Title: Disentangling genetic overlap between Attention-Deficit/Hyperactivity Disorder, literacy and language

Short title: Genetic confounding versus trait-specific 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

Background Interpreting polygenic overlap between ADHD and both literacy- and language-related impairments is challenging as genetic confounding might bias associations. Here, we investigate evidence for links between polygenic ADHD risk and multiple literacy- and language-related abilities (LRAs), beyond genetic effects shared with educational attainment (EA).

Methods Thirteen LRAs were studied in UK children (N≤5,919). Genome-wide summary statistics on clinical ADHD and years-of-schooling were obtained from large consortia (N≤326,041). ADHD-LRA genetic overlap was assessed with ADHD-polygenic scores (ADHD-PGS, *P*-threshold<0.1). Using multivariable regressions (MVR), polygenic links were dissected into trait-specific associations and pleiotropic influences shared with EA.

Results ADHD-PGS were inversely associated with all LRAs, explaining ≤1.6% phenotypic variation. These links are, however, potentially confounded as both ADHD (r_g(SE)=-0.53(SE=0.03)) and LRAs ($(r_g(SE)_{min}$=0.56(0.09)) are genetically correlated with EA. Using MVR, ADHD-specific genetic influences on reading, spelling, phonemic awareness and verbal intelligence were detected, whereas associations with listening comprehension and non-word repetition were fully attributable to pleiotropic effects involving EA. Using conservative ADHD-instruments (*P*-threshold<$5x10^{-8}$), pooled reading performance showed the strongest trait-specific association, corresponding to a 0.35 decrease in Z-scores per log odds in ADHD-liability (*P*=$9.2x10^{-5}$), with little evidence for confounding. Using subthreshold ADHD-instruments (*P*-threshold<0.0015), trait-specific associations had lower magnitude, but higher predictive accuracy (e.g. pooled reading: β=-0.03(SE=0.01), *P*=$1.4x10^{-6}$), while EA-related pleiotropic influences of equal strength and, once standardised, at least equal magnitude were present throughout.

Conclusion Genetic confounding distorts genetic overlap with polygenic ADHD risk and conceals an association profile that primarily involves reading-related impairments, but few language-related problems, with implications for intervention targets.

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 are comorbid for reading disability (RD, also known as developmental dyslexia) and vice versa[4], although the spectrum of affected skills in ADHD may 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[15] has been proposed, that may operate intergenerationally[16]. 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[17–19]. However, the interpretation of polygenic overlap is challenging as associations might be inflated and induced by genetic confounders such as genetically predicted educational attainment (EA). EA is a proxy for cognitive abilities and socioeconomic status[20] and is genetically related to both ADHD and reading difficulties[21,22]. Therefore, EA might obscure the genetic relationship between ADHD and LRAs. However, the extent and nature of the genetic overlap between polygenic ADHD risk and LRAs, accounting for genetic confounding, is 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 beyond pleiotropic effects due to EA.

ADHD polygenic scores (ADHD-PGS) are generated using genome-wide association summary statistics of two large independent ADHD samples, the Psychiatric Genomics Consortium (PGC) and the Danish Lundbeck Foundation Initiative for Integrative Psychiatric Research (iPSYCH). Associations of polygenic ADHD risk 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 techniques[23], analogous to Mendelian Randomization approaches, associations between polygenic ADHD risk and LRA measures are subsequently dissected into trait-specific (associations between ADHD and LRAs independent of EA) and pleiotropic (defined throughout the manuscript as associations that are shared across ADHD, EA and LRAs) links. EA-ADHD pleiotropic effects are estimated utilising EA genome-wide summary statistics from the Social Science Genetic Association Consortium (SSGAC).

Methods and Materials

Literacy- and language-related abilities in the general population

LRA measures were assessed in children and adolescents from ALSPAC, a UK population-based longitudinal pregnancy-ascertained birth cohort (estimated birth date: 1991-1992, Supplementary Information)[24,25]. 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 literacy- and language-related abilities related to reading, spelling and phonemic awareness, and language-related abilities 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 (Supplementary Information). Phenotypic trait correlations, using Pearson-correlation coefficients, were comparable using both untransformed and rank-transformed scores (Table S1). To account for the multiple testing burden, we estimated the effective number of phenotypes studied using Matrix Spectral Decomposition[26] (MatSpD), revealing seven independent measures (experiment-wide error rate of 0.007).

