Title: Disentangling genetically confounded polygenic associations between Attention-Deficit/Hyperactivity Disorder, literacy and language

Short title: Disentangling genetically confounded polygenic associations

Key words: ADHD, literacy, language, EA, ALSPAC, polygenic overlap, genetic confounding

Ellen Verhoef[1,2*], Ditte Demontis[3-5*], Stephen Burgess[6,7], Chin Yang Shapland[1], Philip S. Dale[8], Aysu Okbay[9,10], Benjamin M. Neale[11-13], Stephen V. Faraone[14], iPSYCH-Broad-PGC ADHD Consortium, Evie Stergiakouli[15,16], George Davey Smith[15,17], Simon E. Fisher[1,18], Anders D. Børglum[3-5#], Beate St Pourcain[1,15,18#]

1. Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
2. International Max Planck Research School for Language Sciences, Nijmegen, The Netherlands
3. The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Denmark
4. Centre for Integrative Sequencing, iSEQ, Aarhus University, Aarhus, Denmark
5. Department of Biomedicine - Human Genetics, Aarhus University, Denmark
6. MRC Biostatistics Unit, University of Cambridge, Cambridge, UK
7. Cardiovascular Epidemiology Unit, Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK
8. Speech & Hearing Sciences, University of New Mexico, Albuquerque, USA
9. Department of Complex Trait Genetics, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands

10. Erasmus University Rotterdam Institute for Behavior and Biology, Rotterdam, The Netherlands
11. Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Massachusetts, USA
12. Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA
13. Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA
14. Departments of Psychiatry and Neuroscience and Physiology, SUNY Upstate Medical University, New York, USA
15. MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK
16. School of Oral and Dental Sciences, University of Bristol, Bristol, UK
17. School of Social and Community Medicine, University of Bristol, Bristol, UK
18. Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands

*,# These authors contributed equally to the presented work

Corresponding author: Ellen Verhoef and Beate St Pourcain

Postal address: Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, the Netherlands

e-mail: Ellen.Verhoef@mpi.nl / Phone: +31 24 3521946

e-mail: Beate.StPourcain@mpi.nl / Phone: +31 24 3521964 / Fax: +31 24 3521213

Requests for reprints should be sent to the corresponding author.

Abstract

Interpreting polygenic overlap between ADHD and both literacy- and language-related impairments is challenging as genetic confounding can bias associations. Here, we investigate evidence for links between polygenic ADHD risk and multiple literacy- and language-related abilities (LRAs), assessed in UK children (N≤5,919), conditional on genetic effects shared with educational attainment (EA). Genome-wide summary statistics on clinical ADHD and years-of-schooling were obtained from large consortia (N≤326,041). ADHD-polygenic scores (ADHD-PGS) were inversely associated with LRAs in ALSPAC, most consistently with reading-related abilities, and explained ≤1.6% phenotypic variation. Polygenic links were then dissected into both genetic effects shared with and independent of EA using multivariable regressions (MVR), analogous to Mendelian Randomization approaches accounting for mediating effects. Conditional on EA, polygenic ADHD risk remained associated with multiple literacy-related skills, phonemic awareness and verbal intelligence, but not language-related skills such as listening comprehension and non-word repetition. Pooled reading performance showed the strongest overlap with ADHD independent of EA. Using conservative ADHD-instruments (P-threshold $<5\times10^{-8}$) this corresponded to a 0.35 decrease in Z-scores per log-odds in ADHD-liability ($P=9.2\times10^{-5}$). Using subthreshold ADHD-instruments (P-threshold <0.0015), these associations had lower magnitude, but higher predictive accuracy, with a 0.03 decrease in Z-scores ($P=1.4\times10^{-6}$). Polygenic ADHD-effects shared with EA were of equal strength and at least equal magnitude compared to those independent of EA, for all LRAs studied, and only detectable using subthreshold instruments. Thus, ADHD-related polygenic links are highly susceptible to genetic confounding, concealing an ADHD-specific association profile that primarily involves reading-related impairments, but few language-related problems.

Introduction

Children with Attention-Deficit/Hyperactivity Disorder (ADHD) often experience difficulties mastering literacy- and language-related abilities (LRAs)[1–3]. It has been estimated that up to 40% of children diagnosed with clinical ADHD also suffer from reading disability (RD, also known as developmental dyslexia) and vice versa[4]. The spectrum of affected LRAs in ADHD may, however, also include writing[5,6], spelling[7,8], syntactic[9,10] and phonological[9,10] abilities. Both clinical ADHD and RD are complex childhood-onset neurodevelopmental conditions that affect about 5% and 7% of the general population respectively[11,12]. ADHD is characterised by hyperactive, inattentive and impulsive symptoms[13], whereas decoding and/or reading comprehension deficits are prominent in individuals with RD[14].

To interpret the comorbidity of ADHD and RD a multiple-deficit model including shared underlying aetiologies has been proposed, involving both genetic and environmental influences[15]. This model is supported by twin studies suggesting that the co-occurrence of ADHD symptoms and reading deficits is to a large extent attributable to shared genetic influences[16–18]. However, the interpretation of polygenic overlap is challenging as associations might be inflated or induced by genetic confounders that are genetically related to both ADHD and reading abilities. Such a potential confounder of polygenic ADHD links is genetically predicted educational attainment (EA), a proxy of cognitive abilities and socioeconomic status[19]. Previous research showed evidence for a positive genetic relation between EA and reading abilities[20], as well as a moderate negative genetic correlation between EA and ADHD[21]. Therefore, genetic effects shared with EA might obscure the genetic relationship between ADHD and LRAs. Both the extent and nature of the genetic overlap between ADHD and LRAs, accounting for genetic confounding through EA, are unknown.

Here, we (1) study polygenic links between clinical ADHD and a wide range of population-ascertained literacy- and language-related measures, (2) evaluate whether such links reflect a shared genetic basis with EA and (3) assess whether there is support for shared genetic factors between clinical ADHD and LRAs conditional on genetically predicted EA. ADHD polygenic scores (ADHD-PGS) are generated based on ADHD genome-wide association study (GWAS) summary statistics of two large independent ADHD samples, the Psychiatric Genomics Consortium (PGC) and the Danish Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH) and a combination thereof. Associations between ADHD-PGS and a spectrum of population-based literacy- and language-related measures related to reading, spelling, phonemic awareness, listening comprehension, non-word repetition and verbal intelligence skills, are examined in a sample of children from the UK Avon Longitudinal Study of Parents and Children (ALSPAC). Applying multivariable regression (MVR) techniques, analogous to Mendelian Randomization (MR) approaches accounting for mediating effects[22], we disentangled associations between polygenic ADHD risk and LRA measures into effects independent of and shared with years-of-schooling, using summary statistics from the Social Science Genetic Association Consortium (SSGAC).

Methods and Materials

Literacy- and language-related abilities in the general population

LRAs were assessed in children and adolescents from ALSPAC, a UK population-based longitudinal pregnancy-ascertained birth cohort (estimated birth date: 1991-1992, Supplementary Information)[23,24]. Ethical approval was obtained from the ALSPAC Law-and-Ethics Committee (IRB00003312) and the Local Research-Ethics Committees. Written

informed consent was obtained from a parent or individual with parental responsibility and assent (and for older children consent) was obtained from the child participants.

Phenotype information: Thirteen measures capturing LRAs related to reading, spelling, phonemic awareness, listening comprehension, non-word repetition and verbal intelligence scores were assessed in 7 to 13 year-old ALSPAC participants (N≤5,919, Table 1) using both standardised and ALSPAC-specific instruments (Supplementary Information). Reading measures included comprehension and decoding (based on words and non-words) scores. Detailed descriptions of all LRA measures are available in Table 1 and the Supplementary Information.

All LRA scores were rank-transformed to allow for comparisons of genetic effects across different psychological instruments with different distributions (Supplementary Information). Phenotypic correlations, using Pearson-correlation coefficients, were comparable for untransformed and rank-transformed scores (Table S1). To account for multiple testing, we estimated the effective number of phenotypes studied using Matrix Spectral Decomposition[25] (MatSpD), revealing seven independent measures (experiment-wide error rate of 0.007).

For sensitivity analysis, we excluded 188 children with LRA data available and an ADHD diagnosis at age 7, based on the Development and Wellbeing Assessment (DAWBA)[26] (Supplementary Information).

Genetic analyses: ALSPAC participants were genotyped using the Illumina HumanHap550 quad chip genotyping platforms, and genotypes were called using the Illumina GenomeStudio software. Genotyping, imputation and genome-wide association analysis details are described in the Supplementary Information and Table 2.

Table 1 here

Clinical ADHD summary statistics

Psychiatric Genomics Consortium (PGC). GWAS summary statistics were obtained from a mega-analysis of clinical ADHD[27], conducted by the PGC (4,163 cases and 12,040 controls/pseudo-controls) (Table 2, Supplementary Information, www.med.unc.edu/pgc/).

The Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH). An independent set of ADHD GWAS summary statistics were accessed through the Danish iPSYCH project[28] (14,584 ADHD cases, 22,492 controls) (Table 2, Supplementary Information), using samples from the Danish Neonatal Screening Biobank hosted by Statens Serum Institute[21,29].

Combined PGC and iPSYCH ADHD sample (PGC+iPSYCH). To maximise power, we also analysed meta-GWAS summary statistics from an ADHD sample containing both PGC and iPSYCH participants[21] (20,183 cases, 35,191 controls/pseudo-controls) (Table 2, www.med.unc.edu/pgc/) and its European-only subset (*PGC+iPSYCH(EUR),* 19,099 cases, 34,194 controls/pseudo-controls) (Table 2, www.med.unc.edu/pgc/).

Detailed sample descriptions are available in Table 2 and the Supplementary Information.

Educational attainment summary statistics

GWAS summary statistics for EA[19] (discovery and replication sample combined, excluding ALSPAC and 23andMe samples, N=326,041) were obtained from the SSGAC

consortium. EA was assessed as years of schooling[19]. A detailed sample description is available in Table 2 and the Supplementary Information.

Table 2 here

Genome-wide complex trait analysis

SNP-h^2 and genetic correlations (r_g) between LRAs were estimated using Restricted Maximum Likelihood (REML) analyses[30,31] as implemented in Genome-wide Complex Trait Analysis (GCTA) software[32], including individuals with a genetic relationship <0.05[30]. For this study, we selected only LRAs with evidence for SNP-h^2 and sample size N>4,000 (Table S2).

Linkage Disequilibrium Score regression and correlation

Linkage Disequilibrium Score (LDSC) regression[33] was used to distinguish confounding biases from polygenic influences by examining the LDSC regression intercept. Unconstrained LD-score correlation[34] analysis was applied to estimate r_g (Supplementary Information).

Polygenic scoring analyses

ADHD-PGS[35,36] were created in ALSPAC using PGC, iPSYCH and PGC+iPSYCH GWAS summary statistics (Supplementary Information). ADHD-PGS have been previously linked to ADHD symptoms in ALSPAC participants[37]. Rank-transformed LRAs were regressed on Z-standardised ADHD-PGS (aligned to measure risk-increasing alleles) using ordinary least square (OLS) regression (R:stats library, Rv3.2.0). The proportion of phenotypic variance

explained is reported as OLS-regression-R^2. Beta-coefficients (β) for ADHD-PGS quantify here the change in standard deviation (SD) units of LRA performance per one SD increase in ADHD-PGS.

Multivariable regression analysis

To study the genetic association between ADHD and LRAs conditional on genetic influences shared with EA, we applied MVR. This technique is analogous to MR methodologies accounting for mediating effects[22] and controls for collider bias[38] through the use of GWAS summary statistics. Technically, it involves the regression of regression estimates from independent samples on each other[22] (Supplementary Information). Within this study we use MVR without inferring causality due to violations of classical MR assumptions[22] (see below).

Genetic variant selection: To disentangle ADHD-LRA associations, we selected two sets of genetic variants from the most powerful ADHD GWAS summary statistics (PGC+iPSYCH). The first set contained genome-wide significant variants ($P<5x10^{-8}$, conservative). The second set included variants passing a more lenient *P*-value threshold ($P<0.0015$, subthreshold) to increase power, consistent with current guidelines for the selection of genetic instruments in MR (F-statistic<10)[39]. All sets included independent (PLINK[40] clumping: LD-$r^2<0.25$, ±500 kb), well imputed (INFO[41]>0.8) and common (EAF>0.01) variants. This resulted in 15 conservative and $2{,}689<N_{SNPs}\leq 2{,}692$ subthreshold ADHD-instruments (Table S8).

Estimation of ADHD effects: We extracted regression estimates for selected ADHD-instruments (conservative and subthreshold) from ADHD (PGC+iPSYCH), EA (SSGAC) and LRA (ALSPAC) GWAS summary statistics. Analysing each set of variants independently,

regression estimates for individual LRA measures (β) were regressed on both ADHD (β as lnOR) and EA regression estimates (β) using an OLS regression framework (R:stats library, Rv3.2.0). Outcomes were 1) a MVR regression estimate quantifying the change in SD units of LRA performance per log odds increase in ADHD risk conditional on years of schooling (ADHD effect independent of EA), and 2) a MVR regression estimate quantifying the change in SD units of LRA performance per year of schooling as captured by ADHD instruments (ADHD effect shared with EA). Latter MVR regression estimates capture here genetic confounding, including confounding influences in a narrow sense (i.e. genetically predictable EA causally influences both ADHD and LRA), but also mediating effects (i.e. genetically predictable ADHD causally influences LRA indirectly through EA) and biological pleiotropy (i.e. ADHD risk variants affect ADHD and EA through independent biological pathways). As ADHD risk and EA are inversely genetically related with each other[21], they were converted to quantify a change per missing year of schooling. To compare the magnitude of both MVR estimates, we also conducted analyses using fully standardised EA, ADHD and LRA regression estimates (Supplementary Information).

Finally, MVR regression estimates were meta-analysed across reading-related, spelling-related and all LRA measures (Table 1) using random-effects meta-regression, accounting for phenotypic correlations between LRAs (R:metafor library, Rv3.2.0, Supplementary Information).

Sensitivity analyses: As the directionality of effects cannot be inferred in this study, we examined, in reverse, evidence for ADHD as a genetic confounder of the association between EA and LRAs, using MVR. Two sets of EA instruments (conservative and subthreshold, Table S8) were selected from EA (SSGAC) GWAS summary statistics, analogous to the selection of ADHD instruments, and MVR was conducted as described above.

Attrition analysis

We carried out an *attrition* analysis in ALSPAC studying the genetic association between LRA-missingness and ADHD-PGS (Supplementary Information).

Results

Genetic architecture of literacy- and language-related abilities and clinical ADHD

Phenotypic variation in literacy- and language-related measures (Table 1), including reading abilities (comprehension, accuracy and speed) assessed in words/passages and non-words, spelling abilities (accuracy), phonemic awareness, listening comprehension, non-word repetition and verbal intelligence scores, can be tagged by common variants, with SNP-h^2 estimates between 0.32 (SE=0.07, non-word repetition age 8) and 0.54 (SE=0.07, verbal intelligence age 8) (Table S2; GCTA- and LDSC-based estimations). All LRAs are phenotypically (Table S1) and genetically (Table S3) moderately to strongly interrelated. The observed LDSC-based evidence for genetic liability of clinical ADHD within the PGC (LDSC-h^2=0.08(SE=0.03)) ,iPSYCH (LDSC-h^2=0.26(SE=0.02)) and PGC+iPSYCH samples (Table S4) is consistent with previous reports[21].

Association between ADHD polygenic risk scores and literacy- and language-related abilities

We observed robust evidence for an inverse genetic association between ADHD-PGS and reading accuracy/comprehension age 7 (PGC: OLS-R^2=0.1%, P=4.6x10^{-3}; iPSYCH: OLS-R^2=1.0%, P<1x10^{-10}), reading accuracy age 9 (PGC: OLS-R^2=0.1%, P=5.7x10^{-3}; iPSYCH: OLS-R^2=1.2%, P<1x10^{-10}), and spelling accuracy age 9 (PGC: OLS-R^2=0.2%, P=1.5x10^{-3}; iPSYCH:

OLS-R^2=0.8%, $P<1x10^{-10}$) using independent ADHD discovery samples (Figure 1, Table S5). Polygenic scoring results are presented for a *P*-value threshold of 0.1, but other thresholds provided similar results (data not shown). The strongest evidence for association was observed when ADHD discovery samples were combined (PGC+iPSYCH; Figure 1, Table S5), including those of European ancestry only (PGC+iPSYCH(EUR), Table S5), and genetic overlap was present for all LRAs studied. Results were not affected by the exclusion of children with an ADHD diagnosis in ALSPAC (Table S6).

Figure 1 here

Shared genetic liability between ADHD and LRA with EA

There was strong evidence for a moderate negative genetic correlation (r_g=-0.53(SE=0.03), $P<1x10^{-10}$) between genetically predicted ADHD, as captured by the largest ADHD discovery sample (PGC+iPSYCH), and EA (LDSC-h^2=0.11(SE=0.004)), consistent with previous findings[21]. Likewise, LRAs were moderately to highly positively correlated with EA (e.g. reading speed age 13 r_g=0.80(SE=0.22), $P=3.0x10^{-4}$; Table S7), as previously reported[20]. Additionally, two independent variants reached genome-wide significance for both ADHD[21] and EA[19], consistent with biological pleiotropy[42] (i.e. single genetic loci influencing multiple traits)[42]. These findings indicate complex, potentially reciprocal cross-trait relationships (Figure 2a) and violate MR causal modelling assumptions[22]. Consequently, ADHD instruments are not valid MR instruments as they are not independent of EA, a potential genetic confounder.

