

Cross-Study Differences in the Etiology of Reading Comprehension: a Meta-Analytical Review of Twin Studies

Callie W. Little¹ · Rasheda Haughbrook¹ · Sara A. Hart^{1,2}

Received: 30 March 2016 / Accepted: 24 August 2016 / Published online: 14 September 2016
© Springer Science+Business Media New York 2016

Abstract Numerous twin studies have examined the genetic and environmental etiology of reading comprehension, though it is likely that etiological estimates are influenced by unidentified sample conditions (e.g. Tucker-Drob and Bates, *Psychol Sci*:0956797615612727, 2015). The purpose of this meta-analysis was to average the etiological influences of reading comprehension and to explore the potential moderators influencing these estimates. Results revealed an average heritability estimate of $h^2 = 0.59$, with significant variation in estimates across studies, suggesting potential moderation. Moderation results indicated publication year, grade level, project, zygosity methods, and response type moderated heritability estimates. The average shared environmental estimate was $c^2 = 0.16$, with publication year, grade and zygosity methods acting as significant moderators. These findings support the role of genetics on reading comprehension, and a small significant role of shared environmental influences. The results suggest that our interpretation of how genes and environments influence reading comprehension should reflect aspects of study and sample.

Keywords Meta analysis · Reading comprehension · Heritability · Genetics of reading · Etiology of reading

Edited by Carol Van Hulle.

✉ Callie W. Little
calliew01@gmail.com

¹ Department of Psychology, Florida State University, 1107 W. Call Street, Tallahassee, FL 32306, USA

² Florida Center for Reading Research, Florida State University, 2010 Levy Avenue, Tallahassee, FL 32310, USA

Introduction

Numerous twin studies have been published examining the genetic and environmental contributions to reading, with a more recent focus directed towards reading comprehension (e.g. Harlaar et al. 2010; Logan et al. 2013; Soden et al. 2015). Recent work has highlighted that the genetic and environmental influences on any given outcome derived from genetically sensitive studies are susceptible to moderation by aspects of sample and outcome (Tucker-Drob and Bates 2015). For example, age, nationality and income could influence estimates at the sample level, potentially over or underinflating sources of genetic and environmental influence. Moreover, the type of measurement instrument could potentially influence etiological estimates at the outcome level through the capture of additional constructs (Hart et al. 2013a, b; Keenan et al. 2008). It is important to identify moderators so that we can improve our interpretation of how genes and environment influence the increasingly important construct of reading comprehension.

The increase in genetically-sensitive studies on reading comprehension follows an international collective interest in improving reading comprehension levels (UNESCO 2009; National Center for Education Statistics 2013; OECD 2015). Within the US, the 2013 National Assessment of Educational Progress, reported only 35 % of fourth graders scored at or above proficiency, with 32 % scoring below basic reading levels (National Center for Education Statistics 2013). Students performing at the basic skill level should be able to make simple inferences and interpret meanings of words used in text. Being unable to read at a basic level indicates severe challenges to future academic success within the US (Chall and Jacobs 2003), and, furthermore, to adequate functioning in today's society

(Alfassi 2004). Several reading component skills (i.e. decoding and fluency) have also been identified in association with academic success and wellbeing within the extant literature; however, reading comprehension is commonly considered a broader, representative skill which encompasses these subcomponent skills in both the phenotypic (Kim et al. 2010; Petscher and Kim 2011; Roberts et al. 2005) and genetically-sensitive literature (e.g. Christopher et al. 2013; Little and Hart 2016). At the international level, a 2012 reading assessment, Programme for International Student Assessment (PISA), ranked US 15-year-old girls 18th out of 39 countries on reading comprehension performance and US 15-year-old boys 16th out of 39 countries (OECD 2015). The US PISA scores lie close to the medians suggesting that other nations' reading comprehension scores may also be at risk. Furthermore, at the US and international levels, economic growth has been substantially linked with educational outcomes such as reading comprehension, further suggesting the global importance of reading comprehension ability (Hanushek and Woessmann 2012).

Reading comprehension has been established as an important predictor of academic success and overall wellbeing, which emphasizes the need to understand the underlying etiological influences on reading comprehension (Berkman et al. 2011; Francis et al. 1994). Potential etiological influences include biological, specifically genetics, as well as environmental influences like educational resources available in the home and in schools, or from other aspects such as socioeconomic factors or the neighborhood. Twin studies are a common method of obtaining information on the genetic and environmental etiology of an outcome such as reading comprehension. Twin studies compare monozygotic (MZ) twins, who share 100 % of their segregating genetic material, to dizygotic (DZ) twins, who share approximately 50 % of their segregating genetic material on average (i.e. additive genetic influences, or the average effects of alleles on a trait; Plomin et al. 2013). Both MZ and DZ twin pairs who are reared together are assumed to share 100 % of their shared environmental influences (i.e. influences that serve to make members of a twin pair more similar to each other). Using these known relations between MZ and DZ twins the variance in a trait of interest can be decomposed into genetic influences, shared environmental influences and non-shared environmental influences (i.e. influences that serve to make members of a twin pair less similar to each other; plus error). To the extent that MZ twins are more similar than DZ twins on a particular outcome, additive genetic influences, labelled heritability (h^2), are assumed. Alternatively, if MZ twins are less than two times as similar as DZ twins are on a particular trait, shared environmental (c^2) influences are inferred. When correlations

between MZ pairs do not equal unity, non-shared environmental influences (e^2) are indicated.

Background

Studies using genetically sensitive samples have revealed that genetic influences account for between 30 and 80 % of the variance in reading comprehension, across studies incorporating a wide range of ages and other demographic characteristics (e.g. Byrne et al. 2009; Harlaar et al. 2010; Kovas et al. 2007). In total, the evidence suggests a significant effect of genetic influences on reading comprehension ability, but, reveals a wide range in the published magnitude of these genetic estimates. This range in genetic estimates may be due to moderators within and across the characteristics of these studies, samples and outcomes.

Meta-analytic techniques allow for a quantitative synthesis of extant literature on a specific topic. More specifically, meta-analyses convert statistical measures such as effect sizes or correlation coefficients into a common metric and use this metric to calculate an average weighted estimate across studies (Arthur et al. 2001). A meta-analytic review of twin studies on reading comprehension can serve to determine a weighted average estimate of the genetic and environmental influence across a single population or multiple populations. Additionally, meta-analyses can explore potential heterogeneity within these influences by identifying and systematically testing potential moderators. To date, one meta-analysis of twin studies on educational achievement has been conducted, with six studies with heritability estimates for reading comprehension included (de Zeeuw et al. 2015). The authors report an average heritability estimate of $h^2 = 0.49$, and average shared environment estimate of $c^2 = 0.13$, though there was significant heterogeneity in both of these estimates across studies. The twin samples included in this review were primarily from the United States (US), the United Kingdom (UK) and The Netherlands (NL), and the authors found that country significantly moderated the heritability estimates of reading comprehension.

This meta-analysis was an important first step, but it was limited in two ways. First, this review was part of a larger one concentrating on a range of achievement outcomes, and it left out available studies that included reading comprehension. In particular, samples including children from more diverse racial and ethnic, and socioeconomic backgrounds were not included (e.g. Hart et al. 2010a; Soden et al. 2015), and this omission may have resulted in a restricted overall representation of the relative magnitudes of influence contributing to reading comprehension ability (Tucker-Drob and Bates 2015). Second, no other

moderators beyond country of origin were examined, and it is likely that there are other moderators causing magnitude differences in the etiological influences on reading comprehension between studies.

Moderators

Year of publication is often evaluated as a potential moderator of effect sizes (Cooper et al. 2009, p. 25). Across a range of years, effect sizes may alter due to time-dependent cohort effects such as changes in measurement methods, demographic trends or major societal events which have the potential to influence sample characteristics and sensitivity of effect size calculation (Cooper et al. 2009, p. 463). Coding and evaluating publication year as a potential moderator of heritability can point to areas of influence such as cohort effects within the current synthesis. Another potential moderator of etiological influences on reading comprehension is age. Several previous longitudinal studies have found evidence for the increase of genetic influences, and decrease of shared environmental influences, on reading comprehension across childhood and adolescence (Byrne et al. 2009; Logan et al. 2013). These developmental changes may be due to the activation of new genes which influence reading comprehension and/or due to active gene-environment correlations where individuals select into environments more suitable for the promotion of reading comprehension skills as they mature (Haworth et al. 2010). By examining age (and grade) as a potential moderator in the present study, we are able to test whether the increase of heritability (and decrease of shared environmental influences) found in specific studies in the literature holds across many studies representing a large range of sample characteristics and developmental time points.

Socioeconomic status (SES) is also a potential moderator of the etiological influences of reading comprehension as suggested by results of several studies (e.g. Hart et al. 2013a, b). The broader literature has often found that SES is a significant moderator of the heritability of general cognitive ability, with heritability increasing at higher levels of SES (e.g. Turkheimer et al. 2003). However, the results when looking at reading outcomes has been more mixed, with some work indicating no SES moderation on early reading skills (Tucker-Drob and Bates 2015), some finding increased heritability on reading outcomes at higher levels of SES (Friend et al. 2008) and others finding higher heritability estimates for reading comprehension at lower levels of SES (Hart et al. 2013a, b). Analyzing SES as a moderator of heritability estimates on reading comprehension across a large number of aggregated studies can

serve to elucidate the general direction of influence and to clarify the inconsistencies within the literature. Moreover, social disorganization theory suggests many outcomes, including achievement, are influenced by shared values, social relationships, and the capability to achieve goals shared within the community (Bowen et al. 2002). Social disorganization has been found to be higher for neighborhoods with lower socioeconomic status, leading to increased residential turnover, lower social cohesion and control, higher rates of crime and delinquency as well as reduced access to human and physical resources (Bowen et al. 2002; Bumgarner and Brooks-Gunn 2013). In 2009, US Caucasian families held assets over 20 times greater than African American families on average (Killewald 2013) suggesting that predominantly Caucasian samples may be more likely to have higher average SES, and less potential for greater social disorganization and the associated negative influences. Due to the close relation between race and SES within the US, the racial composition participants was also included as a potential moderator within the current synthesis.

Along with SES, other environmental factors may contribute to differences in estimates across studies. In the meta-analysis conducted by de Zeeuw et al. (2015), country was found to be a significant moderator for heritability, which indicates characteristics of the samples as well as the environment with which the samples may interact could influence the etiology of reading comprehension. Even within some countries, multiple twin projects have published reading comprehension outcomes, suggesting characteristics specific to a particular twin project might significantly moderate etiological influences. For example, in a country as large and diverse as the US, it is possible for twin studies within one geographic location to have different sample characteristics such as SES, racial and ethnic composition and urban or rural settings. Therefore, project may serve to be a potential moderator (mirroring the country as moderator finding of de Zeeuw et al. 2015).

Also, zygosity determination among twin projects relies on different methods: obtaining polymorphic DNA markers via buccal swabs or blood tests or a similarity questionnaire filled out by either twins or parents (Price et al. 2000). Comparisons of these methods have found twin similarity questionnaires to be over 90 % accurate in comparison with DNA-based methods (Kasriel and Eaves 1976; Price et al. 2000); however, a meta-analysis on the genetic and environmental influences of antisocial behavior found a significant moderation of heritability estimates by zygosity determination method (Rhee and Waldman 2002). Given this, zygosity determination method is a potential moderator of the etiology of reading comprehension.

Reading ability level, while generally thought to exist along a continuum (Fletcher et al. 2013; Bishop 2015), is also often classified in categorical terms such as reading disabled or dyslexic, typically developing readers, and gifted or advanced readers (Spencer et al. 2014; Brighton et al. 2015). Although genetic influences on reading status have been found to be consistent for those with differing levels of reading ability and disability (Bishop 2015), shared and non-shared environmental influences such as classroom instruction, home literacy environment or peer groups may differentially impact reading ability status (Rashid et al. 2005; Hart et al. 2013a, b). Within the present meta-analysis, the population of readers in a given sample may differ in ability status; therefore, etiological differences between these populations may exist. Establishing a pattern of moderation for environmental influences, but not heritability, by reading population can provide additional evidence to support a polygenically influenced continuum of reading ability. Finally, the type of outcome or assessment method used to obtain reading comprehension scores may have an impact on the resulting estimates. Unstandardized measurement instruments may lead to greater measurement error than standardized measures. Also, measures of reading comprehension may differ in the skills that they measure within and among developmental time points (Keenan et al. 2008). Reading comprehension consists of multiple cognitive processes such as phonological awareness, decoding and fluency, inferencing, vocabulary, working memory and other executive functioning skills that work in combination to produce text comprehension (Cain and Oakhill 2009; Fletcher et al. 2002; Jerman et al. 2012). Keenan et al. (2008) investigated the contributions of component skills to several reading comprehension assessments and found evidence for differences in the ways these underlying skills were being assessed across measures. Heritability estimates vary across subcomponent skills of reading, reading comprehension and other contributing cognitive functions (de Zeeuw et al. 2015; Polderman et al. 2015), suggesting the potential for moderation by reading comprehension assessments that measure multiple underlying skills.

Current study

The purpose of the current investigation is to conduct a meta-analysis in order to aggregate the genetic and environmental influences of reading comprehension from primary twin studies and to investigate whether sources of heterogeneity at levels of sample, study and outcome significantly moderate these influences. This study is novel in that it provides a comprehensive review of the sizeable literature on reading comprehension measured in

genetically sensitive samples. Also, it is the first meta-analysis to assess moderation of etiological influences on reading comprehension by publication year, age, grade, project, SES, nationality, race, zygosity, population, assessment type and response type. Such a systematic examination of genetically sensitive analyses of reading comprehension is crucial for the identification and reduction of bias among reported estimates that may be caused by these or as of yet, unidentified moderators. Furthermore, establishing trends for changes in the etiological influences on reading comprehension by selected moderators can serve to support or contradict individual conclusions drawn from previous research.