For sensitivity analysis, we excluded 188 children with an ADHD diagnosis at age 7, based on the Development and Wellbeing Assessment (DAWBA)[27], and LRA data available (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. ALSPAC 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). Genome-wide association study (GWAS) summary statistics were obtained from an internationally collaborative mega-analysis of clinical ADHD[28], conducted by the Psychiatric Genomics Consortium (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[29] (14,584 ADHD cases, 22,492 controls) (Table 2, Supplementary Information), using samples from the Danish Neonatal Screening Biobank hosted by Statens Serum Institute[22,30].

Combined PGC and iPSYCH ADHD sample (PGC+iPSYCH). To maximise power, we also analysed meta-GWAS summary statistics from a combined ADHD sample containing both PGC and iPSYCH participants[22] (20,183 cases, 35,191 controls/pseudo-controls) (Table 2, Supplementary Information, www.med.unc.edu/pgc/) and its European-only subset (*PGC+iPSYCH(EUR),* 19,099 cases, 34,194 controls/pseudo-controls) (Table 2, Supplementary Information, 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[20] (discovery and replication sample combined, excluding ALSPAC and 23andMe samples, N=326,041) were obtained from the SSGAC consortium with respect

to years of schooling[20] (Table 2, Supplementary Information). 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 calculated using Restricted Maximum Likelihood (REML) analyses[31,32] as implemented in Genome-wide Complex Trait Analysis (GCTA) software[33], including individuals with a genetic relationship of <0.05[31]. All LRAs within this study were selected to be heritable (N>4,000, Table S2).

Linkage Disequilibrium Score regression and correlation

Linkage Disequilibrium Score (LDSC) regression[34] was used to distinguish confounding from polygenic influences in genome-wide analyses by examining the intercept of the LD score regression (Supplementary Information). Unconstrained LD-score correlation[35] analysis was applied to estimate r_g (Supplementary Information).

Polygenic scoring analyses

To estimate the phenotypic variance in LRAs due to ADHD risk increasing alleles, ADHD-PGS[36,37] were created in ALSPAC using PGC, iPSYCH and PGC+iPSYCH summary statistics (Supplementary Information). ADHD-PGS have been previously linked to ADHD symptoms in ALSPAC participants[38]. Rank-transformed traits 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 units of LRA performance per one standard deviation increase in ADHD-PGS.

Multivariable regression analysis

To estimate shared genetic liability of clinical ADHD and LRAs conditional on genetic influences through EA, we applied multivariable regression as implemented in Mendelian Randomization methodologies[23,39].

Genetic variant selection: We selected two sets of genetic variants from ADHD (PGC+iPSYCH) GWAS summary statistics. The first contained genome-wide significant variants ($P<5x10^{-8}$, conservative variant sets). The second included variants passing a more lenient *P*-value threshold ($P<0.0015$, subthreshold variant sets) to increase the power of the analysis, consistent with current guidelines for the selection of genetic instruments (F-statistic<10)[40]. All sets included independent (PLINK[41] clumping: LD-$r^2<0.25$, ±500 kb), well imputed (INFO[42]>0.8) and common (EAF>0.01) variants only. This resulted in 15 conservative and $2{,}689<N_{SNPs}\leq 2{,}692$ subthreshold ADHD instruments (Table S8).

Estimation of ADHD-specific and EA-ADHD pleiotropic effects: To examine the genetic association of ADHD with LRA scores conditional on EA, we extracted sets of ADHD-associated variants (conservative and subthreshold instruments) from ADHD (PGC+iPSYCH), EA (SSGAC) as well as LRA (ALSPAC) GWAS summary statistics (Supplementary Information). Analysing each set of variants independently, genetic effects for individual LRA measures (β) were regressed on both ADHD genetic effects (β as lnOR) and EA genetic effects (β) using an OLS regression framework (R:stats library, Rv3.2.0)[23]. Regression estimates quantify here the change in standard deviation (SD) units of LRA performance per log odds in ADHD risk, or per year of schooling, respectively. EA-related estimates were converted to quantify a change per missing year of schooling, as ADHD risk and EA are inversely genetically related with each other[22].