Figure 2 here

Multivariable regression analyses

To disentangle the genetic overlap of polygenic ADHD risk with literacy- and language-related measures into genetic ADHD effects independent of and shared with EA, we applied MVR[22] using ADHD instruments based on the most powerful ADHD discovery sample (PGC+iPSYCH) (Figure 2b).

Using conservative ADHD instruments (Table S8), only non-word reading accuracy age 9 was associated with polygenic ADHD conditional on EA, with a decrease of 0.37 SD per log-odds increase in ADHD risk (β_{ADHD}=-0.37(SE=0.09), P=1.7x10^{-3}) (Table 3). Additionally, when meta-analysing LRAs across pre-defined domains, we observed an effect of polygenic ADHD risk on pooled reading performance, independent of EA. This corresponds to a decrease of 0.35 SD in pooled reading performance per log-odds increase in ADHD liability (β_{ADHD}=-0.35(SE=0.09), P=9.2x10^{-5}, P_{het}=0.19). There was little evidence for association with spelling or combined LRA performance conditional on EA (Table 3).

Using subthreshold ADHD instruments (Table S8), polygenic ADHD effects on LRA performance, conditional on EA, were detectable for all reading- and spelling-related measures, phonemic awareness and verbal intelligence, but not language-related skills such as listening comprehension and non-word repetition (Table 3). Observable effects were, however, considerably smaller than those captured by conservative ADHD instruments although the statistical evidence increased, such as a 0.03 SD decrease in non-word reading accuracy at age 9 per log-odds increase in ADHD risk (β_{ADHD}=-0.03(SE=0.01), P=2.4x10^{-4}, Table 3).

Polygenic ADHD effects on literacy- and language-related performance that are shared with EA were identified for all LRAs studied, but only using subthreshold, not

conservative ADHD instruments (Table 3). This translates into, for example, a further 0.46 SD units decrease in non-word reading accuracy at age 9 per missing school year (β_{EA}=-0.46(SE=0.10), P=6.1x10^{-6}, Table 3). Thus, the association between polygenic ADHD risk and listening comprehension and non-word repetition, is fully attributable to genetic effects shared with EA (Table 3). MVR using fully standardised EA, ADHD and LRA regression estimates showed that ADHD effects shared with EA were as large as or even larger than ADHD effects independent of EA (Figure 2c, Table S9).

Using an analogous approach, we also disentangled the genetic overlap of polygenic EA with LRAs into genetic EA effects independent of and shared with ADHD based on EA instruments (Figure S1). As observed using ADHD instruments, there was strong evidence for EA effects shared with ADHD using subthreshold, but not conservative EA instruments (Table S10). Importantly, the magnitude of ADHD genetic effects shared with EA compared to EA genetic effects shared with ADHD was largely consistent with each other in fully standardised analyses (Table S9 and S10).

All MVRs were performed using constrained intercepts[22]. There was little evidence supporting the inclusion of regression intercepts that would imply additional genetic effects using these instruments, not yet captured by either ADHD or EA effect estimates.

Table 3 here

Attrition in ALSPAC

Analyses of sample drop-out in ALSPAC, exemplified by missing reading accuracy and comprehension scores at age 7 (WORD), were carried out to investigate potential sources of

bias. We identified a positive genetic association between sample-dropout and ADHD-PGS (PGC+iPSYCH, OR=1.03(SE=0.005), P=1.4x10^{-8}, Table S11), consistent with previous studies[43]. Disentangling this polygenic link using MVR (Table S12) based on ADHD subthreshold instruments showed a 0.05 increase in log-odds sample drop-out per log-odds increase in liability to ADHD independent of EA (log(OR)=0.05(SE=0.01), P=3.7x10^{-4}). Additionally, a further 1.04 increase in log odds sample drop-out per missing year of schooling (log(OR)=1.04(SE=0.19),P=7.3x10^{-8}) was observed.

Discussion

This study identified strong evidence for an inverse association between polygenic ADHD risk and multiple population-based LRAs using a polygenic scoring approach. However, these associations are genetically confounded through effects shared with genetically predictable EA. Accurate modelling of polygenic links using MVR techniques, conditional on EA, revealed an ADHD-specific association profile that primarily involves literacy-related impairments, but few language-related problems. Once genetic confounding is accounted for, polygenic ADHD risk was most strongly inversely associated with an overarching reading domain, in addition to spelling, phonemic awareness and verbal intelligence, but not listening comprehension and non-word repetition abilities. Importantly, ADHD genetic effects shared with EA inflated or even induced the genetic overlap between polygenic ADHD risk and all LRAs studied.

These findings are consistent with previous clinical and twin studies reporting a shared underlying aetiology between ADHD and literacy-related impairments[1–3,6–8,10,44,45]. They also reflect the observed robust evidence for an association between ADHD-PGS and several reading- and spelling-related abilities, using independent ADHD discovery samples.

The identified association profile suggests that reading-related deficits in ADHD may, as hypothesised for RD, arise from a phonological impairment[46,47], which affects decoding and reading skills[48], but also spelling abilities[49]. However, reading abilities can, once developed, also shape phonological skills[50].

In addition, this study supports the genetic confounding of associations between polygenic ADHD and LRAs through genetically predictable EA, and, equally likely, the genetic confounding of associations between polygenic EA and LRAs through genetically predictable ADHD. The magnitude of these confounding effects was comparable, consistent with a model of reciprocal genetic influences between EA and ADHD (Figure 2a). Importantly, both directions of effect are supported by observational[51] and twin[52] studies. Thus, more generally, our findings support a multiple-deficit model proposing shared underlying aetiologies between ADHD and RD dimensions[15].

Disentangling the underlying aetiological relationships between ADHD, EA and LRAs is challenging due to violations of classical MR assumptions[22]. This study focusses therefore on dissecting the genetic association between ADHD and LRAs into ADHD genetic effects that are shared with or independent of genetic confounders, such as EA, while controlling for collider bias[38]. Using MVR, the detection of these polygenic associations was strongly governed by the choice of genetic variants. Conservative ADHD instruments identified large ADHD-specific effects on reading as a domain, independent of EA, and little evidence for genetic effects shared with EA. However, conservative instruments had limited power[53]. They comprised 15 independent SNPs only, but included genetic variation within *FOXP2,* a gene that has been implicated in childhood apraxia of speech as well as expressive and receptive language impairments (http://omim.org/entry/602081)[54]. Subthreshold instruments including thousands of variants, on the other hand, tagged ADHD-specific polygenic associations

(conditional on EA) of smaller magnitude, but with higher predictive accuracy. Moreover, these instruments detected strong evidence for genetic confounding affecting all polygenic links between ADHD and LRAs studied, with associations of equal strength and at least equal magnitude compared to associations independent of EA. In general, these findings are thus in support of an omnigenic[55] model of complex trait architectures, compatible with a general factor model of psychopathology[56], including ADHD[57]. The omnigenic model construes that only the largest-effect variants will reflect ADHD specificity, and may thus tag the most trait-specific associations between ADHD and reading, independent of EA. The majority of variants, however, will capture pleiotropic (omnigenic) influences pointing to highly interconnected neural networks[55] that give rise to genetic confounding. Consequently, the majority of subthreshold variants, captured by both ADHD and EA subthreshold variants, are likely to represent highly powerful cross-trait genetic predictors that enhance and induce genetic overlap.

Finally, the methodological framework within this study has relevance not only for studies investigating polygenic links between ADHD and LRAs, but all studies examining polygenic overlap between ADHD and phenotypes that share genetically liability with EA or any other source of genetic confounding. Specifically, our findings suggest that lower variant selection thresholds can introduce genetic confounding that needs to be accounted for before identified associations can be interpreted in terms of underlying mechanisms, including shared genetic aetiologies. This is especially important as current guidelines for studying polygenic links with allelic scores recommend aggregating genetic variants across less stringent significance thresholds to maximise genetic association between discovery and target samples[58,59].

This study has several limitations. Firstly, ALSPAC, as other cohort studies, suffers from attrition[43,60]. Sensitivity analyses showed that this is unlikely to bias our findings based on conservative instruments. However, links identified using subthreshold ADHD variants, might have been underestimated given that individuals with a genetic predisposition to ADHD are more likely to drop out[43]. Secondly, the strength of the genetic overlap between polygenic ADHD risk and LRAs may vary according to ADHD symptom domain levels, implicating especially inattentiveness[61], as well as the nature of the literacy- or language-related ability involved (as we observed evidence for effect heterogeneity when combining all LRAs). It is conceivable that also other verbal abilities, not investigated in this study, such as grammar, expressive vocabulary or pragmatic skills, may genetically overlap with ADHD. Finally, we only studied evidence for genetic confounding with respect to genetically predictable EA using ADHD instruments (and vice versa genetically predictable ADHD using EA instruments). However, we found little evidence for the presence of additional unknown confounders, not yet tagged by these instruments. Furthermore, the power of available LRA GWAS statistics is still too low to generate genetic instruments supporting reverse models. Larger and more detailed clinical and population-based samples, as well as extensive multivariate variance analyses of the spectrum of LRAs (that are currently computationally expensive[62]) will help to discover and characterize further overlap between ADHD and literacy- and language-related cognitive processes.