Method

Search procedure and coding scheme

Criteria for studies to be included into the current meta-analysis were: (1) reported twin intraclass correlations or univariate heritability, shared and non-shared environmental estimates from genetically sensitive analyses: (2) a measure of reading comprehension. Eight separate searches using the search terms “genetic influences on reading comprehension” and “reading comprehension twins” were conducted using PsychInfo, ProQuest, EBSCO and ERIC to locate published articles that focused on genetic influences on reading comprehension. Of the original 7186 results returned, a review of the titles and abstracts led to the exclusion of 7134, resulting in 52 articles for coding. An additional 7 studies were located by asking experts in the field and searching through reference lists from the 52 articles found through the database search. Of the final 59, 22 studies were excluded because the measure of reading was not truly a reading comprehension measure or a composite reading score which also included a measure of a different component skill of reading. The remaining 37 studies were coded at three levels: study, sample, and outcome. Study characteristics coded were publication year, zygosity determination method, intraclass correlations or the univariate variance estimate, and sample size. Sample characteristics coded were population (e.g. learning disabled, dyslexic), nationality, project, race, SES, grade, and age. Finally, outcome-level variables were coded for assessment method including type of outcome measure (e.g. researcher-created or standardized) and response type (e.g. cloze, multiple choice).

For all categorical moderators, if 75 % or more of the sample was identified within one category, the entire sample was coded for that category. For example, if 75 % or more of participants within a study were reported as using a zygosity questionnaire, the sample was coded for

‘questionnaire’ as the zygosity determination method. When the sample included a blend (no single category representing 75 % or more of the sample), it was coded as ‘blended’. Tables 1 and 2 presents all studies with coded moderators (categorical and continuous) and the legend presents the identified categories for each categorical moderator. All variables were initially coded by the first author, and 50 % of the studies were subsequently coded by a second, trained individual to assess inter-coder reliability. The inter-coder reliability coefficients ranged from 0.95 to 1.00 for all coded variables. Any discrepancies were resolved through discussion.

Projects

Of the 37 studies coded, nine separate projects were identified: the Florida State Twin Registry (FTP), the Twins Early Development Study (TEDS), the Colorado Learning Disabilities Research Center Twin Study (CLDRC), the Western Reserve Reading and Math Project (WRRMP), the International Longitudinal Twin Study (ILTS), the Minnesota Study of Twins Reared Apart (MISTRA), the National Longitudinal Survey of Youth (NLSY), the Louisville Twin Study (LTS), and two independent studies coded as ‘Other’.

Florida state twin registry (FTP)

The Florida State Twin Registry was established in 2002 through a pilot project focusing on adult twins, but was expanded to include reading outcomes for school-aged children in 2006 through the Learning Disabilities Research Center at Florida State University (Taylor et al. 2013). As of 2013, the total sample consisted of 2591 MZ and DZ twins from 21 counties within the state of Florida. Potential twins were identified via the Progress Monitoring and Reporting Network (PMRN), a state-wide database of standardized assessments, and letters were mailed to these families asking them to participate in the study along with a short zygosity questionnaire. The resulting sample represents a more racially and ethnically diverse population than many existing twin projects (Taylor et al. 2013).

Twins early development study (TEDS)

Twins Early Development Study is a longitudinal twin study based in the United Kingdom (UK) and includes over 13,000 pairs (Oliver and Plomin 2007). Families of twins recruited into TEDS have been followed from early childhood through early adulthood. Twins born in England and Wales between 1994 and 1996 were identified through birth records obtained from the Office of National Statistics (ONS). Details of the recruitment procedures are available

from Trouton et al. (2002). The TEDS sample is representative of the UK with over 90 % of participants identified as Caucasian and approximately 46 % of mothers and 90 % of fathers reported employment (Oliver and Plomin 2007). Twins recruited into TEDS have been assessed on a large battery of health, behavioral, and cognitive traits including reading comprehension and related skills (Oliver and Plomin 2007).

Colorado learning disabilities research center twin study (CLDRC)

The CLDRC was founded as part of the broader learning disabilities center at the University of Colorado–Boulder in 1990 (Olson et al. 2013). This project focuses on the genetic and environmental influences of reading, reading related skills, cognitive and behavioral outcomes (e.g. executive functioning, ADHD) and writing and consists of MZ and DZ twins and their siblings from 27 Colorado school districts (DeFries 1997). Twins were recruited into the project if at least one member of the twin pair was identified as having a reading difficulty or ADHD symptoms at a ratio of 2:1 for control families in which neither member of the twin pair reported problems (Arnett et al. 2015; DeFries 1997). A recent publication reported the sample consisted of 2332 predominantly Caucasian twins (over 90 %) and siblings from English speaking families ranging in age from 8 to 19 years (Arnett et al. 2015).

Western reserve reading and math project (WRRMP)

Based in Ohio, the WRRMP is a longitudinal project of over 400 families of school-aged twins and their siblings which examines reading, math, and related cognitive outcomes (Hart et al. 2009; Petrill et al. 2006). Petrill et al. (2006) describe the initial sample and recruitment process. Participants were recruited through personal interaction, media advertisements, and schools in the Cleveland, Columbus and Cincinnati areas of Ohio along with Western Pennsylvania. Schools in these areas mailed packets to parents of twins who were enrolled in Kindergarten though not yet finished 1st grade to request participation in the study. The resulting sample was predominantly Caucasian (greater than 90 %) and comprised a wide range of SES. Twins were assessed on a range of cognitive skills over 7 years and 9 waves of testing (Hart et al. 2016).

International longitudinal twin study (ILTS)

The ILTS is a longitudinal study which investigates literacy skills and consists of MZ and DZ twin pairs aged 4–10 from the US, Australia, Norway and Sweden (Byrne et al.

Table 1 Coded studies

References	Study type	Project	Zygoty	h^2	c^2	e^2	MZ ICC	DZ ICC	Sample n	Population	Assessment type	Assessment type	Method	SES	Race	Nationality	Mean age	Grade
Matheny and Brown-Dolan (1974)	1	9	1	0.66	0.19	0.15	0.85	0.52	140	–	13	–	–	3	–	–	–	–
Harris (1982)	1	8	4	0.38	0.41	0.21	0.79	0.6	176	0	2	1	1	1	0	1	7.73	2
Stevenson et al. (1987)	1	8	3	0.44	0.27	0.29	0.71	0.49	570	0	12	2	1	–	0	5	13.0	–
Brooks et al. (1990)	1	3	3	0.27	0.19	0.53	–	–	292	0	1	5	1	–	0	1	12	–
Thompson et al. (1991)	1	4	3	0.27	0.66	0.07	–	–	556	0	11	2	1	–	0	1	9.8	–
Alarcón et al. (1994)	5	3	3	0.14	0.22	0.64	0.36	0.29	628	4	1	5	1	–	0	1	12	–
Petrill and Thompson (1994)	1	4	3	0.46	0.41	0.13	–	–	283	0	11	2	1	–	0	1	9.55	–
Rodgers et al. (1994)	1	7	–	0.50	0.14	0.36	–	–	1322	4	1	5	1	3	5	1	–	–
Davis et al. (2001)	1	3	3	0.52	0.00	0.48	0.52	0.26	1174	2	1	5	1	–	0	1	11.7	–
Johnson et al. (2005)	3	6	–	0.51	0.00	0.49	0.36	0.59	239	–	7	7	2	3	–	6	50	–
Keenan et al. (2006)	1	3	3	0.51	0.018	0.31	–	–	382	0	4	–	1	–	0	1	11	–
Harlaar et al. (2007)	1	2	1	0.67	0.11	0.22	0.79	0.47	6832	–	5	4	1	2	0	5	7	–
Petrill et al. (2007)	1	4	1	0.76	0.11	0.13	–	–	422	0	2	1	1	3	0	1	7.2	–
Byrne et al. (2007)	1	5	1	0.76	0.03	0.21	0.77	0.44	638	0	2	1	1	3	0	1	6.1	–
Kovas et al. (2007)	1	2	3	0.43	0.20	0.37	–	–	589	0	1	5	1	–	0	5	10	–
Betjemann et al. (2008)	1	3	–	0.67	0.04	0.29	–	–	314	4	1	5	1	–	0	1	11.0	–
Davis et al. (2008)	1	2	–	0.38	0.25	0.37	–	–	5082	4	1	5	1	1	0	5	–	–

Table 1 continued

References	Study type	Project	Zygoty	h^2	c^2	e^2	MZ ICC	DZ ICC	Sample n	Population	Assessment type	Method	SES	Race	Nationality	Mean age	Grade
Byrne et al. (2009)	1	5	1	0.54	0.18	0.28	0.72	0.45	810	-	2	1	-	0	1	8.4	2
	1	5	1	0.76	0.05	0.29	0.71	0.33	270	-	2	1	-	0	2	7.9	2
	1	5	1	0.60	0.16	0.24	0.76	0.46	150	-	2	1	-	0	3	8.7	2
	1	5	1	0.67	0.07	0.26	-	-	1230	-	2	1	-	0	4	-	-
	1	4	1	0.32	0.42	0.26	0.74	0.58	384	0	2	1	-	0	1	-	1
Byrne et al. (2010)	1	4	1	0.82	0.00	0.28	0.82	0.40	560	0	2	1	-	0	1	-	1
	1	4	1	0.32	0.40	0.28	0.72	0.56	256	0	2	1	-	0	1	-	2
	1	4	1	0.64	0.06	0.30	0.70	0.38	526	0	2	1	-	0	1	-	2
	1	4	1	0.62	0.20	0.18	0.82	0.51	206	0	2	1	-	0	1	-	1
	1	4	1	0.74	0.00	0.26	0.66	0.29	224	0	2	1	-	0	1	-	1
Harlaar et al. (2010)	1	4	1	0.48	0.28	0.24	0.76	0.52	154	0	2	1	-	0	1	-	2
	1	4	1	0.5	0.16	0.34	0.66	0.16	216	0	2	1	-	0	1	-	2
	1	4	1	0.58	0.00	0.42	0.6	0.23	434	0	1	5	3	0	1	9.9	-
	1	4	1	0.75	0.00	0.25	0.73	0.25	404	0	2	1	3	0	1	9.9	-
	5	7	-	0.43	0.26	0.31	-	-	7912	0	1	5	3	5	-	9.6	-
Hart et al. (2010a)	5	7	-	0.32	0.46	0.22	-	-	1666	2	1	5	3	5	-	9.6	-
	1	4	1	0.79	0.00	0.26	-	-	456	0	2	1	3	0	1	9.9	-
Hart et al. (2010b)	1	4	1	0.68	0.00	0.42	-	-	456	0	1	5	3	0	1	9.9	-
	1	2	1	0.40	0.15	0.45	0.55	0.35	9288	-	1	5	-	0	5	12.0	-
Davis et al. (2009)	1	2	1	0.34	0.15	0.21	0.49	0.32	8894	-	10	2	-	0	5	12.0	-
	1	3	3	0.66	0.23	0.10	-	-	644	4	3	2	-	0	1	11.0	-
	1	5	1	0.61	0.11	0.27	-	-	1660	0	2	1	-	0	6	8.0	-
Olson et al. (2011)	1	5	1	0.86	0.09	0.04	-	-	778	0	4	-	-	0	1	10.0	-
	5	8	-	0.53	-	-	-	-	270	0	2	1	-	0	6	-	-
Logan et al. (2011)	1	4	1	0.18	0.00	0.63	-	-	520	0	1	5	3	0	1	9.8	-
	1	4	1	0.00	0.19	0.52	-	-	520	0	2	1	3	0	1	9.8	-
Astrom et al. (2012)	1	3	3	0.67	0.17	0.16	-	-	1348	-	1	5	-	0	-	-	-
	1	4	1	0.39	0.39	0.22	-	-	1168	0	2	1	3	0	1	-	0
	1	4	1	0.32	0.50	0.18	-	-	1168	0	2	1	3	0	1	-	1
Logan et al. (2013)	1	4	1	0.44	0.47	0.09	-	-	1168	0	2	1	3	0	1	-	0
	1	4	1	0.72	0.05	0.23	-	-	1168	0	2	1	3	0	1	-	2
	1	4	1	0.55	0.23	0.22	-	-	1168	0	2	1	3	0	1	-	3
1	4	1	0.82	0.00	0.18	-	-	1168	0	2	1	3	0	1	-	4	

Table 1 continued

References	Study type	Project	Zygoty	h^2	c^2	e^2	MZ ICC	DZ ICC	Sample n	Population	Assessment type	Assessment type	Method	SES	Race	Nationality	Mean age	Grade
Olson et al. (2013)	1	3	3	0.83	0.05	0.12	0.88	0.47	512	0	4	–	1	–	0	1	11.0	–
Christopher et al. (2013)	1	5	1	0.71	0.10	0.19	–	–	960	0	2	1	1	3	0	1	7.4	1
	1	5	1	0.59	0.14	0.27	–	–	964	0	2	1	1	3	0	1	8.5	2
	1	5	1	0.68	0.04	0.28	–	–	928	0	2	1	1	3	0	1	10.5	4
Harlaar et al. (2012)	1	2	1	0.34	0.13	0.53	–	–	9544	–	1	5	1	–	0	5	11.9	–
	1	2	1	0.37	0.18	0.45	–	–	9730	–	10	2	1	–	0	5	11.9	–
Little et al. (2014)	1	1	3	0.55	0.20	0.25	–	–	1382	0	8	2	1	3	5	1	11.7	–
	1	1	3	0.52	0.14	0.34	–	–	1382	0	9	2	1	3	5	1	11.7	–
Hart et al. (2014)	1	1	3	0.55	0.22	0.23	–	–	466	0	8	2	1	3	5	1	10.7	–
Taylor and Hart (2014)	1	1	3	0.53	0.24	0.23	–	–	1350	0	8	2	1	3	5	1	10.1	–
Soden et al. (2015)	1	4	1	0.57	0.24	–	0.80	0.52	448	0	2	1	1	–	0	1	7.1	1
	1	4	1	0.52	0.18	–	0.72	0.46	406	0	2	1	1	–	0	1	–	2
	1	4	1	0.75	0.02	–	0.80	0.35	399	0	2	1	1	–	0	1	–	3
	1	4	1	0.73	0.02	–	0.75	0.33	367	0	2	1	1	–	0	1	–	4
	1	4	1	0.77	0.00	–	0.75	0.27	378	0	2	1	1	–	0	1	–	5
	1	4	1	0.74	0.01	–	0.74	0.35	243	0	2	1	1	–	0	1	–	6
	1	4	1	0.68	0.12	–	0.78	0.47	960	0	2	1	1	–	0	1	7.4	1
	1	4	1	0.58	0.14	–	0.69	0.46	964	0	2	1	1	–	0	1	–	2
	1	4	1	0.70	0.02	–	0.71	0.39	928	0	2	1	1	–	0	1	–	4
Grasby et al. (2015)	1	5	1	0.68	0.00	0.19	0.65	0.36	250	0	2	1	1	–	0	2	8.0	2
	1	5	1	0.45	0.12	0.43	0.51	0.38	250	0	4	2	1	–	0	2	8.7	2

Table 2 Legends

Category	Value description
Study type	1 = Twin 3 = Twin and adoption 5 = Family design
Project	1 = FTP 2 = TEDS 3 = CLDRC 4 = WRRMP 5 = ILTS 6 = MISTRA 7 = NLSY 8 = Other
Zygosity method	1 = Saliva 3 = Questionnaire 5 = Other
Population	0 = Regular 4 = Blended
Assessment	1 = PIAT 2 = WRMT-R 4 = Composite 5 = NC 7 = RCMISTRA 8 = FCAT 9 = FAIR RC 10 = GOAL 11 = MAT 12 = NARA 13 = CAT
Assessment type	1 = Cloze 2 = MC questions 3 = Blended 4 = Rating scale 5 = Picture selection 7 = Short answer/retell
Assessment method	1 = Standardized
SES	2 = High 3 = Blended
Race	0 = Caucasian 5 = Blended
Nationality	1 = US 2 = Australian 3 = Scandinavian 5 = UK 6 = Blended

2007). Grasby et al. (2015) reported twins were recruited in pre-school and followed through the first three years of formal schooling. Information on the sample, recruitment

and measures can be found in the preliminary results reported by Byrne and colleagues (Byrne et al. 2002).