Finally, regression coefficients were meta-analysed across reading-related abilities (decoding and reading comprehension measures)[14], spelling-related abilities and all LRA measures using random-effects meta-regression, accounting for phenotypic correlations between traits (R:metafor library, Rv3.2.0, Supplementary Information). Pooled trait combinations are detailed in Table 1.To compare the magnitude of specific and pleiotropic regression estimates, analyses were also conducted using fully standardised genetic effects (Supplementary Information).

Sensitivity analyses: To confirm the presence of EA-ADHD pleiotropic effects, we studied the genetic association between EA and LRAs conditional on ADHD. We extracted sets of EA-associated variants (conservative and subthreshold instruments, Table S8) from the same GWAS summary statistics and carried out a similar analysis as described for ADHD instruments above.

Attrition analysis

We carried out an *attrition* analysis in ALSPAC studying the genetic association between LRA-missingness and ADHD-PGS (Supplementary Information). Polygenic links were subsequently dissected into trait-specific and EA-ADHD pleiotropic influences as described above (Supplementary Information).

Results

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

All studied 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, are heritable traits 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). All LRAs were both at the phenotypic (Table S1) and genetic (Table S3) level moderately to strongly interrelated. Likewise, there was LDSC-based evidence for genetic liability of clinical ADHD within the PGC (LDSC-h^2=0.08(SE=0.03)) and the iPSYCH (LDSC-h^2=0.26(SE=0.02)) samples and combinations thereof (Table S4), as previously reported[22].

ADHD polygenic risk and literacy- and language-related abilities

We observed consistent evidence for an inverse genetic association between ADHD polygenic risk and performance in reading accuracy/comprehension age 7 (PGC: OLS-R^2=0.1%, P=4.6x10^{-3}; iPSYCH: OLS-R^2=1.0%, P<1x10^{-10}; PGC+IPSYCH: OLS-R^2=1.3%, 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} , PGC+IPSYCH: OLS-R^2=1.5%, 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}, PGC+IPSYCH: OLS-R^2=1.1%, P<1x10^{-10}) using independent ADHD samples (Figure 1, Table S5). Findings did not differ from results based on participants of European ancestry only (PGC+iPSYCH(EUR), Table S5), nor were they affected by the exclusion of children with an ADHD diagnosis (Table S6). Note that ADHD polygenic scores based on iPSYCH and PGC+iPSYCH, but not PGC samples, were associated with all of the LRA measures studied, with the strongest effects being observed using the PGC+iPSYCH sample (Table S5).

Figure 1 here

Shared genetic liability with EA

We studied the genetic overlap between ADHD and EA (LDSC-h^2= 0.11(SE=0.004)) and observed strong evidence for a moderate negative genetic correlation (PGC+iPSYCH, r_g=-0.53(SE=0.03), $P<1x10^{-10}$) , consistent with previous findings[22]. 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[21]. More importantly, we observed evidence for the presence of EA-ADHD pleiotropic effects (i.e. single genetic loci influencing multiple traits[43]) as two independent variants are associated with both ADHD[22] and EA[20] at the genome-wide significant level (Supplementary Information). These findings indicate complex cross-trait relationships that are potentially reciprocal (Figure 2a).

Figure 2 here

Multivariable regression analyses

To disentangle the polygenetic overlap of ADHD risk with literacy- and language-related measures into trait-specific associations (between ADHD and LRAs independent of EA) in contrast to pleiotropic links (assessed as genetic effects shared between ADHD, LRAs and EA), we applied multivariable regression analyses[23] using ADHD instruments (Figure 2b). As sensitivity analyses, we confirmed the presence of shared EA-ADHD effects using EA instruments (Figure S1).

Using conservative ADHD instruments (Table S8), a trait-specific association of 0.37 SD units decrease per log odds increase in ADHD risk (β_{ADHD}=-0.37(SE=0.09), $P=1.7x10^{-3}$) was observed for non-word reading accuracy at age 9, but not for other LRAs (Table 3). There was little evidence for pleiotropic effects (Table 3), which was confirmed using conservative EA instruments (Table S9).

