psychiatry, 48*, 1192–1199.
- Harlaar, N., Deater-Deckard, K., Thompson, L.A., DeThorne, L.S., & Petrill, S.A. (2011). Associations between reading achievement and independent reading in early elementary school: A genetically informative cross-lagged study. *Child Development, 82*, 2123–2137.
- Harlaar, N., Trzaskowski, M., Dale, P.S., & Plomin, R. (2014). Word reading fluency: Role of genome-wide single-nucleotide polymorphisms in developmental stability and correlations with print exposure. *Child Development, 85*, 1190–1205.
- Hatcher, P.J., Hulme, C., & Ellis, A.W. (1994). Ameliorating early reading failure by integrating the teaching of reading and phonological skills: The phonological linkage hypothesis. *Child Development, 65*, 41–57.
- Heath, A.C., Kessler, R.C., Neale, M.C., Hewitt, J.K., Eaves, L.J., & Kendler, K.S. (1993). Testing hypotheses about direction of causation using cross-sectional family data. *Behavior Genetics, 23*, 29–50.
- Leppänen, U., Aunola, K., & Nurmi, J.E. (2005). Beginning readers' reading performance and reading habits. *Journal of Research in Reading, 28*, 383–399.
- Martin, N.W., Hansell, N.K., Wainwright, M.A., Shekar, S.N., Medland, S.E., Bates, T.C., ... & Wright, M.J. (2009). Genetic covariation between the Author Recognition Test and reading and verbal abilities: What can we learn from the analysis of high performance? *Behavior Genetics, 39*, 417–426.
- McDonald Connor, C., Morrison, F.J., Fishman, B., Crowe, E.C., Al Otaiba, S., & Schatschneider, C. (2013). A longitudinal cluster-randomized controlled study on the accumulating effects of individualized literacy instruction on students' reading from first through third grade. *Psychological Science, 24*, 1408–1419.
- McGeown, S.P., Osborne, C., Warhurst, A., Norgate, R., & Duncan, L.G. (2016). Understanding children's reading activities: Reading motivation, skill and child characteristics as predictors. *Journal of Research in Reading, 39*, 109–125.
- Mol, S.E., & Bus, A.G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin, 137*, 267.
- Muthén, L.K., & Muthén, B.O. (1998-2015). *Mplus User's Guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Nation, K. (2017). Nurturing a lexical legacy: Reading experience is critical for the development of word reading skill. *npj Science of Learning, 2*, 3.
- Olson, R.K., Keenan, J.M., Byrne, B., & Samuelsson, S. (2014). Why do children differ in their development of reading and related skills? *Scientific Studies of Reading, 18*, 38–54.
- Perfetti, C. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading, 11*, 357–383.
- Pfost, M., Dörfler, T., & Artelt, C. (2013). Students' extracurricular reading behavior and the development of vocabulary and reading comprehension. *Learning and Individual Differences, 26*, 89–102.
- Plomin, R., DeFries, J.C., & Loehlin, J.C. (1977). Genotype-environment interaction and correlation in the analysis of human behavior. *Psychological Bulletin, 84*, 309.
- Plomin, R., Shakeshaft, N.G., McMillan, A., & Trzaskowski, M. (2014). Nature, nurture, and expertise. *Intelligence, 45*, 46–59.
- Rebollo, I., de Moor, M.H., Dolan, C.V., & Boomsma, D.I. (2006). Phenotypic factor analysis of family data: Correction of the bias due to dependency. *Twin Research and Human Genetics, 9*, 367–376.
- Regtvoort, A., Zijlstra, H., & van der Leij, A. (2013). The effectiveness of a 2-year supplementary tutor-assisted computerized intervention on the reading development of beginning readers at risk for reading difficulties: A randomized controlled trial. *Dyslexia, 19*, 256–280.
- Rietveld, M., van Der Valk, J., Bongers, I., Stroet, T., Slagboom, P., & Boomsma, D. (2000). Zygosity diagnosis in young twins by parental report. *Twin Research and Human Genetics, 3*, 134–141.
- Sénéchal, M., & LeFevre, J.-A. (2002). Parental involvement in the development of children's reading skill: A five-year longitudinal study. *Child Development, 73*, 445–460.
- Share, D.L., & Shalev, C. (2004). Self-teaching in normal and disabled readers. *Reading and Writing, 17*, 769–800.
- Tamura, N., Castles, A., & Nation, K. (2017). Orthographic learning, fast and slow: Lexical competition effects reveal the time course of word learning in developing readers. *Cognition, 163*, 93–102.

- Thomsen, S.F., van der Sluis, S., Stensballe, L.G., Posthuma, D., Skytthe, A., Kyvik, K.O., ... & Bisgaard, H. (2009). Exploring the association between severe respiratory syncytial virus infection and asthma: A registry-based twin study. *American Journal of Respiratory and Critical Care Medicine*, *179*, 1091–1097.
- Toulopoulou, T., Van Haren, N., Zhang, X., Sham, P., Cherny, S., Campbell, D., ... & Pol, H. (2015). Reciprocal causation models of cognitive vs volumetric cerebral intermediate phenotypes for schizophrenia in a pan-European twin cohort. *Molecular Psychiatry*, *20*, 1386.
- van Beijsterveldt, C.E.M., Groen-Blokhuis, M., Hottenga, J.J., Franic, S., Hudziak, J.J., Lamb, D., ... & Schutte, N. (2013). The Young Netherlands Twin Register (YNTR): Longitudinal twin and family studies in over 70,000 children. *Twin Research and Human Genetics*, *16*, 252.
- van Bergen, E. (2017). Children who read more (i.e. have more print exposure) are better readers. Why do you think this is? Retrieved from <https://twitter.com/drElsje/status/874538280628486144> [Twitter poll 13th June 2017].
- van Bergen, E., van der Leij, A., & de Jong, P.F. (2014). The intergenerational multiple deficit model and the case of dyslexia. *Frontiers in Human Neuroscience*, *8*, 1–13.
- Verhoeven, L. (1995). *Drie-Minuten-Toets. Handleiding [Three-Minutes-Test. Manual]*. Arnhem, the Netherlands: Cito.
- Verhoeven, L., & van Leeuwe, J. (2009). Modeling the growth of word-decoding skills: Evidence from Dutch. *Scientific Studies of Reading*, *13*, 205–223.
- Wise, B.W., Ring, J., & Olson, R.K. (2000). Individual differences in gains from computer-assisted remedial reading. *Journal of Experimental Child Psychology*, *77*, 197–235.

Accepted for publication: 12 March 2018