Minnesota study of twins reared apart (MISTRA)

The MISTRA is unique from the other twin projects included in this meta-analysis in that it includes twins ranging in age from 18 to 79 rather than school-aged children (Johnson et al. 2005). Initiated in 1979, MISTRA includes twins who were separated near birth along with friends, spouses and family members (Johnson et al. 2005). Recruitment occurred through multiple sources and processes resulting in a diverse sample of over 230 twin pairs from multiple countries (Bouchard et al. 1990). Twins within this sample were assessed for a large and comprehensive variety of health, behavioral and cognitive measures including reading comprehension (Bouchard et al. 1990; Johnson et al. 2005).

National longitudinal survey of youth (NLSY)

The NLSY contains a large, longitudinal, nationally representative sample of 12,686 US men and women that was established in 1979 (Baker 1993). Individuals who were between the ages of 14 and 22 on December, 31st 1978 were recruited to participate in the study. From 1986, children of women in the original study have been assessed on measures of health, behavior, environmental conditions, home observations and cognitive measures. In an effort to include all children from each home in the sample, data were collected on siblings and cousins as well. More information about the full battery of assessments, recruitment, and sample demographics can be found online from the Center for Human Resource Research (2006). Genetically-sensitive studies utilizing twins, siblings and cousins from the NLSY child sample were included in our meta-analysis (Hart et al. 2010a; Rodgers et al. 1994). Etiological estimates in these studies were obtained through a linking algorithm developed to identify the level of genetic relatedness among many of the kinship pairs available in the NLSY data set (Rodgers 1996).

Louisville twin study (LTS)

The Louisville Twin Study began identification of twins in the greater Louisville area from hospital birth records in 1957 and was updated to include recruitment through Department of Health records in 1965 (Vandenberg et al. 1968). Over 500 families with twins up to age 15 were recruited and assessed on a battery of measures including cognition, personality, physical development and environmental influences (Rhea, 2015).

Measures

California achievement test (CAT)

The CAT (1963 norms) is a normed, standardized test of reading comprehension for grades K through 12 (Tiegs and Clark 1977). Children are assessed through multiple choice and open-ended questions.

Florida assessment for instruction in reading (FAIR)

The reading comprehension subtest of the FAIR is a computer-administered assessment of reading comprehension given to school-aged students in the state of Florida during the 2009–2010 through the 2012–2013 school years. Students read through narrative or expository passages and then answer multiple choice questions about the passages. The generic estimate of reliability from item-response theory ranges from 0.88 to 0.92 from grades 3 to 10 (http://www.fcrr.org/fair/Technical%20manual%20-%203-12-FINAL_2012.pdf).

Florida comprehensive assessment test (FCAT)

The FCAT is a standardized assessment that was given to students within the state of Florida, annually. The reading portion of the FCAT consists of several narrative and expository passages followed by multiple choice comprehension questions. Reliability estimates for FCAT Reading Comprehension ranged from 0.80 in 3rd grade to 0.87 in 10th grade in 2010–2011 (<http://readibank.com/wp-content/uploads/2015/04/FCAT-Reliability-and-Validity-Report.pdf>).

Global online assessment for learning (GOAL)

The GOAL is administered to students in the UK. This measure is designed to assess both literal and inferential comprehension of words, sentences, and short paragraphs via multiple choice questions (Global Online Assessment for Learning 2002).

Metropolitan achievement test (MAT)

The MAT is a norm-referenced, multiple choice assessment of reading comprehension. Reliability across forms and tests ranges from 0.79 to 0.98 for the sixth edition (Canney 1989).

MISTRA reading comprehension (MISTRA-RC)

A recall-format assessment of reading comprehension created for the MISTRA study, the MISTRA-RC contains

three passages which participants read aloud then recount from memory (Johnson et al. 2005).

National curriculum (NC)

The NC is a teacher assessment of student's reading comprehension abilities that is based on key stages within the UK National Curriculum (Department for Education and Employment, 2000). Teachers rate student's reading ability on five point Likert scales. Reliability between the NC reading tests and the NC teacher assessments have been reported at 0.80 (Dale et al. 2005).

Neale analysis of reading Ability (NARA)

With NARA, individuals read six passages of increasing difficulty and respond to comprehension questions from the passages. Reliability scores for the comprehension section range from 0.93 to 0.95 (Neale 1966, 1999).

Peabody individual achievement test (PIAT)

The PIAT is a norm-referenced assessment which includes a subtest of reading comprehension (Dunn and Markwardt 1970). Participants read several short passages then select the one of four pictures that best represents the meaning of the passage. Test–retest reliability for PIAT reading comprehension subtest is 0.64 (Dunn and Markwardt 1970) and for the PIAT-R ranged from 0.86 to 0.94 by grade (Markwardt 1989).

Woodcock–Johnson reading mastery test (WRMT-R)

The WRMT-R is a cloze format reading comprehension measure which requires participants to read multiple short passages and fill in the appropriate missing word for each passage (Woodcock, 1987). Reported split-half reliabilities for this assessment range from 0.73 to 0.94 (Woodcock 1987).

Analyses

A meta-analysis of 37 genetically sensitive studies was conducted to estimate the average magnitude of genetic and environmental influences on reading comprehension. Effect sizes were obtained from univariate heritability (h^2), shared environmental (c^2) and non-shared environmental (e^2) estimates reported in the synthesized studies or calculated using Falconer's formula from intraclass correlations reported in the studies (Falconer 1960). Univariate estimates were obtained from manifest reading comprehension measures rather than from latent variables in all but one

case¹ where estimates from observed variables were unavailable. Individual and aggregated effect sizes were standardized using Fisher's z transformation before analyses (Lipsey and Wilson 2001). Fisher's z transformation corrects for non-normality within the r distribution, stabilizing the variance of r across a normal distribution. The formula used for the z transformation was:

$$z = 0.5 \ln \frac{1 + ESr}{1 - ESr}$$

where ESr is the effect size of the etiological estimate of reading comprehension (i.e. heritability, shared environmental and nonshared environmental influence). This transformation was done for each etiological estimate for each study.

In order to interpret the results, following analyses, z -transformed effect sizes were converted back to regular h^2 , c^2 , and e^2 values using the following formula (Hedges and Olkin 1985).²

$$r = \frac{e^{2ESzr} - 1}{(e^{2ESzr} + 1)}$$

where $ESzr$ is the z -transformed effect size of the etiological estimate of reading comprehension (i.e. heritability, shared environmental and nonshared environmental influence). This conversion was done for each etiological estimate calculated after the meta-analytic analyses were conducted.

There were two potential sources of sample dependence that were accounted for. First, several of the 9 projects' samples were used across multiple studies. Second, within single studies, multiple estimates of the etiological influences on reading comprehension were sometimes reported (e.g. in longitudinal studies where they were reported by time point). To account for these potential sources of dependence, aggregation of etiological influences was done across three steps. First, effect sizes and sample sizes were aggregated across projects by taking an average. Four sets of estimates were averaged across FTP, 9 across TEDS, 11 across CLDRC, 30 across WRRMP, 15 across ILTS, 3 across NLSY and 3 from studies coded as 'Other.' Two projects, MISTRA and LTS, donated a single estimate each and did not need to be aggregated. Secondly, effect sizes and sample sizes were aggregated across grade levels, in that 2 sets of estimates were averaged for Kindergarten, 9 across 1st grade, 14 across 2nd grade, 2 across 3rd grade, and 4 across 4th grade. For grades 5 and 6 only one estimate was available; therefore, no aggregation was

necessary. These two sets of aggregated values were entered first into two fixed-effects models, then into two random-effects models in order to derive the weighted average effect sizes for h^2 , c^2 , and e^2 , first by project, and then by grade. The random-effects model accounts for variance within the effect sizes across studies rather than assume fixed variance across samples (Hedges 1983).

As a third and final step, in order to test for all levels of moderators across reported effect sizes, rather than aggregating, we included all possible etiological estimates of reading comprehension into subsequent analyses, leading to a total of 77 heritability estimates, and 76 shared and non-shared environmental estimates. Potential moderators for heritability and shared environment were assessed using a two-level, mixed-effects model which allows for population effect sizes to be predicted from between study variance from study characteristics (Borenstein et al. 2009). Non-shared environmental estimates also contain error, which may potentially confound moderator analyses which rely on parsing out variance due to true heterogeneity versus random error; therefore, we elected to exclude e^2 estimates from the moderator analyses.³

Potential moderators were examined using Q , I^2 , and T^2 statistics. The Q statistic and corresponding p value is an overall indicator of either the presence or absence of significant heterogeneity among effect sizes and I^2 represents the magnitude of heterogeneity present (Q_M ; Borenstein et al. 2009). The Q statistic can also test for the presence of residual heterogeneity not accounted for by moderators being tested (Q_E). The I^2 statistic represents the proportion of variance that is due to heterogeneity versus chance and ranges from 0 to 100 % with values closer to zero indicating variance is most likely due to random error and values closer to 100 indicating variance is more likely due to true heterogeneity (Higgins et al. 2003). The T^2 statistic represents the true variance from the observed studies (Borenstein et al. 2009).

In order to test for potential publication bias, a Rosenthal fail-safe N test was conducted for the averaged estimates of heritability and shared environmental influence. This analysis calculates the number of studies with null results that would be needed to raise Type 1 error to a significance level of $p < 0.05$ (Rosenthal 1979). Additionally, a funnel plot was used to also determine the level of potential publication bias. For funnel plots where standard error is plotted along the y-axis and the effect size is plotted along the x-axis, the resulting symmetry of the plotted points on either side of the mean can be used to

¹ Betjemann et al. 2011.

² e^2 is included in this formula as an example. Calculations were also conducted using a^2 and c^2 where appropriate.

³ One study (Betjemann et al. 2011) used a latent variable of reading comprehension. It is important to note that latent variables reduce the presence of measurement error, generally yielding lower non-shared environment and higher heritability estimates compared to the single measures.

evaluate the presence of publication bias (Cooper et al. 2009). Studies with larger sample sizes generate more precise estimates and usually appear at the top of the graph and those with smaller sample sizes generate less precise estimates and appear towards the bottom of the graph. More precise estimates should fall closer to the mean, resulting in a funnel-shaped plot. Analyses were run utilizing the metafor package in R statistical software (R Development Core Team 2011; Viechtbauer 2010).

Results

The aggregated etiological estimates were calculated in three steps, representing three levels of the data. First, effect sizes aggregated by individual project were analyzed. Starting with heritability estimates, a fixed-effects test of homogeneity was significant, indicating a random-effects analysis should be conducted to determine the nature of the variance between studies $Q_M(8) = 119.37$, $p < 0.01$. A follow-up, random effects model indicated an average heritability estimate of 0.54 (0.47–0.59), SE 0.04. Next, the same steps were conducted for environmental estimates revealing an average shared environmental estimate of 0.18 (0.12–0.24), SE 0.03; $Q_M(8) = 75.45$, $p < 0.01$ and an average non-shared environmental estimate of 0.30 (0.24–0.36), SE 0.03; $Q_M(8) = 73.04$, $p < 0.01$. These results suggested that approximately 54 % of individual differences in reading comprehension were due to heritability, 18 % due to shared environmental influences and 30 % due to non-shared environmental influences. Figures 1, 2 and 3 present forest plots of effect sizes by project.

Next, results from effect sizes aggregated by grade level indicated an average heritability estimate of 0.65 (0.56–0.73), SE 0.08; $Q_M(6) = 176.84$, $p < 0.01$, and an average shared environmental estimate of 0.14 (0.02–0.26), SE 0.06; $Q_M(6) = 133.53$, $p < 0.01$. For non-shared environmental estimates, a fixed-effects model was non-significant for the presence of heterogeneity $Q_M(6) = 7.10$, $p = 0.31$; therefore, the average weighted estimate was evaluated under a fixed-effects model $e^2 = 0.22$ (0.19–0.25), SE 0.02. Figures 4, 5 and 6 present forest plots of effect sizes by grade.

Finally, following analyses of aggregated estimates, all of the reported effect sizes from included studies were averaged and follow up moderator analyses on the heritability and shared environmental influences were conducted. The average heritability of reading comprehension using all available estimates under a random-effects model was $h^2 = 0.59$ (0.55–0.63), SE 0.03; $Q_M(76) = 5809.38$, $p < 0.01$. A Rosenthal fail-safe N test of publication bias determined that a total of 878,345 studies with null results

would be needed to nullify the average heritability estimate. Due to a large number of studies with large sample sizes, a funnel plot of heritability revealed a slightly non-funnel shape (Fig. 7), though the plotted effect sizes appeared symmetric on either sides of the mean, suggesting no publication bias.