Subthreshold ADHD instruments (Table S8) captured trait-specific associations between polygenic ADHD and all LRAs assessed, except for listening comprehension and non-word repetition

both at age 8 (Table 3). Trait-specific links were, however, considerably smaller than those captured by conservative ADHD instruments, with a 0.03 SD unit decrease in non-word reading accuracy age 9 per log odds increase in ADHD risk, although the statistical evidence increased (β_{ADHD}=-0.03(SE=0.01), P=2.4x10^{-4}, Table 3). In addition, there was evidence for a shared polygenetic EA-ADHD effect on literacy- and language-related performance for all measures studied (Table 3). For example, as there is an inverse genetic correlation between ADHD and years of schooling[22], this translates into a further 0.46 SD units decrease in non-word reading accuracy at age 9 per missing school year ($\beta_{EA\text{-}ADHD}$=-0.46(SE=0.10), P=6.1x10^{-6}, Table 3). The association between polygenic ADHD risk and both listening comprehension and non-word repetition was fully attributable to shared genetic effects with EA (Table 3). Overall, EA-ADHD pleiotropic effects were confirmed using subthreshold EA instruments (Table S9).

To investigate whether trait-specific links between polygenic ADHD risk and both literacy- and language-related abilities were consistent across LRA measures, estimates were meta-analysed across pre-defined domains (Table 1). Using conservative ADHD instruments (Table S8), a trait-specific association was only observed between polygenic ADHD risk and pooled reading performance (β_{ADHD}=-0.35(SE=0.09), P=9.2x10^{-5}), with little evidence for heterogeneity, (P_{het}=0.19, Table 3). Trait-specific associations for pooled spelling performance and global LRA performance were not significant and both showed evidence for heterogeneity (Table 3). Consistent with results for individual LRA measures, trait-specific and EA-ADHD pleiotropic influences were observed for pooled reading, pooled spelling and global LRA performance using subthreshold ADHD instruments, with little evidence for heterogeneity (Table 3).

Finally, to compare the magnitude of trait-specific and EA-ADHD pleiotropic multivariable regression estimates, effect estimates were standardised (Supplementary Information, Figure 2c). This showed that shared EA-ADHD effects, were as large, or even larger, than ADHD-specific effects (Figure 2c, Table S10), when assessing ADHD-LRA associations with subthreshold ADHD instruments.

EA-ADHD pleiotropic effects estimated from either ADHD or EA subthreshold variants were largely consistent with each other (Table S9, Table S10).

All multivariable regressions were performed using constrained intercepts[23] (Supplementary Information). There was little evidence supporting the inclusion of regression intercepts that would imply additional, not yet captured, pleiotropic effects.

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[44]. Dissecting this polygenic link using multivariable regression (Table S13) showed a 0.05 increase in log odds sample drop-out per log odds increase in liability to ADHD (ADHD-specific effect: log(OR)=0.05(SE=0.01), P=3.7x10^{-4}) and a further 1.04 increase in log odds sample drop-out per missing year of schooling (Pleiotropoic EA-ADHD effect: log(OR)=1.04(SE=0.19),P=7.3x10^{-8}), when analysing ADHD subthreshold variants. Neither effect was detectable using conservative ADHD instruments.

Discussion

This study identified strong evidence for an inverse association between polygenic ADHD risk and several population-based literacy- and language-related abilities, using multiple discovery samples. We confirmed that these polygenic links are susceptible to confounding due to strong genetic correlations between EA and both polygenic ADHD risk and LRAs. Accurate modelling of polygenic links showed that there are trait-specific associations between polygenic ADHD risk and reading, spelling, phonemic awareness and verbal intelligence scores, but not listening comprehension and non-word repetition. However, the detectable genetic overlap was strongly inflated or even induced by genetic confounding[45], when investigated with subthreshold ADHD instruments.

Disentangling the underlying aetiological relationships between ADHD, EA and LRAs is challenging as the observed genetic interrelationships are complex, implicating reciprocal processes that prevent causal modelling. Specifically, processes of vertical pleiotropy (i.e. ADHD risk variants impacting on LRA measures through their effect on EA) are indistinguishable from processes of horizontal pleiotropy (i.e. shared risk-associated variants directly influencing both ADHD and EA), thus violating classical Mendelian Randomization assumptions[23]. In other words, individual ADHD GWAS signals ($P \leq 5 \times 10^{-8}$) are not valid instruments to infer causality between ADHD and LRAs, due to cross-trait-associations with EA at the genome-wide significant level, and modelling approaches therefore need to allow for the presence of pleiotropic influences. In addition, exposure (polygenic ADHD risk) and outcome (literacy- and language-related measures) were not observed within the same set of individuals. However, this does not preclude the dissection of genetic relationships between ADHD and LRAs measures into trait-specific and pleiotropic links, while controlling for collider bias[46]. This has important implications for current research into ADHD-specific literacy- and language-related impairments.