Conclusion

Polygenic ADHD links with LRAs are susceptible to genetic confounding through EA, especially when investigated with genetic variants typically selected for polygenic scoring approaches, inflating and biasing genetic overlap. Accounting for these unspecific genetic

effects, reveals an ADHD-specific association profile that primarily involves reading-related impairments, but few language-related problems.

Acknowledgements

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The UK Medical Research Council and the Wellcome Trust (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. The ALSPAC GWAS data was generated by Sample Logistics and Genotyping Facilities at the Wellcome Trust Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe.

This research was facilitated by the Social Science Genetic Association Consortium (SSGAC) and Psychiatric Genomics Consortium (PGC) by providing access to genome-wide summary statistics.

This publication is the work of the authors and EV and BSTP will serve as guarantors for the contents of this paper. EV, CYS, BSTP and SEF are supported by the Max Planck Society. The iPSYCH project is funded by the Lundbeck Foundation (grant no R102-A9118 and R155-2014-1724) and the universities and university hospitals of Aarhus and Copenhagen. Genotyping of iPSYCH and PGC samples was supported by grants from the Lundbeck Foundation, the Stanley Foundation, the Simons Foundation (SFARI 311789 to Mark J Daly), and NIMH (5U01MH094432-02 to Mark J Daly). The Danish National Biobank resource was supported by the Novo Nordisk Foundation. Data handling and analysis was supported by NIMH (1U01MH109514-01 to Michael O'Donovan and Anders D Børglum). High-performance computer capacity for handling and statistical analysis of

iPSYCH data on the GenomeDK HPC facility was provided by the Centre for Integrative Sequencing, iSEQ, Aarhus University, Denmark (grant to Anders D Børglum).

An earlier version of this manuscript is available as a preprint on bioRxiv.

iPSYCH-Broad-PGC ADHD Consortium members

(listed alphabetically)

Esben Agerbo[1,10,11], Thomas Damm Als[1,2,3],Marie Bækved-Hansen[1,12], Rich Belliveau[5], Anders D. Børglum[1,2,3], Jonas Bybjerg-Grauholm[1,12], Felecia Cerrato[5], Kimberly Chambert[5], Claire Churchhouse[4,5,13], Søren Dalsgaard[10], Mark J. Daly[4,5,13], Ditte Demontis[1,2,3], Ashley Dumont[5], Jacqueline Goldstein[4,5,13], Jakob Grove[1,2,3,14], Christine S. Hansen[1,12,15], Mads Engel Hauberg[1,2,3], Mads V. Hollegaard[1,12], David M. Hougaard[1,12], Daniel P. Howrigan[4,5], Hailiang Huang[4,5], Julian Maller[5,16], Alicia R. Martin[4,5,13], Joanna Martin[5,6,7], Manuel Mattheisen[1,2,3,8,9], Jennifer Moran[5], Ole Mors[1,21], Preben Bo Mortensen[1,2,10,11], Benjamin M. Neale[4,5,13], Merete Nordentoft[1,19], Jonatan Pallesen[1,2,3], Duncan S. Palmer[4,5], Carsten Bøcker Pedersen[1,10,11], Marianne Giørtz Pedersen[1,10,11], Timothy Poterba[4,5,13], Jesper Buchhave Poulsen[1,12], Stephan Ripke[4,5,13,17], Elise B. Robinson[4,18], F. Kyle Satterstrom[4,5,13], Christine Stevens[5], Patrick Turley[4,5], Raymond K. Walters[4,5], Thomas Werge[1,15,20]

1. The Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH, Denmark

2. Centre for Integrative Sequencing, iSEQ, Aarhus University, Aarhus, Denmark

3. Department of Biomedicine - Human Genetics, Aarhus University, Aarhus, Denmark

4. Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA

5. Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA

6. Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

7. MRC Centre for Neuropsychiatric Genetics & Genomics, School of Medicine, Cardiff University, Cardiff, United Kingdom

8. Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden

9. Stockholm Health Care Services, Stockholm County Council, Stockholm, Sweden

10. National Centre for Register - Based Research, Aarhus University, Aarhus, Denmark

11. Centre for Integrated Register - based Research, Aarhus University, Aarhus, Denmark

12. Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark

13. Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA

14. Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark

15. Institute of Biological Psychiatry, MHC Sct. Hans, Mental Health Services Copenhagen, Roskilde, Denmark

16. Genomics plc, Oxford, United Kingdom

17. Department of Psychiatry, Charite Universitatsmedizin Berlin Campus Benjamin Franklin, Berlin, Germany

18. Department of Epidemiology, Harvard Chan School of Public Health, Boston, Massachusetts, USA

19. Mental Health Services in the Capital Region of Denmark, Mental Health Center Copenhagen, University of Copenhagen, Copenhagen, Denmark

20. Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark

21. Psychosis Research Unit, Aarhus University Hospital, Risskov, Denmark

Financial Disclosures

The authors declare no conflict of interest.

References

1. Geurts HM, Embrechts M. Language profiles in ASD, SLI, and ADHD. *J Autism Dev Disord*. 2008;38(10):1931-1943. doi:10.1007/s10803-008-0587-1

2. Helland WA, Posserud M-B, Helland T, Heimann M, Lundervold AJ. Language Impairments in Children With ADHD and in Children With Reading Disorder. *J Atten Disord*. October 2012:1087054712461530. doi:10.1177/1087054712461530

3. Germanò E, Gagliano A, Curatolo P. Comorbidity of ADHD and dyslexia. *Dev Neuropsychol*. 2010;35(5):475-493. doi:10.1080/87565641.2010.494748

4. Maughan B, Carroll J. Literacy and mental disorders. *Curr Opin Psychiatry*. 2006;19(4):350-354. doi:10.1097/01.yco.0000228752.79990.41

5. Mayes SD, Calhoun SL. Learning, attention, writing, and processing speed in typical children and children with ADHD, autism, anxiety, depression, and oppositional-defiant disorder. *Child Neuropsychol J Norm Abnorm Dev Child Adolesc*. 2007;13(6):469-493. doi:10.1080/09297040601112773

6. Mayes SD, Calhoun SL. Frequency of reading, math, and writing disabilities in children with clinical disorders. *Learn Individ Differ*. 2006;16(2):145-157. doi:10.1016/j.lindif.2005.07.004

7. Czamara D, Tiesler CMT, Kohlböck G, et al. Children with ADHD Symptoms Have a Higher Risk for Reading, Spelling and Math Difficulties in the GINIplus and LISAplus Cohort Studies. *PLOS ONE*. 2013;8(5):e63859. doi:10.1371/journal.pone.0063859

8. Åsberg J, Kopp S, Berg-Kelly K, Gillberg C. Reading comprehension, word decoding and spelling in girls with autism spectrum disorders (ASD) or attention-deficit/hyperactivity disorder (AD/HD): performance and predictors. *Int J Lang Commun Disord*. 2010;45(1):61-71. doi:10.3109/13682820902745438

9. Green BC, Johnson KA, Bretherton L. Pragmatic language difficulties in children with hyperactivity and attention problems: an integrated review. *Int J Lang Commun Disord*. 2014;49(1):15-29. doi:10.1111/1460-6984.12056

10. Hawkins E, Gathercole S, Astle D, The Calm Team, Holmes J. Language Problems and ADHD Symptoms: How Specific Are the Links? *Brain Sci*. 2016;6(4). doi:10.3390/brainsci6040050

11. Polanczyk G, de Lima MS, Horta BL, Biederman J, Rohde LA. The Worldwide Prevalence of ADHD: A Systematic Review and Metaregression Analysis. *Am J Psychiatry*. 2007;164(6):942-948. doi:10.1176/ajp.2007.164.6.942

12. Peterson RL, Pennington BF. Developmental dyslexia. *Lancet Lond Engl*. 2012;379(9830):1997-2007. doi:10.1016/S0140-6736(12)60198-6

13. Thapar A, Cooper M. Attention deficit hyperactivity disorder. *The Lancet*. 2016;387(10024):1240-1250. doi:10.1016/S0140-6736(15)00238-X

14. Gough PB, Tunmer WE. Decoding, Reading, and Reading Disability. *Remedial Spec Educ*. 1986;7(1):6-10. doi:10.1177/074193258600700104