Figure 8 displays a forest plot of heritability estimates by study and 95 % confidence intervals. The overall range of heritability estimates spanned from 0.14 to 0.84. In order to determine which study features contributed to heterogeneity among heritability estimates, studies were analyzed at the moderator level using a mixed model approach. Publication year, age, and grade were entered as continuous moderators into the model and all other moderators were entered as categorical. Table 3 displays the results of these analyses. Based on omnibus tests of heterogeneity, of the continuous moderators, grade and publication year were significant sources of heterogeneity between estimates of heritability and of the categorical moderators project, zygosity determination method, and response type were significant. For publication year, results indicated that each one unit increase in publication year corresponded to a 0.01 (95 % CI 0.005–0.02) increase in heritability and with grade, each one unit increase in grade level corresponded with a 0.07 (95 % CI 0.02–0.13) unit increase in heritability. Figures 9 and 10 illustrate the increases in heritability by year and grade, respectively. The test of residual heterogeneity was significant for year [$Q_E(75) = 5763.60$, $p < 0.0001$] and for grade [$Q_E(31) = 839.48$, $p < 0.0001$] indicating that other moderators not considered may be influencing estimates of heritability above each of these. The results of heritability estimates with a 95 % confidence interval, corresponding p values and standard errors by project, zygosity, and response type are included in Table 4. Estimates of

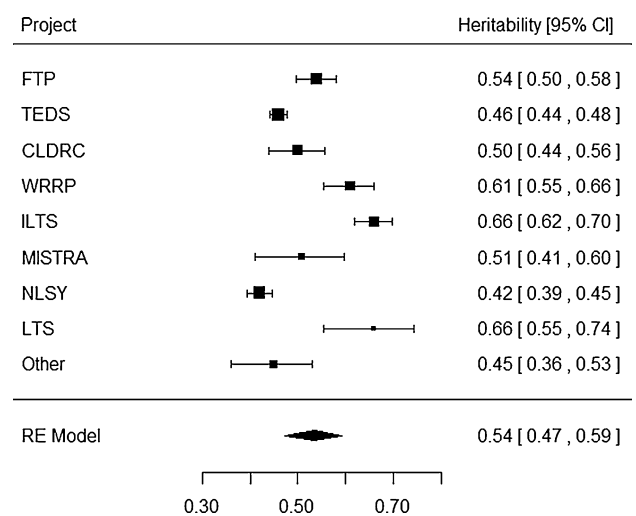


Fig. 1 Forest plot of heritability estimates aggregated by project

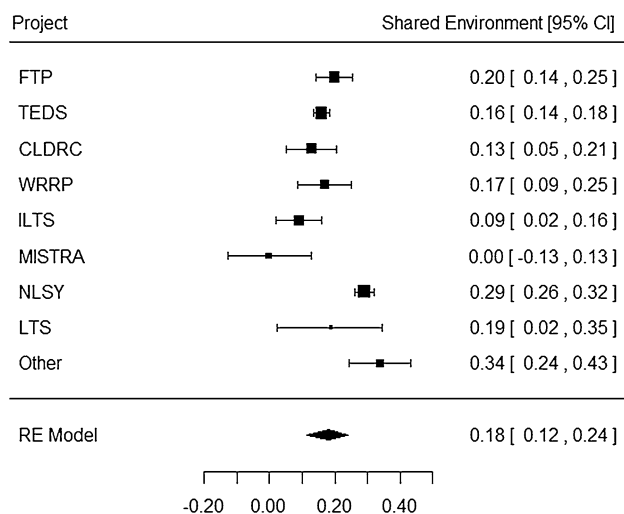


Fig. 2 Forest plot of shared environmental estimates aggregated by project

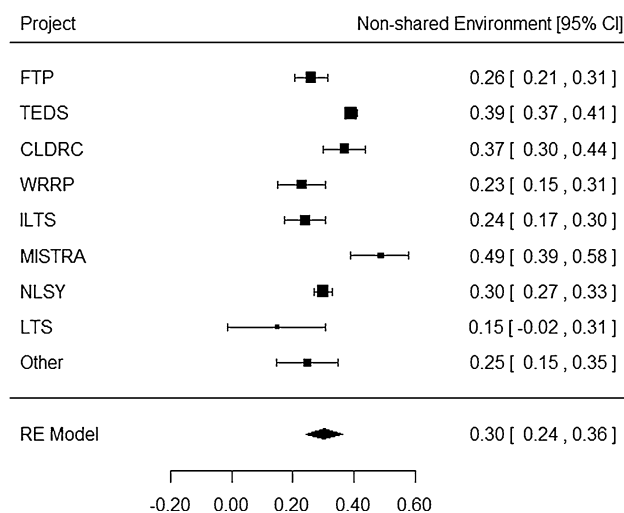


Fig. 3 Forest plot of non-shared environmental estimates aggregated by project

heritability under all moderators were significant with values ranging from 0.42 to 0.66.

Turning to the shared environmental estimates, a fixed-effects test of homogeneity was again significant, $Q_M(75) = 1405.84$, $p < 0.01$ and a follow-up random-effects model was run, indicating that the average estimate of shared environmental influences was $c^2 = 0.16$ (0.13–0.20), SE 0.02. Results of the Rosenthal fail-safe N test indicated that 60,354 studies with null results would be needed to nullify the average estimate. Figure 11 displays a funnel plot of shared environmental estimates. Estimates on either side of the mean were asymmetrical which suggests the presence of publication bias, with a slightly higher number of studies reporting estimates lower than the weighted average. Shared environmental estimates ranged

from 0.00 to 0.66. Figure 12 presents a forest plot of shared environmental estimates by study along with 95 % confidence intervals. Potential moderators for shared environmental estimates were tested using a mixed model approach with age, grade and year entered as continuous variables and the rest entered as categorical. Results revealed that publication year and grade were significant sources of heterogeneity from among continuous moderators (Figs. 13 and 14) with a 0.01 (95 % CI -0.01 to -0.002) decrease in shared environmental estimates for every 1 unit increase in publication year and a 0.06 (95 % CI -0.10 to -0.03) decrease in shared environmental estimates for every 1 unit increase in grade. Within the categorical moderators of shared environmental estimates, zygosity determination method was the sole moderator. Results from the omnibus tests of moderation are presented in Table 5 and the results of shared environmental estimates with a 95 % confidence interval, corresponding p values and standard errors by zygosity determination method are listed in Table 6.

Finally, non-shared environmental estimates were averaged under a random effects model [$Q_M(75) = 2831.33$, $p < 0.01$], with results indicating that the average estimate of non-shared environmental influences was $e^2 = 0.29$ (0.26–0.32), SE 0.02. A Rosenthal fail-safe N test indicated that 226,174 studies with null results would be needed to nullify the average estimate, and a funnel plot (Fig. 15) suggested some publication bias may be present in favor of estimates below the weighted average. Figure 16 displays a forest plot of non-shared environmental estimates. However, due to the presence of error within the non-shared environmental estimates, follow-up tests of moderation were not conducted.

Discussion

The purpose of this meta-analysis was to aggregate the genetic and environmental influences of reading comprehension from primary twin studies and to investigate whether potential sources of heterogeneity significantly moderate these influences. Results revealed that the average magnitude of heritability was large, $h^2 = 0.59$, with significant variation in estimates across studies. Furthermore, results indicated a small, yet significant average shared environmental contribution to reading comprehension, $c^2 = 0.16$, with less variability present across studies. A funnel plot of shared environmental estimates suggested the presence of some publication bias in favor of studies reporting lower values. The majority of the projects included in this synthesis were homogeneous (population $>75\%$) within categories of race, reading population, and potentially SES, which may serve to reduce overall

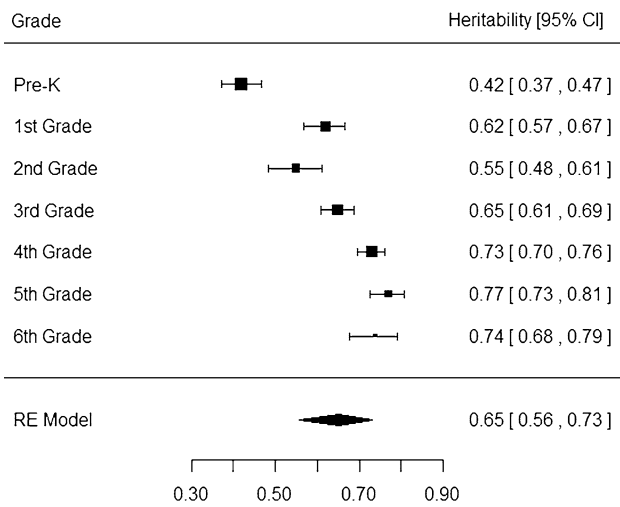


Fig. 4 Forest plot of heritability estimates aggregated by grade

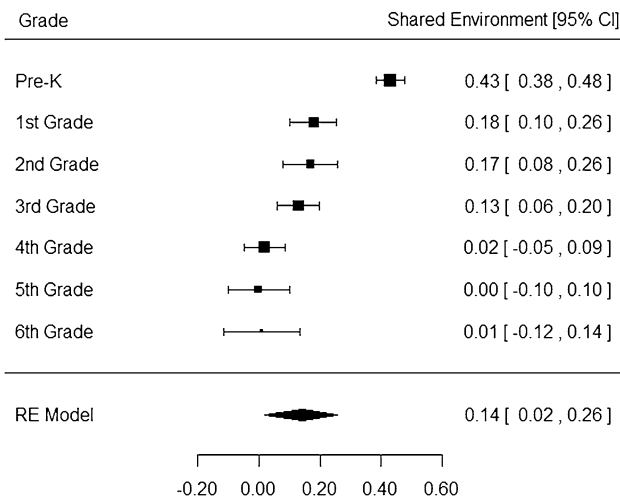


Fig. 5 Forest plot of shared environmental estimates aggregated by grade

variability within shared environmental conditions and result in lower published shared environmental estimates from these projects. The heritability result mirrors other meta-analyses of twin data, which have found an average of 50 % heritability across a wide range of physical, behavioral and cognitive traits (de Zeeuw et al. 2015; Polderman et al. 2015), and the significant shared environmental influence for reading comprehension mirrors de Zeeuw et al. (2015). Heritability estimates were found to be moderated by grade level, publication year, project, zygosity determination method, and response type. Grade, publication year, and zygosity determination method were significant moderators of shared environmental estimates, showing an inverse pattern to heritability.

The magnitude of genetic influences on reading comprehension and related skills including general cognitive ability has been found to increase across the lifespan

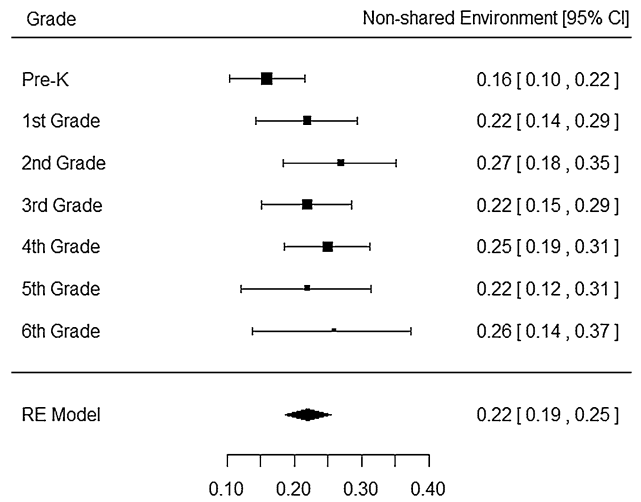


Fig. 6 Forest plot of non-shared environmental estimates aggregated by grade

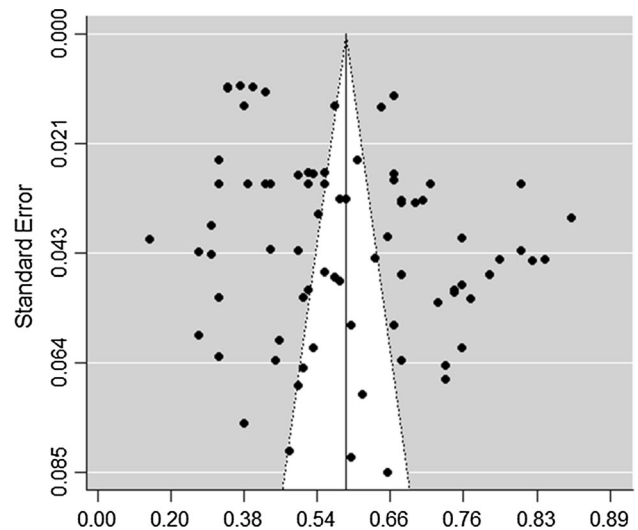


Fig. 7 Funnel plot of heritability estimates

(Byrne et al. 2009; Hart et al. 2013a, b; Haworth et al. 2010). This increase in heritability is commonly mirrored with a concurrent decrease in shared environmental influences. Importantly, the current results follow these previously established patterns of change, with heritability estimates of reading comprehension increasing with each grade-level increase and the shared environmental influences decreasing across the grades. This pattern, demonstrated across the large variety of samples and age ranges collected for the current meta-analysis, provides further evidence for the increasing role of genetics, and decreasing shared environment, in reading comprehension ability throughout childhood and adolescence. For educational practices, establishing this pattern convincingly has implications for the role of schools, classrooms and teachers. The increase in heritability may be indicating