Firstly, the detection of trait-specific versus pleiotropic polygenic liabilities was strongly governed by the choice of genetic variants. Conservative ADHD instruments identified large trait-

specific associations between ADHD and pooled reading abilities with little evidence for genetic confounding, but had limited power[47]. These ADHD instruments comprised only 15 independent SNPs including 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)[48]. Subthreshold variants, on the other hand, captured trait-specific associations between ADHD and most LRAs with high predictive accuracy and lower magnitude. However, even though small, the existence of trait-specific associations between ADHD and multiple LRAs provides evidence for genetic mechanisms underlying the comorbidity between ADHD and literacy- and language-related impairments, affecting multiple facets[1–3,5–10]. It could be speculated whether reading-related deficits in ADHD, as hypothesised for RD, arise from a phonological impairment[49,50], which affects decoding and reading abilities[51]. Although phonological processing is important for developing reading abilities[52,53], it is possible that phonemic awareness skill is a reflection of decoding ability at this age, as reading abilities, once developed, also affect phonological skills[54]. Problems with phonological short term memory[55] and listening comprehension[56] have also been reported in children with ADHD , but our results do not support the existence of trait-specific genetic links .

Secondly, subthreshold variants captured, irrespective of the studied LRA, strong EA-ADHD pleiotropic effects. This directly supports a multiple-deficit model proposing shared underlying aetiologies between ADHD and RD dimensions[15,16], through their genetic links with EA. It is, for example, possible that ADHD symptoms may impact on classroom behavior, effectively reducing children's ability to follow instructions that in turn impairs their acquisition of skills for learning (e.g. reading)[10,57,58]. Alternatively, consistent with an intergenerational multiple-deficit model proposed for reading disability[16], children growing up in disadvantaged environments, that is genetically predictable through parental polygenic EA scores[59], might be more vulnerable to psychiatric illness such as ADHD[60] that may in turn also affect subsequent literacy- and language-related outcomes. Furthermore, interplay between ADHD and EA is conceivable where adolescents with ADHD are more likely to leave school at an earlier age and pass on an increased genetic load for ADHD to their own

children[61]. These processes may involve cognitive mechanisms such as processing speed[62] that have been linked to both EA[63] and shared genetic influences between ADHD and RD[64]. Thus, our findings support the presence of ADHD-specific reading-related deficits, in need of targeted interventions, against a background of overarching cognitive but unspecific trait interrelationships (that might be improved through general training programmes aiming at schooling outcomes[65]).

In a general context, the findings of this study are in support of an omnigenic[66] model underlying complex traits. This implies that only the largest-effect variants will reflect ADHD specificity, and may thus tag the most trait-specific associations. The majority of variants, however, will capture pleiotropic (omnigenic) influences without trait-specific functions, pointing to highly interconnected neural networks[66] that give rise to genetic confounding. This implies that the proportion of ADHD-specific variants among selected conservative instruments is high, whereas the ratio of these variants among selected subthreshold instruments is predictably lower. Consequently, the majority of subthreshold variants are likely to represent highly powerful cross-trait genetic predictors, without trait-specific function, that enhance and distort genetic overlap. This interpretation is also compatible with a general factor model of psychopathology[67], which has been extended to the genetic architecture of ADHD[68].

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. Within our study, EA-ADHD pleiotropic effects were at least as large, or even larger, than trait-specific associations between ADHD and LRAs when using subthreshold instruments. This inflates or even induces genetic overlap, as for example between polygenic ADHD risk and both listening comprehension and non-word repetition. Thus, predictive accuracy between phenotypes does not necessarily relate to trait-specific associations, and PGS analyses cannot be interpreted in terms of underlying mechanisms[69]. 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[70,71]. Our analyses suggest that genetic overlap between phenotypes needs to be accounted for genetic confounding by dissecting genetic overlap into trait-specific and pleiotropic influences.

A limitation of our study is that ALSPAC, as other cohort studies, suffers from attrition[44,72]. Sensitivity analyses showed that this is unlikely to bias our findings based on conservative instruments. In contrast, trait-specific and pleiotropic 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[44]. We are also aware that the strength of the genetic overlap between polygenic ADHD risk and literacy- and language-related abilities may vary according to ADHD symptom domain levels, implicating especially inattentiveness[73], as well as the nature of the trait 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.