15. Pennington BF. From single to multiple deficit models of developmental disorders. *Cognition*. 2006;101(2):385-413. doi:10.1016/j.cognition.2006.04.008

16. Martin NC, Levy F, Pieka J, Hay DA. A Genetic Study of Attention Deficit Hyperactivity Disorder, Conduct Disorder, Oppositional Defiant Disorder and Reading Disability: Aetiological overlaps and implications. *Int J Disabil Dev Educ*. 2006;53(1):21-34. doi:10.1080/10349120500509992

17. Willcutt EG, Pennington BF, DeFries JC. Twin study of the etiology of comorbidity between reading disability and attention-deficit/hyperactivity disorder. *Am J Med Genet*. 2000;96(3):293-301. doi:10.1002/1096-8628(20000612)96:3<293::AID-AJMG12>3.0.CO;2-C

18. Willcutt EG, Pennington BF, Olson RK, DeFries JC. Understanding comorbidity: A twin study of reading disability and attention-deficit/hyperactivity disorder. *Am J Med Genet B Neuropsychiatr Genet*. 2007;144B(6):709-714. doi:10.1002/ajmg.b.30310

19. Okbay A, Beauchamp JP, Fontana MA, et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*. 2016;533(7604):539-542. doi:10.1038/nature17671

20. Selzam S, Dale PS, Wagner RK, et al. Genome-Wide Polygenic Scores Predict Reading Performance Throughout the School Years. *Sci Stud Read*. March 2017:1-16. doi:10.1080/10888438.2017.1299152

21. Demontis D, Walters RK, Martin J, et al. Discovery Of The First Genome-Wide Significant Risk Loci For ADHD. *bioRxiv*. June 2017:145581. doi:10.1101/145581

22. Burgess S, Thompson DJ, Rees JMB, Day FR, Perry JR, Ong KK. Dissecting Causal Pathways Using Mendelian Randomization with Summarized Genetic Data: Application to Age at Menarche and Risk of Breast Cancer. *Genetics*. January 2017:genetics.300191.2017. doi:10.1534/genetics.117.300191

23. Fraser A, Macdonald-Wallis C, Tilling K, et al. Cohort Profile: the Avon Longitudinal Study of Parents and Children: ALSPAC mothers cohort. *Int J Epidemiol*. 2013;42(1):97-110. doi:10.1093/ije/dys066

24. Boyd A, Golding J, Macleod J, et al. Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol*. 2013;42(1):111-127. doi:10.1093/ije/dys064

25. Li J, Ji L. Adjusting multiple testing in multilocus analyses using the eigenvalues of a correlation matrix. *Heredity*. 2005;95(3):221-227. doi:10.1038/sj.hdy.6800717

26. Goodman R, Ford T, Richards H, Gatward R, Meltzer H. The Development and Well-Being Assessment: Description and Initial Validation of an Integrated Assessment of Child and Adolescent Psychopathology. *J Child Psychol Psychiatry*. 2000;41(5):645-655. doi:10.1111/j.1469-7610.2000.tb02345.x

27. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet*. 2013;45(9):984-994. doi:10.1038/ng.2711

28. Mors O, Perto GP, Mortensen PB. The Danish Psychiatric Central Research Register. *Scand J Public Health*. 2011;39(7 Suppl):54-57. doi:10.1177/1403494810395825

29. Pedersen CB, Bybjerg-Grauholm J, Pedersen MG, et al. The iPSYCH2012 case-cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Mol Psychiatry*. September 2017. doi:10.1038/mp.2017.196

30. Yang J, Benyamin B, McEvoy BP, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet*. 2010;42(7):565-569. doi:10.1038/ng.608

31. Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinforma Oxf Engl*. 2012;28(19):2540-2542. doi:10.1093/bioinformatics/bts474

32. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88(1):76-82. doi:10.1016/j.ajhg.2010.11.011

33. Bulik-Sullivan BK, Loh P-R, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291-295. doi:10.1038/ng.3211

34. Bulik-Sullivan B, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015;47:1236–1241. doi:10.1038/ng.3406

35. International Schizophrenia Consortium, Purcell SM, Wray NR, et al. Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. *Nature*. 2009;460(7256):748-752. doi:10.1038/nature08185

36. Dudbridge F. Power and Predictive Accuracy of Polygenic Risk Scores. *PLOS Genet*. 2013;9(3):e1003348. doi:10.1371/journal.pgen.1003348

37. Stergiakouli E, Davey Smith G, Martin J, et al. Shared genetic influences between dimensional ASD and ADHD symptoms during child and adolescent development. *Mol Autism*. 2017;8:18. doi:10.1186/s13229-017-0131-2

38. Aschard H, Vilhjálmsson BJ, Joshi AD, Price AL, Kraft P. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *Am J Hum Genet*. 2015;96(2):329-339. doi:10.1016/j.ajhg.2014.12.021

39. Pierce BL, Ahsan H, Vanderweele TJ. Power and instrument strength requirements for Mendelian randomization studies using multiple genetic variants. *Int J Epidemiol*. 2011;40(3):740-752. doi:10.1093/ije/dyq151

40. Purcell S, Neale B, Todd-Brown K, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet*. 2007;81(3):559-575.

41. Howie BN, Donnelly P, Marchini J. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLOS Genet*. 2009;5(6):e1000529. doi:10.1371/journal.pgen.1000529

42. Davey Smith G, Hemani G. Mendelian randomization: genetic anchors for causal inference in epidemiological studies. *Hum Mol Genet*. 2014;23(R1):R89-R98. doi:10.1093/hmg/ddu328

43. Taylor A, Jones H, Sallis H, et al. The molecular genetics of participation in the Avon Longitudinal Study of Parents and Children. *bioRxiv*. October 2017:206698. doi:10.1101/206698

44. Cheung CHM, Wood AC, Paloyelis Y, et al. Aetiology for the covariation between combined type ADHD and reading difficulties in a family study: the role of IQ. *J Child Psychol Psychiatry*. 2012;53(8):864-873. doi:10.1111/j.1469-7610.2012.02527.x

45. Paloyelis Y, Rijsdijk F, Wood AC, Asherson P, Kuntsi J. The Genetic Association Between ADHD Symptoms and Reading Difficulties: The Role of Inattentiveness and IQ. *J Abnorm Child Psychol*. 2010;38(8):1083-1095. doi:10.1007/s10802-010-9429-7

46. Carroll JM, Snowling MJ. Language and phonological skills in children at high risk of reading difficulties. *J Child Psychol Psychiatry*. 2004;45(3):631-640. doi:10.1111/j.1469-7610.2004.00252.x

47. *Developmental Disorders of Language Learning and Cognition*. https://books.google.com/books/about/Developmental_Disorders_of_Language_Lear.html?id=GfahZ6HS1rYC. Accessed January 19, 2018.

48. Rack J, Snowling M, Olson R. The Nonword Reading Deficit in Developmental Dyslexia - a Review. *Read Res Q*. 1992;27(1):28-53. doi:10.2307/747832

49. Treiman R. *Beginning to Spell: A Study of First-Grade Children*. Oxford University Press; 1992.

50. Kamhi AG, Catts HW. *Language and Reading Disabilities*. 3 edition. Boston: Pearson; 2011.

51. Larsson H, Sariaslan A, Långström N, D'Onofrio B, Lichtenstein P. Family income in early childhood and subsequent attention deficit/hyperactivity disorder: a quasi-experimental study. *J Child Psychol Psychiatry*. 2014;55(5):428-435. doi:10.1111/jcpp.12140

52. de Zeeuw EL, van Beijsterveldt CEM, Ehli EA, de Geus EJC, Boomsma DI. Attention Deficit Hyperactivity Disorder Symptoms and Low Educational Achievement: Evidence Supporting A Causal Hypothesis. *Behav Genet*. 2017;47(3):278-289. doi:10.1007/s10519-017-9836-4

53. Schatzkin A, Abnet CC, Cross AJ, et al. Mendelian Randomization: How It Can—and Cannot—Help Confirm Causal Relations between Nutrition and Cancer. *Cancer Prev Res (Phila Pa)*. 2009;2(2):104-113. doi:10.1158/1940-6207.CAPR-08-0070

54. Lai CS, Fisher SE, Hurst JA, Vargha-Khadem F, Monaco AP. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature*. 2001;413(6855):519-523. doi:10.1038/35097076

55. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*. 2017;169(7):1177-1186. doi:10.1016/j.cell.2017.05.038

56. Lahey BB, Applegate B, Hakes JK, Zald DH, Hariri AR, Rathouz PJ. Is There a General Factor of Prevalent Psychopathology during Adulthood? *J Abnorm Psychol*. 2012;121(4):971-977. doi:10.1037/a0028355

57. Franke B. What's in a Name: the "Omnigenic" Model as a Point of Departure for Polygenic Psychiatric Disorders. *JPBS*. 2017. doi:10.20900/jpbs.20170014(S1-S8)

58. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinformatics*. 2015;31(9):1466-1468. doi:10.1093/bioinformatics/btu848

59. Wray NR, Lee SH, Mehta D, Vinkhuyzen AAE, Dudbridge F, Middeldorp CM. Research Review: Polygenic methods and their application to psychiatric traits. *J Child Psychol Psychiatry*. 2014;55(10):1068-1087. doi:10.1111/jcpp.12295

60. Martin J, Tilling K, Hubbard L, et al. Association of Genetic Risk for Schizophrenia With Nonparticipation Over Time in a Population-Based Cohort Study. *Am J Epidemiol*. 2016;183(12):1149-1158. doi:10.1093/aje/kww009

61. Greven CU, Harlaar N, Dale PS, Plomin R. Genetic Overlap between ADHD Symptoms and Reading is largely Driven by Inattentiveness rather than Hyperactivity-Impulsivity. *J Can Acad Child Adolesc Psychiatry*. 2011;20(1):6-14.