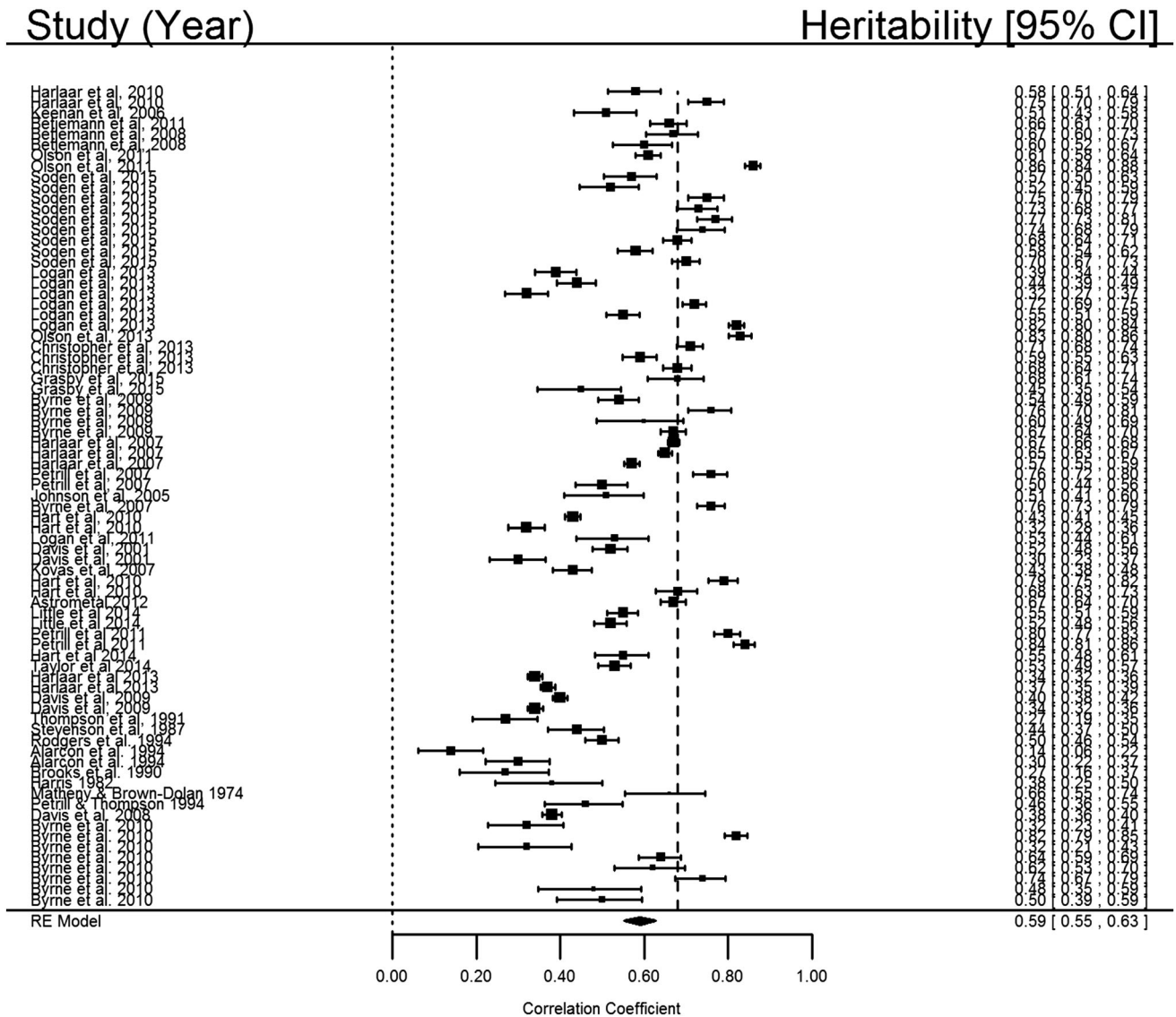


Fig. 8 Forest plot of heritability estimates by study

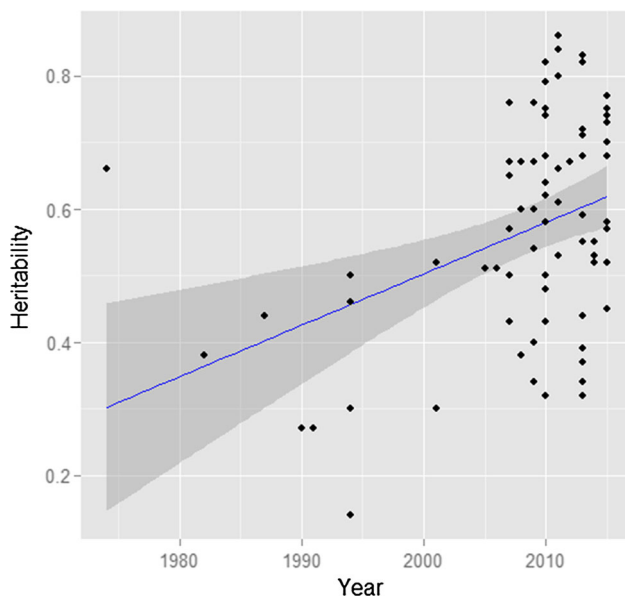
novel genetic influences are coming online as children age. A longitudinal examination of reading development found evidence of novel genetic influences for reading fluency across grades 1 through 3, suggesting new processes related to reading (i.e. additional component skills or general cognitive processes) may be activating or increasing their contribution to reading development as children age (Hart et al. 2013a, b). Additionally, several multivariate genetically sensitive studies have found overlapping genetic influences between several reading related skills (e.g. Byrne et al. 2005; Harlaar et al. 2007; Little and Hart 2016) and longitudinal studies have found both overlapping genetic influences between initial time points and across developmental time periods (e.g. Christopher et al. 2013; Logan et al. 2013), providing evidence for stability of genetic influences across skills and across development in

addition to innovative influences occurring at developmental stages. The accumulation of reading-related skills and their genetic influences across development coupled with potential new genetic influences on reading comprehension may present some explanation for the increasing role of heritability in older samples within the present results. Additionally, gene-environment correlations may be causing an artificial increase in the heritability estimate as children are increasingly surrounded by environments correlated with their reading comprehension skills. However, the presence of gene-environment correlation indicates further utility for interventions targeted to improving reading-related environmental factors (Olson et al. 2014). For example, interventions which focus on increasing exposure to reading activities or literacy-rich environments may have increased potential to improve reading ability for

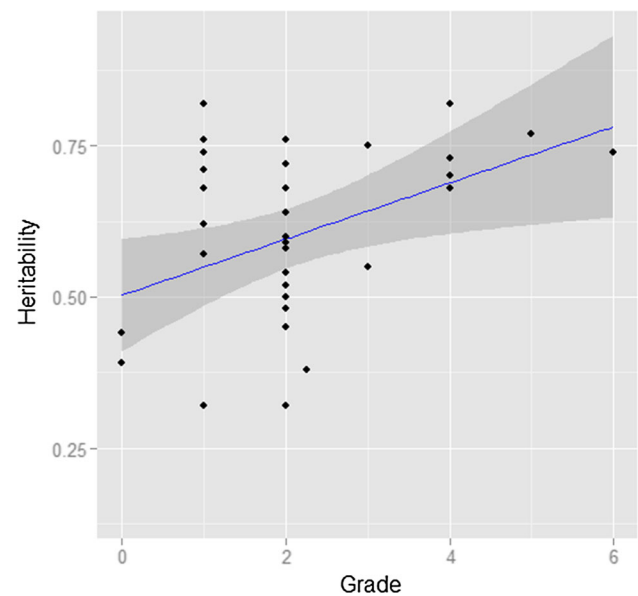
Table 3 Moderators of heritability estimates for reading comprehension

	<i>n</i>	$Q_M(df)$	<i>p</i>	T^2	I^2 (%)	R^2 (%)
Continuous predictors						
Year	77	10.34(1)	0.001	0.060	98.8	11.4
Age	56	0.69(1)	0.405	0.073	99.2	0
Grade	33	6.74(1)	0.009	0.046	96.4	16.15
Categorical predictors						
Project	77	17.78(8)	0.023	0.059	98.8	11.97
SES	32	3.44(2)	0.180	0.058	98.7	4.26
Racial composition	75	2.37(1)	0.124	0.067	99.0	1.90
Nationality	73	5.41(5)	0.368	0.068	98.9	0.85
Zygoty method	69	10.81(2)	0.005	0.062	98.8	11.67
Population	62	3.18(2)	0.204	0.073	98.3	2.06
Assessment type	76	0.18(1)	0.669	0.069	99.0	0.00
Response type	73	19.61(4)	<0.001	0.050	98.5	18.7

Significant ($p < 0.05$) moderators are indicated in bold. R^2 indicates the total amount of heterogeneity accounted for by the specified moderator

**Fig. 9** Increase in heritability by publication year

children with genetic predisposition for low reading ability and who may be more likely to avoid reading-related environments on their own. On the other side, the decrease in the shared environmental influence suggests that the environmental input on reading comprehension is stabilizing across multiple years of formalized education in reading. No matter the causal reason for this pattern, instructional approaches may need to become more individualized over time to account for these increasingly genetically-influenced individual differences. Individualized instructional practices have shown evidence of successfully improving student's reading comprehension skills longitudinally, such that students who received more years

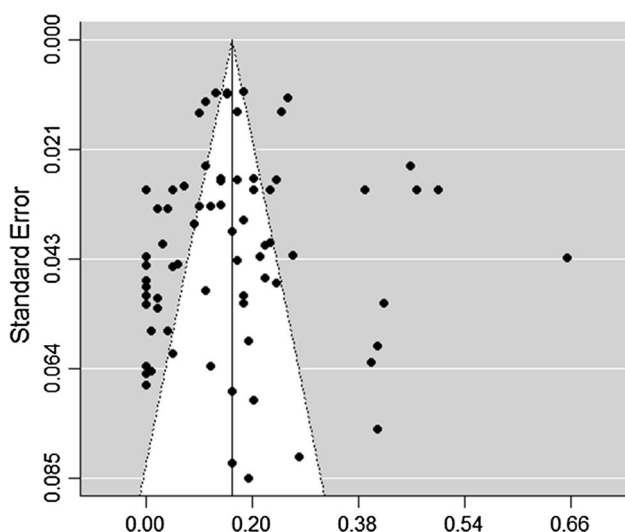
**Fig. 10** Increasing heritability by grade. Note: 0 = Kindergarten

of individualized instruction outperformed those who received individualized instruction at lower doses (Connor et al. 2013). This increased effectiveness may be related to a rise in the stability of instructional practices for students when these practices are catered to individual student needs instead of whole class needs. For example, classroom-level instructional practices may be influenced by different subsets of students across the school year, which may alter the overall pace of instruction (e.g. slowing down for struggling students or speeding through lessons to keep up with fast learners). Using individualized instructional practices can assist teachers and practitioners in maintaining stable, individually-paced instructional plans throughout the school year.

Table 4 Heritability estimates for reading comprehension by moderators

	h^2	SE	p	Lower CI	Upper CI
Project					
FTP	0.54	0.123	<0.001	0.35	0.69
TEDS	0.47	0.081	<0.001	0.34	0.59
CLDRC	0.53	0.075	<0.001	0.41	0.62
WRRMP	0.64	0.045	<0.001	0.58	0.69
ILTS	0.67	0.064	<0.001	0.60	0.74
MISTRA	0.51	0.252	0.025	0.07	0.79
NLSY	0.42	0.080	0.002	0.17	0.62
LTS	0.66	0.258	0.002	0.28	0.86
Other	0.45	0.101	<0.001	0.20	0.65
Zygosity					
Blood grouping	0.64	0.036	<0.001	0.59	0.68
Questionnaire	0.49	0.061	<0.001	0.39	0.57
Response type					
Cloze	0.64	0.036	<0.001	0.60	0.68
Multiple choice	0.48	0.068	<0.001	0.36	0.57
Rating scale	0.63	0.129	<0.001	0.45	0.76
Picture selection	0.49	0.053	<0.001	0.40	0.56
Short answer/retell	0.51	0.233	0.016	0.11	0.77

FTP Florida Twin Project, TEDS Twins Early Development Study, CLDRC Colorado Learning Disabilities Research Center Twin Study, WRRMP Western Reserve Reading and Math Project, ILTS International Longitudinal Twin Study, MISTRA Minnesota Study of Twins Reared Apart, NLSY National Longitudinal Survey of Youth, LTS Louisville Twin Study. Caucasian $\geq 75\%$ of the sample was reported as Caucasian. Blended no race was reported as making up 75 % or more of the sample

**Fig. 11** Funnel plot of shared environmental estimates

Results suggested heritability estimates increased slightly from earlier to later years of publication, and a reciprocal opposite effect was seen for the shared environment. Although publication year ranged from 1987 to 2015, the majority of studies were published after 2009, potentially skewing these results. A post hoc analysis revealed a skewness estimate of -2.16 for publication year and Fig. 3 demonstrates the presence of skewness. Several of the included projects (ILTS, FTP and WRRMP) showed an increase in publication rate after 2009, contributing to this skew and potentially to the moderation of heritability and the shared environment. Project was also a significant moderator for heritability, with ILTS, WRRMP and CLDRC reporting the highest average heritability estimates. Higher heritability estimates for ILTS and WRRMP, which contributed the majority of the studies published after 2009, may be partially responsible for the moderation of heritability by publication year. Beyond what is directly reported in our selected publications, twin projects may differ in several other aspects such as assessment administration method (e.g. home visits, online portals or mailed questionnaires), regional differences (e.g. urban or rural), or other unknown factors. The current meta-analysis explored several potential moderators that were nested within project, but results indicated additional, unmeasured moderators may be influencing heterogeneity of heritability estimates between projects, suggesting further exploration of project characteristics is warranted. Interestingly, shared-environmental estimates were not significantly moderated by project, indicating the potential sources of moderation for heritability estimates will not include those that influence the shared-environment. This pattern of results illustrates the need for further examination of between-project differences and suggests future areas of exploration should concentrate on agents that influence heritability, alone.