Larger and more detailed clinical samples as well as extensive multivariate variance analyses of the spectrum of literacy- and language-related abilities (that are currently computationally expensive[74]) will help to discover and characterize further overlap between ADHD and literacy- and language-related cognitive processes.

Conclusion

There is an ADHD-specific spectrum of literacy- and language-related impairments, in particular related to reading performance involving decoding and comprehension abilities. However, the association between polygenic ADHD risk and literacy- and language-related measures is prone to genetic confounding and ADHD-specific influences can only be identified using accurate dissections of polygenic links.

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. van Bergen E, van der Leij A, de Jong PF. The intergenerational multiple deficit model and the case of dyslexia. *Front Hum Neurosci*. 2014;8. doi:10.3389/fnhum.2014.00346

17. 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

18. 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

19. 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

20. 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

21. 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

22. 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

23. 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

24. 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

25. 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

26. 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

27. 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

28. 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

29. 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

30. 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

31. 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

32. 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

33. 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

34. 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

35. 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

36. 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

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

38. 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

39. Burgess S, Thompson SG. Multivariable Mendelian Randomization: The Use of Pleiotropic Genetic Variants to Estimate Causal Effects. *Am J Epidemiol*. 2015;181(4):251-260. doi:10.1093/aje/kwu283

40. 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

41. 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.

42. 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

43. 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

44. 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

45. Verbanck M, Chen C-Y, Neale B, Do R. Widespread pleiotropy confounds causal relationships between complex traits and diseases inferred from Mendelian randomization. *bioRxiv*. June 2017:157552. doi:10.1101/157552

46. 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

47. 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

48. 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

49. 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

50. *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.

51. 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

52. Adams MJ. *Beginning to Read: Thinking and Learning about Print*. MIT Press; 1994.

53. Rieben L, Perfetti CA, eds. *Learning To Read: Basic Research and Its Implications*. 1 edition. Place of publication not identified: Routledge; 2016.

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

55. Bolden J, Rapport MD, Raiker JS, Sarver DE, Kofler MJ. Understanding Phonological Memory Deficits in Boys with Attention-Deficit/Hyperactivity Disorder (ADHD): Dissociation of Short-term Storage and Articulatory Rehearsal Processes. *J Abnorm Child Psychol*. 2012;40(6):999-1011. doi:10.1007/s10802-012-9619-6

56. McInnes A, Humphries T, Hogg-Johnson S, Tannock R. Listening comprehension and working memory are impaired in attention-deficit hyperactivity disorder irrespective of language impairment. *J Abnorm Child Psychol*. 2003;31(4):427-443.

57. Rapport MD, Scanlan SW, Denney CB. Attention-deficit/hyperactivity disorder and scholastic achievement: a model of dual developmental pathways. *J Child Psychol Psychiatry*. 1999;40(8):1169-1183.

58. 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

59. Ayorech Z, Krapohl E, Plomin R, Stumm S von. Genetic Influence on Intergenerational Educational Attainment. *Psychol Sci*. 2017;28(9):1302-1310. doi:10.1177/0956797617707270

60. Reiss F. Socioeconomic inequalities and mental health problems in children and adolescents: a systematic review. *Soc Sci Med 1982*. 2013;90:24-31. doi:10.1016/j.socscimed.2013.04.026

61. Russell G, Ford T, Rosenberg R, Kelly S. The association of attention deficit hyperactivity disorder with socioeconomic disadvantage: alternative explanations and evidence. *J Child Psychol Psychiatry*. 2014;55(5):436-445. doi:10.1111/jcpp.12170

62. McGrath LM, Pennington BF, Shanahan MA, et al. A multiple deficit model of Reading Disability and Attention-Deficit/Hyperactivity Disorder: Searching for shared cognitive deficits. *J Child Psychol Psychiatry*. 2011;52(5):547-557. doi:10.1111/j.1469-7610.2010.02346.x

63. Ihle A, Oris M, Fagot D, et al. Associations of educational attainment and cognitive level of job with old age verbal ability and processing speed: The mediating role of chronic diseases. *Appl Neuropsychol Adult*. April 2017:1-7. doi:10.1080/23279095.2017.1306525