62. St Pourcain B, Eaves LJ, Ring SM, et al. Developmental changes within the genetic architecture of social communication behaviour: A multivariate study of genetic variance in unrelated individuals. *Biol Psychiatry*. September 2017. doi:10.1016/j.biopsych.2017.09.020

63. *WORD, Wechsler Objective Reading Dimensions Manual*. Psychological Corporation; 1993.

64. Nunes T, Bryant P, Olsson J. Learning Morphological and Phonological Spelling Rules: An Intervention Study. *Sci Stud Read*. 2003;7(3):289-307. doi:10.1207/S1532799XSSR0703_6

65. Neale MD. *Neale Analysis of Reading Ability: Second Revised British Edition*. London: NFER-Nelson; 1997.

66. *TOWRE, Test of Word Reading Efficiency: Examiner's Manual*. PRO-ED; 1999.

67. Rosner J, Simon DP. The Auditory Analysis Test: An Initial Report. *J Learn Disabil*. 1971;4(7):384-392. doi:10.1177/002221947100400706

68. Rust J, Wechsler D, Psychological Corporation. *WOLD: Wechsler Objective Language Dimensions*. London; [Hong Kong: Psychological Corp.; Artsberg Enterprises Ltd. [distributor; 1996.

69. Gathercole SE, Willis CS, Baddeley AD, Emslie H. The Children's Test of Nonword Repetition: a test of phonological working memory. *Mem Hove Engl*. 1994;2(2):103-127. doi:10.1080/09658219408258940

70. Prifitera A, Saklofske DH. *WISC-III Clinical Use and Interpretation: Scientist-Practitioner Perspectives*. Academic Press; 1998.

Figure legends

Figure 1: Phenotypic variance in literacy- and language-related abilities explained by polygenic ADHD risk

Abbreviations: a, accuracy; c, comprehension; s, speed; WORD, Wechsler Objective Reading Dimension; NBO, Nunes, Bryant and Olson (ALSPAC specific instrument); NARA II, The Neale Analysis of Reading Ability- Second Revised British Edition; TOWRE, Test Of Word Reading Efficiency; NW, non-word; NB, Nunes and Bryant (ALSPAC specific instrument); PhonAware, phonemic awareness; AAT, Auditory Analysis Test; WOLD, Wechsler Objective Language Dimensions; NWR, Non-word Repetition Test; VIQ, verbal intelligence quotient; WISC-III, Wechsler Intelligence Scale for Children III; PGC, Psychiatric Genomics Consortium; iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research; ADHD, Attention-Deficit/Hyperactivity Disorder.

a) Schematic representation of polygenic scoring analyses. ADHD polygenic scores were created in ALSPAC using PGC, iPSYCH and PGC+iPSYCH GWAS summary statistics. Rank-transformed LRAs were regressed on Z-standardised ADHD-PGS using ordinary least square regression. b) Phenotypic variance in literacy- and language-related abilities explained by polygenic ADHD risk.

* - Evidence for association between LRAs and polygenic ADHD risk as observed in PGC ADHD, iPSYCH ADHD and PGC+iPSYCH ADHD samples. Note that all LRAs were associated with polygenic ADHD risk in iPSYCH ADHD and PGC+iPSYCH ADHD passing the experiment-wide error rate ($P<0.007$).

Figure 2: Genetic relationships between ADHD, educational attainment and literacy- and language-related abilities

Abbreviations: ADHD, Attention-Deficit/Hyperactivity Disorder; EA, educational attainment; LRAs, literacy and language-related abilities; PGC, Psychiatric Genomics Consortium; iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research; SSGAC, Science Genetic Association Consortium; ALSPAC, Avon Longitudinal Study of Parents And Children; MVR, multivariable regression

(a) Hypothesized biological model of genetic relationships between ADHD, EA, and LRAs, including complex, pleiotropic and reciprocal genetic links, (b) Schematic MVR model assessing polygenic ADHD-LRA overlap independent of and shared with genetic effects for EA, (c) MVR estimates of ADHD-specific effects independent of EA and ADHD effects shared with EA on LRAs using standardised ADHD instruments: Sets of conservative ($P<5x10^{-8}$) and subthreshold ($P<0.0015$) ADHD instruments were extracted from ADHD (PGC+iPSYCH), EA (SSGAC) and LRAs (ALSPAC) GWAS summary statistics. ADHD-specific effects independent of EA (β_{ADHD}) and ADHD effects shared with EA (β_{EA}) on LRAs were estimated with MVRs. To compare the magnitude of β_{ADHD} and β_{EA}, MVR analyses were conducted using standardised regression estimates (Supplementary Methods). β_{ADHD} estimates measure the change in LRA Z-score per Z-score in ADHD liability. β_{EA} estimates measure the change in LRA Z-scores per Z-score in missing school years. MVR estimates based on raw genetic effect estimates are provided in Table 3. Pooled estimates for reading, spelling and global LRA measures (Table 1) were obtained through random-effects meta-regression. Only effects passing the experiment-wide significance threshold ($P<0.007$) are shown with corresponding 95% confidence intervals. There is no causality inferred.

* ADHD effects shared with EA were assessed through EA genetic effect estimates of ADHD-associated variants.

Tables

Table 1: Literacy and literacy- and language-related oral language abilities in the Avon Longitudinal Study of Parents and Children

LRA (psychological instrument)	Mean Score (SE)	Mean Age (SE)	N (%males)	Trait combinations
Reading accuracy and comprehension (WORD[63]), words	28.44 (9.24)	7.53 (0.31)	5,891 (50.6)	Reading; Global literacy- and language-related ability
Reading accuracy (ALSPAC specific: NBO[64]), words	7.55 (2.44)	9.87 (0.32)	5,738 (49.3)	
Reading speed[1] (NARA II[65]), passages	105.50 (12.47)	9.88 (0.32)	5,189 (49.1)	
Reading accuracy[1] (NARA II[65]), passages	104.11 (13.62)	9.88 (0.32)	5,201 (49.1)	
Reading speed (TOWRE[66]), words	82.58 (10.28)	13.83 (0.20)	4,247 (48.4)	
Non-word reading accuracy (ALSPAC specific: NBO[64])	5.24 (2.48)	9.87 (0.32)	5,731 (49.2)	
Non-word reading speed (TOWRE[66])	50.82 (9.38)	13.83 (0.20)	4,237 (48.3)	
Spelling accuracy (ALSPAC specific: NB)	7.89 (4.39)	7.53 (0.31)	5,800 (50.2)	Spelling
Spelling accuracy (ALSPAC specific: NB)	10.27 (3.43)	9.87 (0.32)	5,728 (49.2)	
Phonemic awareness (AAT[67])	20.23 (9.51)	7.53 (0.31)	5,919 (50.6)	
Listening comprehension (WOLD[68])	7.50 (1.96)	8.63 (0.30)	5,473 (49.9)	
Non-word repetition (NWR[69])	7.26 (2.51)	8.63 (0.30)	5,464 (49.9)	
Verbal intelligence[1] (WISC-III[70])	107.85 (16.74)	8.64 (0.31)	5,456 (49.7)	

1. Scores were derived using age norms and adjusted for sex and principal components only before transformation

Abbreviations: LRAs, literacy- and language-related abilities; WORD, Wechsler Objective Reading Dimension; ALSPAC, Avon Longitudinal study of Parents and Children; NBO, ALSPAC-specific assessment developed by Nunes, Bryant and Olson; NARA II, The Neale Analysis of Reading Ability- Second Revised British Edition; TOWRE, Test Of Word Reading Efficiency; NB, ALSPAC-specific assessment developed by Nunes and Bryant; AAT, Auditory Analysis Test; WOLD, Wechsler Objective Language Dimensions; NWR, Nonword Repetition Test; WISC-III, Wechsler Intelligence Scale for Children III.