Nested within project was zygosity determination method which also served as a moderator for heritability estimates. Zygosity determination method was previously found to be a moderator of heritability of anti-social behavior (Rhee and Waldman 2002). Blood grouping methods such as saliva samples have been suggested to result in higher effect sizes (McCartney et al. 1990) and within the present study, studies using blood grouping methods reported overall higher heritability estimates than those with less stringent methods (e.g. questionnaire). Additionally, shared environmental estimates showed the inverse pattern, such that projects using questionnaire-based methods reported higher shared environmental estimates than those using blood-grouping methods from saliva samples. This difference suggests that projects using less stringent methods for zygosity determination may be

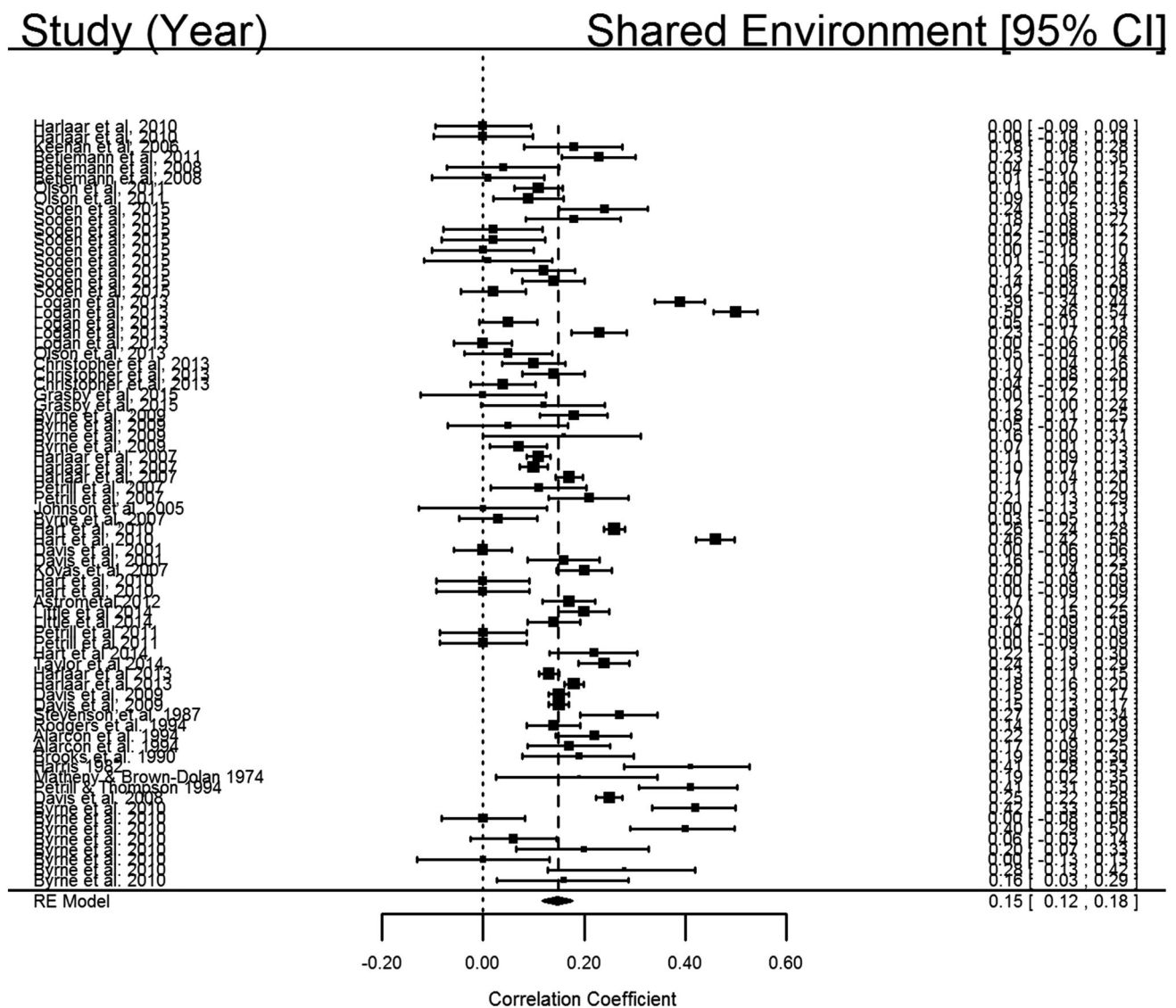


Fig. 12 Forest plot of shared environmental estimates

underestimating the level of genetic influences on reading comprehension, and a more in depth evaluation of zygosity determination methods is warranted to determine why this is the case. Reading population was not found to be a significant moderator of heritability or shared-environment. This supports a conceptualization of reading ability and disability falling along the same continuum, and as characterized by similar influences (Bishop 2015).

Response type was also a significant moderator of heritability. Measures of reading comprehension may be tapping into different underlying constructs such as decoding or reading-related skills such as executive functioning (Keenan et al. 2008). Studies within this meta-analysis assessed reading comprehension using several different response types. Heritability estimates were highest for cloze and teacher rating scales, followed by picture

selection, short answer/retell and multiple choice. Previous investigations have found that decoding relates more strongly to reading comprehension when reading comprehension is assessed with a cloze test (Francis et al. 2005), suggesting that cloze tests of reading comprehension tap into multiple skills and may be subject to greater sources of genetic influence. Genetic influences on decoding have been shown to range from 0.27 to 0.88 (e.g. Logan et al. 2013; Byrne et al. 2007; Olson et al. 2011; van Leeuwen et al. 2009) and up to 94 % of this influence has been found to overlap with reading comprehension skills (Naples et al. 2012), further suggesting that the higher heritability estimates for cloze type assessments may be partially due to the genetic influences of decoding ability. Teacher-based ratings of student academic achievement have been found to be fairly accurate in comparison to achievement tests, though subject to moderation based on the amount of

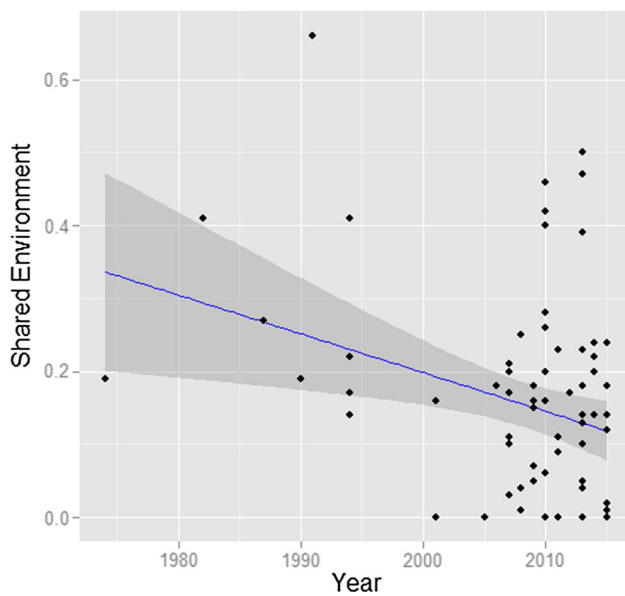


Fig. 13 Decreasing shared environmental influences by year

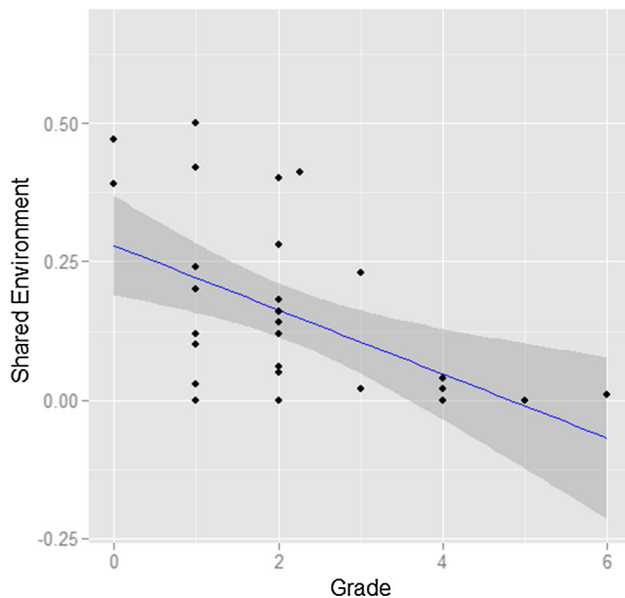


Fig. 14 Decreasing shared environmental influences by grade. Note: 0 = Kindergarten

information available and the type of achievement test used according to a recent meta-analysis (Südkamp et al. 2012). Within the present synthesis, only one teacher rating scale was included, which corresponded with only one study, and therefore may be subject to characteristics specific to that sample or rating scale (Harlaar et al. 2007). Additionally, survey-based measures such as rating scales are subject to effects from contrast effects or rater-bias which may serve to over or under estimate genetic influences (Nadder et al. 1998), but more information is needed to determine

whether these influences were present for the NC Rating scale used by the TEDS sample (Harlaar et al. 2007). Picture selection, multiple choice and short answer/retell resulted in the lowest average estimates of genetic influence. Selecting from a set of pictures or responses may not require an additional cognitive load to reading a passage and selecting the correct response from one's own mental lexicon, therefore, the underlying range of cognitive abilities used with these response types may be lower. Short answer/retell was only used within the MISTRA sample and no reliability information was available on the measure, as it was created for that study (Johnson et al. 2005). More investigation into this type of reading comprehension measure may be necessary to determine why recall resulted in lower estimates of heritability. Cloze, multiple choice and picture selection were used across several studies and samples indicating less chance of these being confounded with study or sample-specific variables.

Non-shared environmental influences on reading comprehension were found to moderate, with the largest aggregated estimates contributed by MISTRA, CLDRC and TEDS. Notably, these projects, along with ILTS, also reported the lowest levels of shared environmental estimates at the aggregated level. Specific, non-shared environmental influences on reading comprehension are difficult to disentangle from error in these models; however, it is possible that some child-specific factors such as peer influences may be present over and above error, and may merit further investigation (Asbury et al. 2008).

Three salient discrepancies arose within the results of this meta-analysis. Firstly, although grade was a significant moderator of heritability and shared environmental estimates, age was not. The majority of studies included utilized a large range of ages and did not divide the samples into age-bands, but rather reported the mean age from the entire sample. This resulted in similar reported means across multiple studies and reduced the amount of variability present for the age variable. However, several of the included studies did report grade-level specific estimates allowing for more overall variability for grade. Additionally, many twin studies included in this meta-analysis reported regressing the outcome variable by age in order to reduce potential age effects within the data, but this was not reported to be done for grade. The larger amount of variability present for grade, along with the presence of significant moderation suggests that developmental differences in both heritability and shared environmental influences may indeed be present, despite the lack of significant moderation by age.

Secondly, project and response type were significant moderators of heritability estimates, but not for shared environmental influences. The mathematical relatedness between genetic and shared environmental estimates

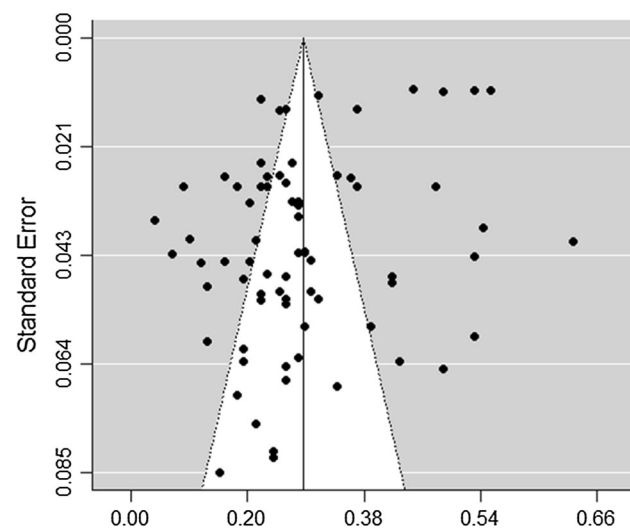
Table 5 Moderators of shared environmental estimates for reading comprehension

	<i>n</i>	$Q_M(df)$	<i>p</i>	T^2	I^2 (%)	R^2 (%)
Continuous predictors						
Year	76	7.08(1)	0.008	0.020	96.7	7.65
Age	56	1.81(1)	0.178	0.024	97.5	1.6
Grade	33	11.44(1)	<0.001	0.019	91.6	27.61
Categorical predictors						
Project	76	10.95(8)	0.204	0.021	96.7	4.07
SES	32	2.27(2)	0.322	0.023	97.1	0.00
Racial composition	74	2.32(1)	0.128	0.022	96.9	2.05
Nationality	72	2.40(5)	0.792	0.022	96.8	0.00
Zygosity method	69	8.56(2)	0.014	0.020	96.3	8.96
Population	61	0.51(2)	0.774	0.028	95.7	0.00
Assessment type	75	0.99(1)	0.320	0.022	97.0	0.00
Response type	72	7.19(4)	0.126	0.022	96.7	4.28

Significant ($p < 0.05$) moderators are indicated in bold. R^2 indicates the total amount of heterogeneity accounted for by the specified moderator

Table 6 Shared environmental estimates for reading comprehension by zygosity

Zygosity method	c^2	SE	<i>p</i>	Lower CI	Upper CI
Blood grouping	0.13	0.021	<0.001	0.09	0.17
Questionnaire	0.23	0.035	<0.001	0.16	0.29
Other	0.41	0.160	0.006	0.12	0.64

**Fig. 15** Funnel plot of non-shared environmental estimates

suggests that as one changes in magnitude so should the other, resulting in potential moderation effects for both. However, due to the selection of multiple studies from the same projects, there was fluctuation among significance and magnitude of shared environmental estimates within

projects that may have been equivalent to that across projects. Whereas, heritability estimates were consistently significant across all projects and moderators. Shared environmental estimates across projects ranged from 0.00 to 0.66, for FTP from 0.14 to 0.24, for CLDRC from 0.00 to 0.23 and for WRRMP from 0.00 to 0.66, for example. Tests of moderation examine between studies heterogeneity, but the within study heterogeneity present among the shared environmental estimates may prevent accurate estimation of true moderator effects within the present meta-analysis. Furthermore, funnel plots indicated the presence of publication bias for shared environmental estimates, but not estimates of heritability, suggesting gaps in published estimates of shared environmental estimates may be influencing the ability to accurately test moderators of these estimates.

Lastly, although SES was not found to be a significant moderator within the current meta-analysis, studies examining social disorganization theory have found evidence that poor access to resources in home, school and neighborhood environments may negatively influence outcomes such as achievement (Bowen et al. 2002; Bumgarner and Brooks-Gunn 2013). Furthermore, evidence from genetically sensitive studies has indicated etiological influences may be moderated by SES status under certain conditions (Hart et al. 2013a, b). The majority of studies included in the present study did not directly report SES; however, resulting in low power to detect moderation.