64. Willcutt EG, Betjemann RS, McGrath LM, et al. Etiology And Neuropsychology Of Comorbidity Between RD And ADHD: The Case For Multiple-Deficit Models. *Cortex J Devoted Study Nerv Syst Behav*. 2010;46(10):1345-1361. doi:10.1016/j.cortex.2010.06.009

65. Daley D, Birchwood J. ADHD and academic performance: why does ADHD impact on academic performance and what can be done to support ADHD children in the classroom? *Child Care Health Dev*. 2010;36(4):455-464. doi:10.1111/j.1365-2214.2009.01046.x

66. 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

67. 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

68. 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)

69. Krapohl E, Patel H, Newhouse S, et al. Multi-polygenic score approach to trait prediction. *Mol Psychiatry*. August 2017:mp2017163. doi:10.1038/mp.2017.163

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

71. 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

72. 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

73. 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.

74. 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

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

76. 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

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

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

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

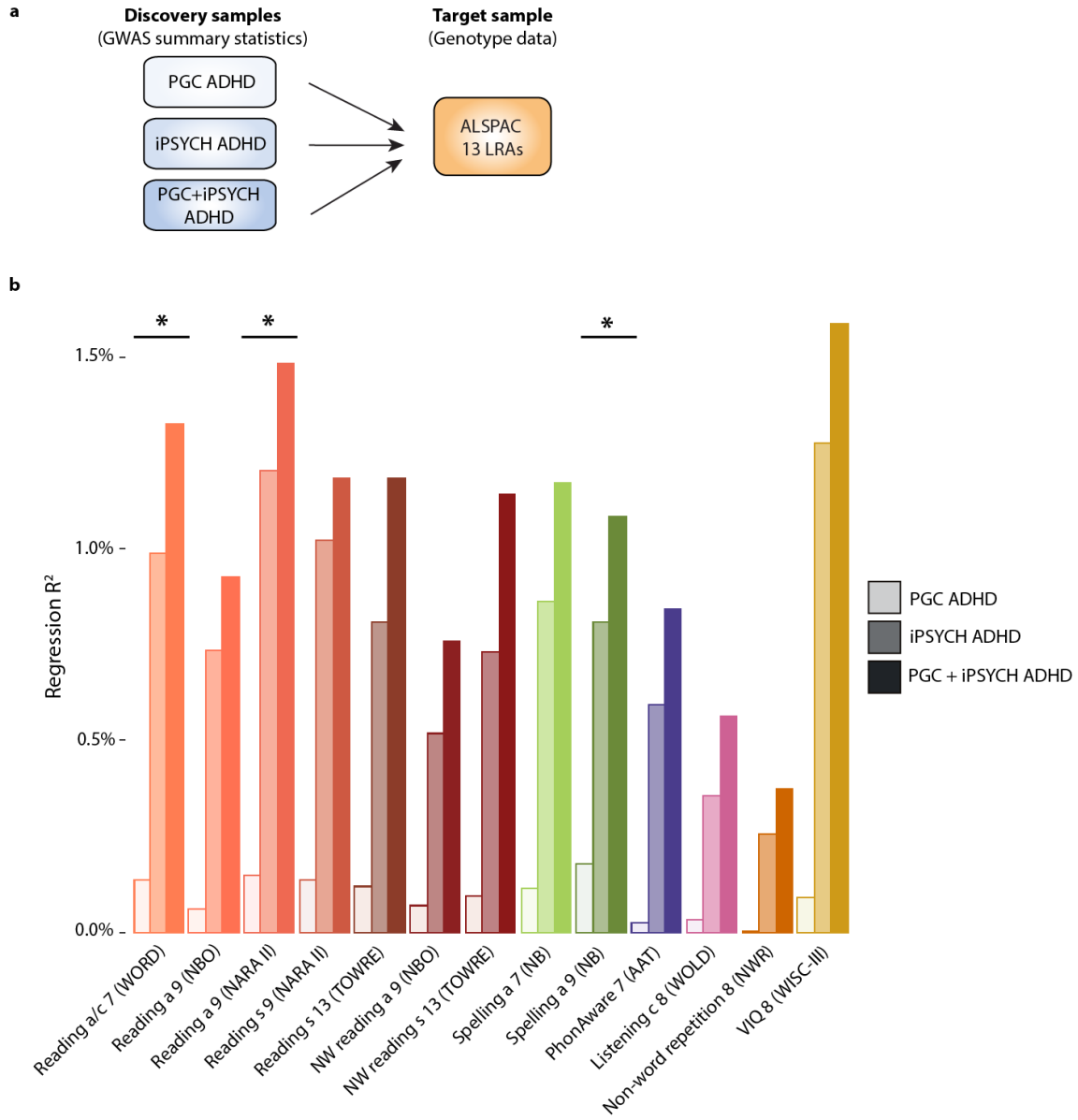
80. Rust J, Wechsler D, Psychological Corporation. *WOLD: Wechsler Objective Language Dimensions.* London; [Hong Kong: Psychological Corp. ; Artsberg Enterprises Ltd. [distributor; 1996.