Thirteen LRAs capturing aspects related to reading, spelling, phonemic awareness, listening comprehension, non-word repetition and verbal intelligence were assessed in 7 to 13 year-old ALSPAC participants using both standardised and ALSPAC-specific instruments (Supplementary Information).

Table 2: Sample description

Phenotype	Sample	Source	Ethnicity	Imputation reference panel	N
LRAs	ALSPAC	General population	White European	HRC r1.1	≤ 5,891
ADHD	PGC	Clinical sample	Predominantly white European	HapMap phase 3	16,203 (N_{cases}=4,163)
	iPSYCH	Clinical sample	White European	1000 Genomes phase 3	37,076 (N_{cases}=14,584)
	PGC + iPSYCH (EUR)	Clinical sample	White European	1000 Genomes phase 3	53,293 (N_{cases}=19,099)
	PGC + iPSYCH	Clinical sample	Predominantly white European	1000 Genomes phase 3	55,374 (N_{cases}=20,183)
EA	SSGAC	Predominantly general population	White European	1000 Genomes phase 3[a]	326,041

a - Predominantly 1000 Genomes phase 3[19]

Abbreviations: LRAs, literacy- and language-related abilities; ADHD, Attention-Deficit/Hyperactivity Disorder; EA, educational attainment; ALSPAC, Avon Longitudinal study of Parents and Children; PGC, Psychiatric Genomics Consortium; iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research; EUR, European ancestry; SSGAC, Social Science Genetic Consortium; HRC, The Haplotype Reference Consortium

Note that there is no overlap between LRA, ADHD and EA samples.

Table 3: Multivariable regression analysis of polygenic associations between ADHD and literacy- and language-related abilities (raw estimates)

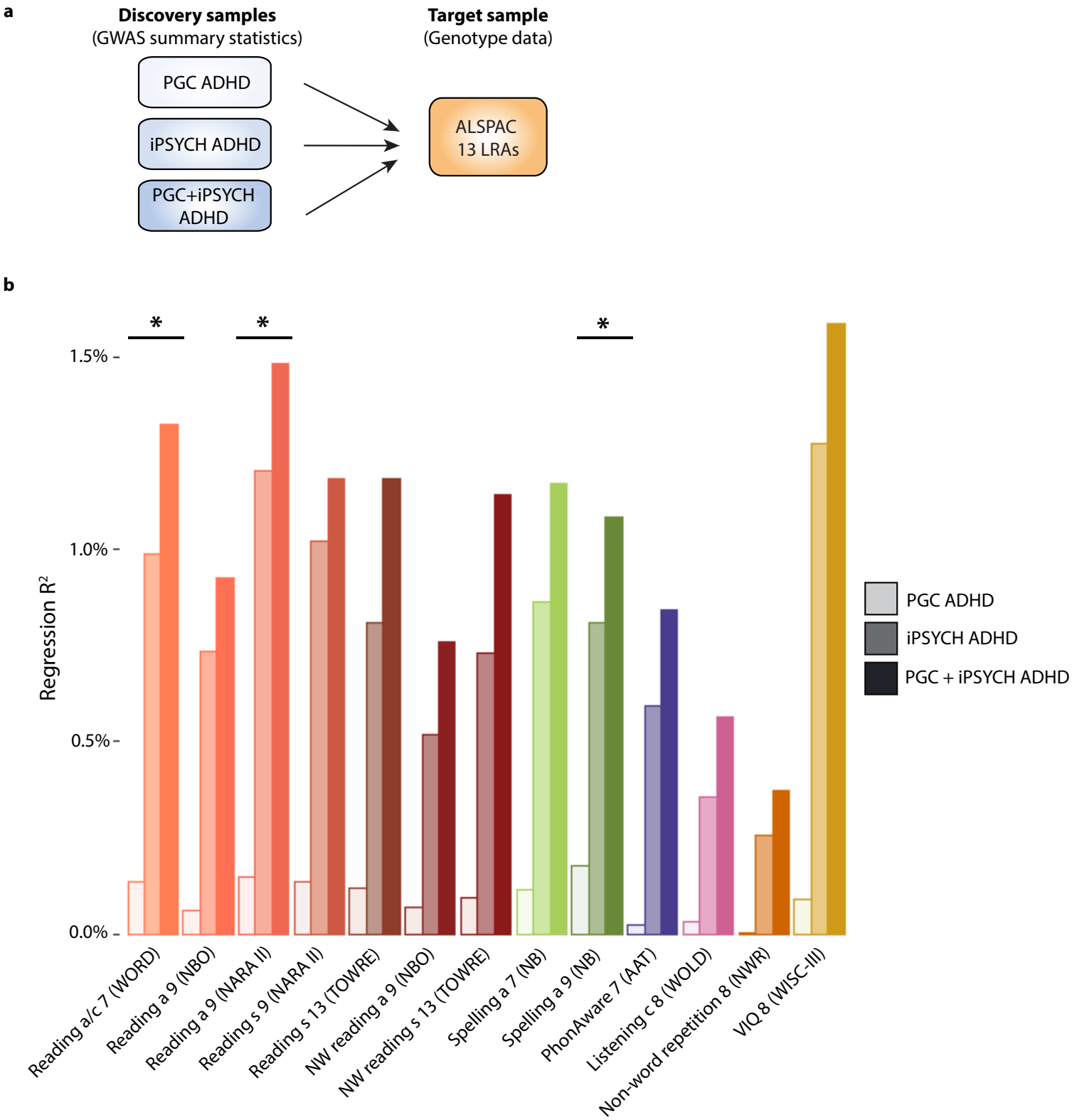
LRAs	ADHD (β_{ADHD}) (ADHD-specific effects independent of EA)						EA (β_{EA}) (EA genetic effects of ADHD-associated variants)[1]					
	Conservative instruments (P_{thr}<5x10^{-8})			Subthreshold instruments (P_{thr}<0.0015)			Conservative instruments (P_{thr}<5x10^{-8})			Subthreshold instruments (P_{thr}<0.0015)		
	β (SE)	*P*	P_{het}	β (SE)	*P*	P_{het}	β (SE)	*P*	P_{het}	β (SE)	*P*	P_{het}
Reading a/c 7 (WORD)	-0.18 (0.13)	0.20	-	-0.03 (0.01)	1.4x10^{-5}	-	-0.09 (1.19)	0.94	-	-0.63(0.10)	1.2x10^{-9}	-
Reading a 9 (NBO)	-0.28 (0.09)	0.01	-	-0.03 (0.01)	2.1x10^{-4}	-	1.33 (0.84)	0.14	-	-0.46 (0.10)	6.3x10^{-6}	-
Reading s 9 (NARA II)	-0.29 (0.13)	0.04	-	-0.03 (0.01)	3.0x10^{-5}	-	1.55 (1.16)	0.21	-	-0.51 (0.11)	2.5x10^{-6}	-
Reading a 9 (NARA II)	-0.34 (0.12)	0.01	-	-0.03 (0.01)	3.8x10^{-4}	-	1.57 (1.11)	0.18	-	-0.64 (0.11)	4.1x10^{-9}	-
Reading s 13 (TOWRE)	-0.36 (0.15)	0.03	-	-0.04 (0.01	2.7x10^{-5}	-	2.06 (1.35)	0.15	-	-0.46 (0.12)	1.8x10^{-4}	-
NW reading a 9 (NBO)	-0.37 (0.09)	1.7 x10^{-3}	-	-0.03 (0.01)	2.4x10^{-4}	-	1.54 (0.85)	0.09	-	-0.46 (0.10)	6.1x10^{-6}	-
NW reading s 13 (TOWRE)	-0.21 (0.17)	0.25	-	-0.03 (0.01)	2.4x10^{-4}	-	1.43 (1.58)	0.38	-	-0.45 (0.12)	1.9x10^{-4}	-
Spelling a 7 (NB)	-0.05 (0.12)	0.72	-	-0.05 (0.01)	1.4x10^{-5}	-	-1.12 (1.13)	0.34	-	-0.63 (0.10)	1.2x10^{-9}	-
Spelling a 9 (NB)	-0.23 (0.08)	0.01	-	-0.04 (0.01)	8.7x10^{-7}	-	1.43 (0.72)	0.38	-	-0.45 (0.10)	1.3x10^{-5}	-
PhonAware 7 (AAT)	-0.22 (0.15)	0.16	-	-0.02 (0.01)	1.9x10^{-3}	-	-0.31 (1.33)	0.82	-	-0.67 (0.10)	<1x10^{-10}	-
Listencing c 8 (WOLD)	-0.03 (0.08)	0.75	-	-0.02 (0.01)	0.02	-	-1.47 (0.70)	0.06	-	-0.55 (0.11)	4.5x10^{-7}	-