In addition to these limitations, the I^2 estimates are all above 80 % suggesting a high amount of true heterogeneity between studies, even in the presence of non-significant tests of moderation. Rucker et al. (2008) highlighted that these estimates are sensitive to sample size and level of sampling error such that as sample size increases I^2 values can increase to non-meaningful levels. Many of the studies included

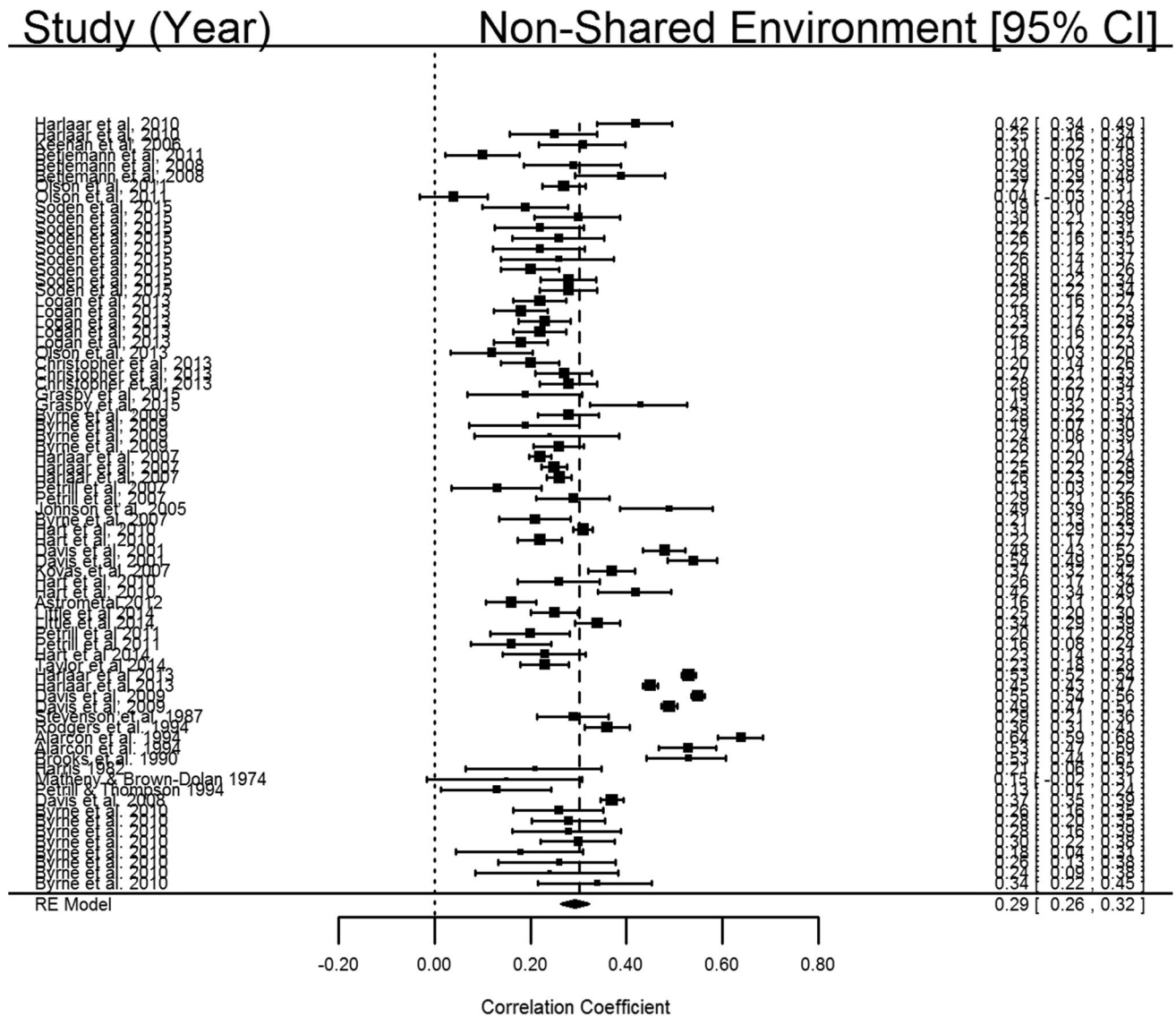


Fig. 16 Forest plot of non-shared environmental estimates

within the current meta-analysis utilized large sample sizes, suggesting the need to interpret I^2 values with caution.

This meta-analysis built on previous research examining the influences of nature and nurture of reading comprehension. The findings support the large role of genetic influences on reading comprehension and a small but significant role of shared environmental influences. Moreover, several aspects of sample and outcome were identified as having an impact on estimates of heritability and shared environment. Identification of these moderators and how they influence heritability has relevance to our interpretation of how genes and environment influence reading comprehension and is able to inform the design and implementation of future genetically sensitive studies. Additionally, the relative contributions of these influences

showed evidence of change across development which suggests implications for educational practice and policy such as individualized instruction. Future directions include examining the role of genes and environment across more clearly delineated levels of age and SES and identifying and testing additional moderators of these influences on reading comprehension.

Acknowledgments The first author was supported by Predoctoral Interdisciplinary Fellowship [funded by the Institute of Education Sciences, US Department of Education (186000-520-025833)].

Compliance with Ethical Standards

Conflict of Interest Author Callie Little declares she has no conflict of interest. Author Sara Hart declares she has no conflict of interest. Author Rasheda Haughbrook declares she has no conflict of interest.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Human and Animal Rights and Informed Consent This article does not contain any studies with animals performed by any of the authors. Informed consent was obtained from all individual participants included in the studies conducted by the authors within this meta-analysis. All other investigations included within this meta-analysis reported obtaining informed consent from individuals included within their analyses.

References

*References marked with an asterisk indicate studies included in the meta-analysis

- *Alarcon M, DeFries JC, Gillis JJ (1994) Familiar resemblance for measures of reading performance in families of reading-disabled and control twins. *Read Writ Interdiscip J* 6:93–101
- Alfassi M (2004) Reading to learn: effects of combined strategy instruction on high school students. *J Educ Res* 97(4):171–185
- Arnett AB, Pennington BF, Willcutt EG, DeFries JC, Olson RK (2015) Sex differences in ADHD symptom severity. *J Child Psychol Psychiatry* 56(6):632–639
- Arthur W, Bennett W, Huffcutt AI (2001) Conducting meta-analysis using SAS. Erlbaum, Mahwah, NJ
- Asbury K, Almeida D, Hibel J, Harlaar N, Plomin R (2008) Clones in the classroom: a daily diary study of the nonshared environmental relationship between monozygotic twin differences in school experience and achievement. *Twin Res Hum Genet* 11(06):586–595
- *Astrom RL, Wadsworth SJ, Olson RK, Willcutt EG, DeFries JC (2012) Genetic and environmental etiologies of reading difficulties: DeFries–Fulker analysis of reading performance data from twin pairs and their non-twin siblings. *Learn Individ Differ* 22(3):365–369
- Baker PC, Keck CK, Mott FL, Quinlan SV (1993) NLSY child handbook: a guide to the 1986–1990 National Longitudinal Survey of Youth child data. The Center for Human Resource Research, The Ohio State University, Columbus, OH
- Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K (2011) Low health literacy and health outcomes: an updated systematic review. *Ann Intern Med* 155(2):97–107
- *Betjemann RS, Keenan JM, Olson RK, DeFries JC (2011) Choice of reading comprehension test influences the outcomes of genetic analyses. *Sci Stud Read* 15(4):363–382
- *Betjemann RS, Willcutt EG, Olson RK, Keenan JM, DeFries JC, Wadsworth SJ (2008) Word reading and reading comprehension: stability, overlap and independence. *Read Writ* 21(5):539–558
- Bishop DVM (2015) The interface between genetics and psychology: lessons from developmental dyslexia. *Proc R Soc B* 282(1806):20143139
- Borenstein M, Hedges LV, Higgins JPT, Rothstein HR (2009) *Introduction to meta-analysis*. Wiley, West Sussex
- Bowen NK, Bowen GL, Ware WB (2002) Neighborhood social disorganization, families, and the educational behavior of adolescents. *J Adolesc Res* 17(5):468–490
- Bouchard TJ, Segal NL, Lykken DT (1990) Sources of human psychological differences: the Minnesota study of twins reared apart. *Science* 250:223–228
- Brighton CM, Moon TR, Huang FH (2015) Advanced readers in reading first classrooms who was really “Left Behind”? Considerations for the field of gifted education. *J Educ Gifted* 38(3):257–293
- *Brooks A, Fulker DW, DeFries JC (1990) Reading Performance and general cognitive ability: a multivariate genetic analysis of twin data. *Personal Individ Differ* 11(2):141–146
- Bumgarner E, Brooks-Gunn JEANNE (2013) Socioeconomic status and student achievement. *Int Guide Stud Achiev*:92–94
- *Byrne B, Coventry WL, Olson RK, Samuelsson S, Corley R, Willcutt EG, DeFries JC (2009) Genetic and environmental influences on aspects of literacy and language in early childhood: continuity and change from preschool to Grade 2. *J Neurolinguistics* 22(3):219–236
- *Byrne B, Coventry WL, Olson RK, Wadsworth SJ, Samuelsson S, Petrill SA, Willcutt EG, Corley R (2010) “Teacher effects” in early literacy development: Evidence from a study of twins. *J Educ Psychol* 102(1):32–42
- Byrne B, Delaland C, Fielding-Barnsley R, Quain P, Samuelsson S, Høien T, Willcutt E (2002) Longitudinal twin study of early reading development in three countries: preliminary results. *Ann Dyslexia* 52(1):47–73
- *Byrne B, Samuelsson S, Wadsworth S, Hulslander J, Corley R, DeFries JC, Olson RK (2007) Longitudinal twin study of early literacy development: Preschool through grade 1. *Read Writ* 20(1–2):77–102
- Byrne B, Wadsworth S, Corley R, Samuelsson S, Quain P, DeFries JC, Willcutt E, Olson RK (2005) Longitudinal twin study of early literacy development: preschool and kindergarten phases. *Sci Stud Read* 9(3):219–235
- Cain K, Oakhill J (2009) Reading comprehension development from 8 to 14 years: the contribution of component skills and processes. In: Wagner RK, Schatschneider C, Phythian-Sence C (eds) *Beyond decoding. The behavioral and biological foundations of reading comprehension*. Guilford Press, New York, pp 143–175
- Canney G (1989) Test review: metropolitan achievement tests (MAT6) reading diagnostic tests. *J Read*:148–150
- Center for Human Resource Research (2006). NLSY79 users guide. <ftp://www.nlsinfo.org/pub/usersvc/NLSY79/NLSY79%202004%20User%20Guide/79text/front.htm>. Accessed 11 Sept 2015
- Chall JS, Jacobs VA (2003) Poor children’s fourth-grade slump. *Am Educ* 27(1):14–17
- *Christopher ME, Hulslander J, Byrne B, Samuelsson S, Keenan JM, Pennington B, Olson RK (2013) Modeling the etiology of individual differences in early reading development: evidence for strong genetic influences. *Sci Stud Read* 17(5):350–368
- Connor CM, Morrison FJ, Fishman B, Crowe EC, 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. *Psychol Sci* 24(8):1408–1419
- Cooper H, Hedges LV, Valentine JC (eds) (2009) *The handbook of research synthesis and meta-analysis*. Russell Sage Foundation, New York
- Dale PS, Harlaar N, Plomin R (2005) Telephone testing and teacher assessment of reading skills in 7-year-olds: I. Substantial correspondence for a sample of 5544 children and for extremes. *Read Writ* 18(5):385–400
- *Davis CJ, Knopik VS, Olson RK, Wadsworth SJ, DeFries JC (2001) Genetic and environmental influences on rapid naming and reading ability: a twin study. *Ann Dyslexia* 51(1):231–247
- *Davis OSP, Kovas Y, Harlaar N, Busfield P, McMillan A, Frances J, Plomin R (2008) Generalist genes and the Internet generation:

- etiology of learning abilities by web testing at age 10. *Genes Brain Behav* 7(4):455–462
- *Davis OS, Haworth CM, Plomin R (2009) Learning abilities and disabilities: generalist genes in early adolescence. *Cogn Neuropsychiatry* 14(4–5):312–331
- Department for Education and Employment (2000) QCA key stage 1: assessment and reporting arrangements. HMSO, London
- de Zeeuw EL, de Geus EJ, Boomsma DI (2015) Meta-analysis of twin studies highlights the importance of genetic variation in primary school educational achievement. *Trends Neurosci Educ* 4(3):69–76. doi:10.1016/j.tine.2015.06.001
- DeFries J (1997) Colorado learning disabilities research center. *Learn Disabil Multidiscip J* 8(1):7–19
- Dunn LM, Markwardt FC (1970) Peabody individual achievement test. American Guidance Service, New York
- Falconer DS (1960) Introduction to quantitative genetics. Ronald Press, New York, NY
- Fletcher JM, Foorman BR, Boudousquie A, Barnes MA, Schatschneider C, Francis DJ (2002) Assessment of reading and learning disabilities a research-based intervention-oriented approach. *J Sch Psychol* 40(1):27–63
- Fletcher JM, Stuebing KK, Morris RD, Lyon GR (2013) Classification and definition of learning disabilities: a hybrid model. In Swanson HL, Harris K (eds) *Handbook of learning disabilities*, 2nd edn. Guilford Press, New York, NY, p 33–50
- Francis DJ, Fletcher JM, Catts HW, Tomblin JB (2005) Dimensions affecting the assessment of reading comprehension. In Stahl SA, Paris SG (eds) *Children's reading comprehension and assessment*, Erlbaum, Mahwah, NJ, p 369–394
- Francis DJ, Shaywitz SE, Stuebing KK, Shaywitz BA, Fletcher JM (1994) The measurement of change: assessing behavior over time and within a developmental context. In Lyon GR (ed) *Frames of reference for the assessment of learning disabilities: new views on measurement*, Brookes, Baltimore, p 29–58
- Friend A, DeFries JC, Olson RK (2008) Parental education moderates genetic influences on reading disability. *Psychol Sci* 19(11):1124–1130
- GOAL (2002) GOAL formative assessment: key stage 3. Hodder & Stoughton, London
- *Grasby KL, Byrne B, Olson RK (2015) Validity of large-scale reading tests: a phenotypic and behaviour—genetic analysis. *Austral J Educ* 59(1):5–21
- Hanushek EA, Woessmann L (2012) Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *J Econ Growth* 17(4):267–321
- *Harlaar N, Cutting L, Deater-Deckard K, DeThorne LS, Justice LM, Schatschneider C, Petrill SA (2010) Predicting individual differences in reading comprehension: a twin study. *Ann Dyslexia* 60(2):265–288
- *Harlaar N, Dale PS, Plomin R (2007) From learning to read to reading to learn: substantial and stable genetic influence. *Child Dev* 78(1):116–131
- *Harlaar N, Kovas Y, Dale PS, Petrill SA, Plomin R (2012) Mathematics is differentially related to reading comprehension and word decoding: evidence from a genetically sensitive design. *J Educ Psychol* 104(3):622
- *Harris EL (1982) Genetic and environmental influences on reading achievement: a study of first- and second-grade twin children. *Acta Genet Med Gemollol* 31:64–116
- Hart SA, Logan JA, Soden-Hensler B, Kershaw S, Taylor J, Schatschneider C (2013a) Exploring how nature and nurture affect the development of reading: an analysis of the Florida Twin Project on reading. *Dev Psychol* 49(10):1971
- Hart SA, Logan JA, Thompson LA, Kovas Y, McLoughlin G, Petrill S (2016) A latent profile analysis of math achievement, numerosity, and math anxiety in twins. *J Educ Psychol* 108(2):181–193
- *Hart SA, Mikolajewski AJ, Johnson W, Schatschneider C, Taylor J (2014) Examining transactional influences between reading achievement and antisocially-behaving friends. *Personal Individ Differ* 71:9–14
- *Hart SA, Petrill SA, Dush CMK (2010a) Genetic influences on language, reading, and mathematics skills in a national sample: an analysis using the National Longitudinal Survey of Youth. *Lang Speech Hear Serv Sch* 41:118–128. doi:10.1044/0161-1461
- *Hart SA, Petrill SA, Thompson LA (2010b) A factorial analysis of timed and untimed measures of mathematics and reading abilities in school aged twins. *Learn Individ Differ* 20(2):63–69
- Hart SA, Petrill SA, Thompson LA, Plomin R (2009) The ABCs of math: a genetic analysis of mathematics and its links with reading ability and general cognitive ability. *J Educ Psychol* 101(2):388–402
- Hart SA, Soden B, Johnson W, Schatschneider C, Taylor J (2013b) Expanding the environment: gene × school-level SES interaction on reading comprehension. *J Child Psychol Psychiatry*. doi:10.1111/jcpp.12083
- Haworth CM, Wright MJ, Luciano M, Martin NG, de Geus EJ, van Beijsterveldt CE, Bartels M, Posthuma D, Boomsma DI, Davis OS, Kovas Y, Corley RP, Defries JC, Hewitt JK, Olson RK, Rhea SA, Wadsworth SJ, Iacono WG, McGue M, Thompson LA, Hart SA, Petrill SA, Lubinski D, Plomin R (2010) The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Mol Psychiatry* 15(11):1112–1120
- Hedges LV (1983) A random effects model for effect sizes. *Psychol Bull* 93(2):388
- Hedges LV, Olkin I (1985) *Statistical methods for meta-analysis*. Academic Press, Orlando
- Higgins JP, Thompson SG, Deeks JJ, Altman DG (2003) Measuring inconsistency in meta-analyses. *Brit Med J* 327(7414):557
- Jerman O, Reynolds C, Swanson HL (2012) Does growth in working memory span or executive processes predict growth in reading and math in children with reading disabilities? *Learn Disabil Q* 35(3):144–157
- *Johnson W, Bouchard TJ, Segal NL, Samuels J (2005) General intelligence and reading performance in adults: Is the genetic factor structure the same as for children? *Personal Individ Differ* 38(6):1413–1428
- Kasriel J, Eaves L (1976) The zygosity of twins: further evidence on the agreement between diagnosis by blood groups and written questionnaires. *J Biosoci Sci* 8(03):263–266
- Keenan JM, Betjemann RS, Olson RK (2008) Reading comprehension tests vary in the skills they assess: differential dependence on decoding and oral comprehension. *Sci Stud Read* 12(3):281–300. doi:10.1080/10888430802132279
- *Keenan JM, Betjemann RS, Wadsworth SJ, DeFries JC, Olson RK (2006) Genetic and environmental influences on reading and listening comprehension. *J Res Read* 29(1):75–91
- Killewald A (2013) Return to being black, living in the red: a race gap in wealth that goes beyond social origins. *Demography* 50(4):1177–1195
- Kim YS, Petscher Y, Schatschneider C, Foorman B (2010) Does growth rate in oral reading fluency matter in predicting reading comprehension achievement? *J Educ Psychol* 102(3):652–667. doi:10.1037/a0019643
- *Kovas Y, Haworth CMA, Harlaar N, Petrill SA, Dale PS, Plomin R (2007) Overlap and specificity of genetic and environmental influences on mathematics and reading disability in 10-year-old twins. *J Child Psychol Psychiatry* 48(9):914–922
- Lipsey MW, Wilson DB (2001) *Practical meta-analysis*. Applied social research methods series, vol 49. Sage, Thousand Oaks

- Little CW, Hart SA (2016) Examining the genetic and environmental associations among spelling, reading fluency, reading comprehension and a high stakes reading test in a combined sample of third and fourth grade students. *Learn Individ Differ* 45:25–32
- *Little CW, Hart SA, Schatschneider C, Taylor J (2014) Examining associations among ADHD, homework behavior, and reading comprehension a twin study. *J Learn Disabil*:0022219414555715
- *Logan JA, Hart SA, Cutting L, Deater-Deckard K, Schatschneider C, Petrill S (2013) Reading development in young children: genetic and environmental influences. *Child Dev* 84(6):2131–2144
- *Logan J, Petrill SA, Flax J, Justice LM, Hou L, Bassett AS, Bartlett CW (2011) Genetic covariation underlying reading, language and related measures in a sample selected for specific language impairment. *Behav Genet* 41(5):651–659
- Markwardt FC (1989) Peabody individual achievement test-revised: PIAT-R. American Guidance Service, Circle Pines
- *Matheny AP, Brown-Dolan A (1974) A twin study of genetic influences in reading achievement. *J Learn Disabil* 7(2):43–46
- McCartney K, Harris MJ, Bernieri F (1990) Growing up and growing apart: a developmental meta-analysis of twin studies. *Psychol Bull* 107(2):226
- Nadder TS, Silberg JL, Eaves LJ, Maes HH, Meyer JM (1998) Genetic effects on ADHD symptomatology in 7-to 13-year-old twins: results from a telephone survey. *Behav Genet* 28(2):83–99
- Naples A, Katz L, Grigorenko EL (2012) Lexical decision as an endophenotype for reading comprehension: an exploration of an association. *Dev Psychopathol* 24(04):1345–1360
- National Center for Education Statistics (2013) The Nation's Report Card: a first look: 2013 mathematics and reading (NCES 2014-451). Institute of Education Sciences, US Department of Education, Washington, DC
- Neale MD (1966) Neale analysis of reading ability: manual of directions and norms. Macmillan Education, London
- Neale MD (1999) Neale analysis of reading ability: manual. ACER Press, Australian Council for Educational Research Limited, 19 Prospect Hill Road, Camberwell, Melbourne, Victoria 3124, Australia
- OECD (2015) PISA 2012 Results in Focus
- Oliver BR, Plomin R (2007) Twins' Early Development Study (TEDS): a multivariate, longitudinal genetic investigation of language, cognition and behavior problems from childhood through adolescence. *Twin Res Hum Genet* 10(1):96–105
- *Olson RK, Hulslander J, Christopher M, Keenan JM, Wadsworth SJ, Willcutt EG, DeFries JC (2013) Genetic and environmental influences on writing and their relations to language and reading. *Ann Dyslexia* 63(1):25–43
- Olson RK, Keenan JM, Byrne B, Samuelsson S (2014) Why do children differ in their development of reading and related skills? *Sci Stud Read* 18(1):38–54
- *Olson RK, Keenan JM, Byrne B, Samuelsson S, Coventry WL, Corley R, Hulslander J (2011) Genetic and environmental influences on vocabulary and reading development. *Sci Stud Read* 15(1):26–46
- *Petrill S, Logan J, Hart S, Vincent P, Thompson L, Kovas Y, Plomin R (2011) Math fluency is etiologically distinct from untimed math performance, decoding fluency, and untimed reading performance: Evidence from a twin study. *J Learn Disabil*:0022219411407926
- *Petrill SA, Deater-Deckard K, Thompson LA, Schatschneider C, DeThorne LS, Vandenberg DJ (2007) Longitudinal genetic analysis of early reading: the Western Reserve reading project. *Read Writ* 20(1–2):127–146
- Petrill SA, Deater-Deckard K, Thompson LA, DeThorne LS, Schatschneider C (2006) Reading skills in early readers: genetic and shared environmental influences. *J Learn Disabil* 39(1):48–55
- *Petrill SA, Thompson LA (1994). The effect of gender upon heritability and common environmental estimates in measures of scholastic achievement. *Personal Individ Differ* 16(4):631–640
- Petscher Y, Kim YS (2011) The utility and accuracy of oral reading fluency score types in predicting reading comprehension. *J Sch Psychol* 49(1):107–129. doi:10.1016/j.jsp.2010.09.004
- Plomin R, DeFries JC, Knopik VS, Neiderhiser J (2013) Behavioral genetics. Palgrave Macmillan, Basingstoke
- Polderman TJ, Benyamin B, de Leeuw CA, Sullivan PF, van Bochoven A, Visscher PM, Posthuma D (2015) Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet*
- Price TS, Freeman B, Craig I, Petrill SA, Ebersole L, Plomin R (2000) Infant zygosity can be assigned by parental report questionnaire data. *Twin Res* 3(03):129–133
- Rashid FL, Morris RD, Sevcik RA (2005) Relationship between home literacy environment and reading achievement in children with reading disabilities. *J learn disabil* 38(1):2–11
- Reading performance (PISA) (indicator). doi:10.1787/79913c69-en. Accessed 04 Sept 2015
- Rhea SA (2015) Reviving the Louisville twin study: an introduction. *Behav Genet* 45(6):597–599
- Rhee SH, Waldman ID (2002) Genetic and environmental influences on antisocial behavior: a meta-analysis of twin and adoption studies. *Psychol Bull* 128(3):490
- Roberts G, Good R, Corcoran S (2005) Story retell: a fluency-based indicator of reading comprehension. *Sch Psychol Q* 20(3):304–317. doi:10.1521/scpq.2005.20.3.304
- *Rodgers JL, Rowe DC, May K (1994) DF analysis of NLSY IQ/achievement data: Nonshared environmental influences. *Intelligence* 19(2):157–177
- Rodgers JL (1996) NLSY linking algorithm. Unpublished document. Department of Psychology, University of Oklahoma
- Rosenthal R (1979) The file drawer problem and tolerance for null results. *Psychol Bull* 86(3):638
- Rücker G, Schwarzer G, Carpenter JR, Schumacher M (2008) Undue reliance on I2 in assessing heterogeneity may mislead. *BMC Med Res Methodol* 8(1):79
- R Development Core Team (2011) R: a language and environment for statistical computing [Computer software]. R Foundational for Statistical Computing, Vienna. <http://www.R-project.org>
- *Soden B, Christopher ME, Hulslander J, Olson RK, Cutting L, Keenan JM, Thompson LA, Wadsworth SJ, Willcutt EG, Petrill SA (2015) Longitudinal stability in reading comprehension is largely heritable from grades 1 to 6. *PLoS One* 10(1):e0113807
- Spencer M, Wagner RK, Schatschneider C, Quinn JM, Lopez D, Petscher Y (2014) Incorporating RTI in a hybrid model of reading disability. *Learn Disabil Q* 37(3):161–171. doi:10.1177/0731948714530967
- *Stevenson J, Graham P, Fredman G, McLoughli V (1987) A twin study of genetic influences on reading and spelling ability and disability. *J Child Psychol Psychiatry* 28(2):229–247
- Südkamp A, Kaiser J, Möller J (2012) Accuracy of teachers' judgments of students' academic achievement: a meta-analysis. *J Educ Psychol* 104(3):743
- *Taylor J, Hart SA (2014) A chaotic home environment accounts for the association between respect for rules disposition and reading comprehension: a twin study. *Learn Individ Differ* 35:70–77
- Taylor JE, Hart SA, Mikolajewski AJ, Schatschneider C (2013) An update on the Florida State Twin Registry. *Twin Res Hum Genet* 16(Special Issue 01):471–475. doi:10.1017/thg.2012.74
- *Thompson LA, Dettmerman DK, Plomin R (1991) Associations between cognitive abilities and scholastic achievement: genetic overlap but environmental differences. *Psychol Sci* 2(3):158–165
- Tiegs E, Clark W (1977) California Achievement Test. CTB, Monterey

- Trouton A, Spinath FM, Plomin R (2002) Twins early development study (TEDS): a multivariate, longitudinal genetic investigation of language, cognition and behavior problems in childhood. *Twin Res* 5(05):444–448
- Tucker-Drob EM, Bates TC (2015) Large cross-national differences in gene \times socioeconomic status interaction on intelligence. *Psychol Sci*:0956797615612727
- Turkheimer E, Haley A, Waldron M, D'Onofrio B, Gottesman II (2003) Socioeconomic status modifies heritability of IQ in young children. *Psychol Sci* 14(6):623–628
- UNESCO (2009) Investing in cultural diversity and intercultural dialogue. Executive summary. UNESCO world report. UNESCO, Paris
- Vandenberg SG, Stafford RE, Brown AM (1968) The Louisville Twin Study. In: Vandenberg SG (ed) *Progress in human behavior genetics*. Johns Hopkins Press, Baltimore, pp 153–204
- van Leeuwen M, van den Berg SM, Peper JS, Pol HEH, Boomsma DI (2009) Genetic covariance structure of reading, intelligence and memory in children. *Behav Genet* 39(3):245–254
- Viechtbauer W (2010) Conducting meta-analyses in R with the metafor package. *J Stat Softw* 36(3):1–48
- Woodcock RW (1987) *Woodcock reading mastery tests-revised*. American Guidance Service, Circle Pines