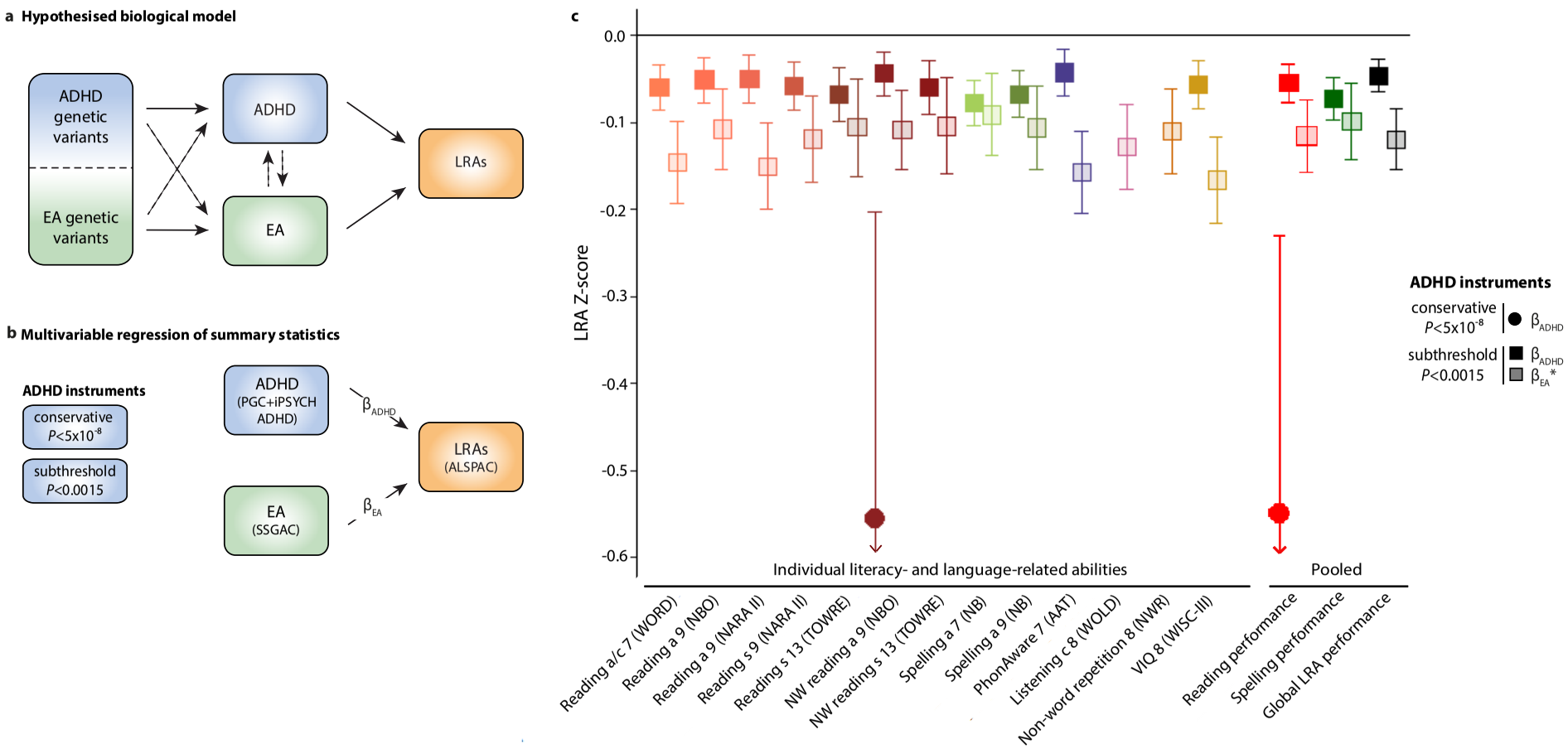
Non-word repetition 8 (NWR)	-0.14 (0.14)	0.34	-	-0.02 (0.01)	0.01	-	-0.16 (1.30)	0.90	-	-0.47 (0.11)	$1.5x10^{-5}$	-
VIQ 8 (WISC-III)	-0.23 (0.15)	0.16	-	-0.03 (0.01)	$5.0x10^{-5}$	-	-0.36 (1.37)	0.80	-	-0.71 (0.11)	$<1x10^{-10}$	-
Pooled reading performance	-0.35 (0.09)	$9.2x10^{-5}$	0.19	-0.03 (0.01)	$1.4x10^{-6}$	0.79	1.67 (0.78)	0.03	0.31	-0.50 (0.09)	$4.9x10^{-8}$	0.09
Pooled spelling performance	-0.18 (0.11)	0.10	0.03	-0.04 (0.01)	$1.1x10^{-8}$	0.28	0.40 (1.25)	0.75	0.001	-0.42 (0.10)	$1.3x10^{-5}$	0.39
Global LRA performance	-0.18 (0.07)	0.01	0.007	-0.03 (0.01)	$1.2x10^{-6}$	0.06	0.29 (0.68)	0.67	0.002	0.51 (0.08)	$<1x10^{-10}$	0.002

1. ADHD genetic effects shared with EA as assessed through EA genetic effect estimates of ADHD-associated variants

Abbreviations: LRAs, literacy- and language-related abilities; ADHD, Attention-Deficit/Hyperactivity Disorder; EA, educational attainment; P_{thr}, *P*-value threshold; *P*, *P*-value ; P_{het}, heterogeneity *P*-value; a, accuracy; c, comprehension; s, speed; WORD, Wechsler Objective Reading Dimension; NBO, Nunes, Bryant and Olson (ALSPAC specific instrument); NARA II, The Neale Analysis of Reading Ability- Second Revised British Edition; TOWRE, Test Of Word Reading Efficiency; NW, non-word; NB, Nunes and Bryant (ALSPAC specific instrument); PhonAware, phonemic awareness, AAT, Auditory Analysis Test; WOLD, Wechsler Objective Language Dimensions; NWR, Nonword Repetition Test; VIQ, verbal intelligence quotient; WISC-III, Wechsler Intelligence Scale for Children III; MVR - Multivariable regression.

Sets of conservative ($P<5x10^{-8}$) and subthreshold ($P<0.0015$) ADHD instruments were extracted from ADHD (PGC+iPSYCH), EA (SSGAC) and LRAs (ALSPAC) GWAS summary statistics. ADHD-specific effects independent of EA (β_{ADHD}) and ADHD effects shared with EA (β_{EA}) on LRAs were estimated with MVRs (Figure 2b). ADHD effects shared with EA were assessed through EA genetic effect estimates of ADHD-associated variants and presented with respect to missing school years. β_{ADHD} quantifies the change in LRA Z-score per log odds increase in ADHD liability. β_{EA} quantifies the change in LRA Z-score per missing year of schooling. Pooled estimates for reading, spelling and global LRAs (Table 1) were obtained through random-effects meta-regression. Evidence for effect heterogeneity (P_{het}) was monitored through Cochran's Q-test. Effects passing the experiment-wide significance threshold ($P<0.007$) are depicted in grey. To compare effect sizes of β_{ADHD} and β_{EA}, MVR was carried out using standardised genetic effect estimates for which results are provided in Figure 2c.


a
Discovery samples
(GWAS summary statistics)
Target sample
(Genotype data)
PGC ADHD
iPSYCH ADHD
PGC+iPSYCH ADHD
ALSPAC
13 LRAs
b
*
*
*
1.5%
1.0%
0.5%
0.0%
Regression R²
PGC ADHD
iPSYCH ADHD
PGC + iPSYCH ADHD
Reading a/c 7 (WORD)
Reading a 9 (NBO)
Reading a 9 (NARA II)
Reading s 9 (NARA II)
Reading s 13 (TOWRE)
NW reading a 9 (NBO)
NW reading s 13 (TOWRE)
Spelling a 7 (NB)
Spelling a 9 (NB)
PhonAware 7 (AAT)
Listening c 8 (WOLD)
Non-word repetition 8 (NWR)
VIQ 8 (WISC-III)


a Hypothesised biological model
ADHD genetic variants
EA genetic variants
ADHD
EA
LRAs
b Multivariable regression of summary statistics
ADHD instruments
conservative P<5x10-8
subthreshold P<0.0015
ADHD (PGC+iPSYCH ADHD)
EA (SSGAC)
βADHD
βEA
LRAs (ALSPAC)
c
LRA Z-score
0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
Individual literacy- and language-related abilities
Pooled
Reading a/c 7 (WORD)
Reading a 9 (NBO)
Reading a 9 (NARA II)
Reading s 9 (NARA II)
Reading s 13 (TOWRE)
NW reading a 9 (NBO)
NW reading s 13 (TOWRE)
Spelling a 7 (NB)
Spelling a 9 (NB)
PhonAware 7 (AAT)
Listening c 8 (WOLD)
Non-word repetition 8 (NWR)
VIQ 8 (WISC-III)
Reading performance
Spelling performance
Global LRA performance
ADHD instruments
conservative P<5x10-8
βADHD
subthreshold P<0.0015
βADHD
βEA*