81. 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

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

Figure legends

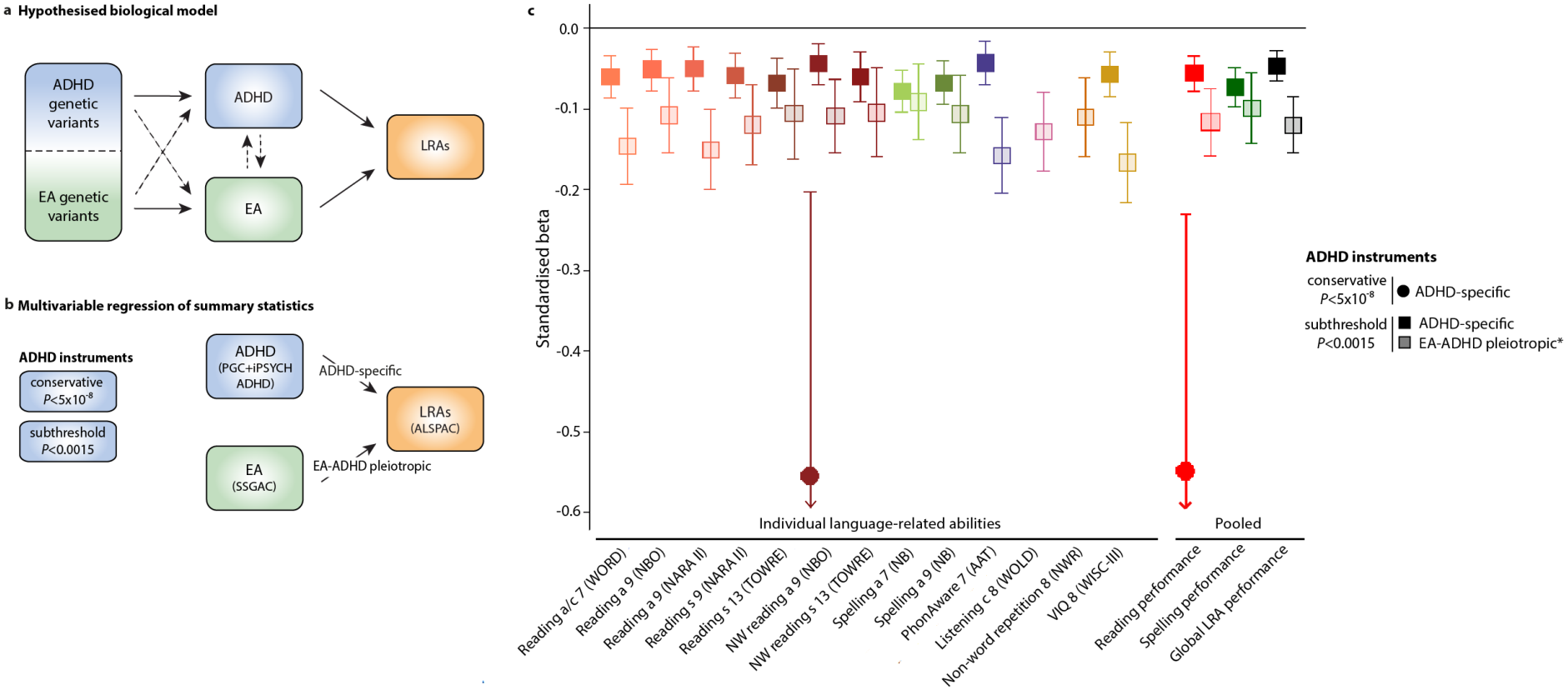
Figure 1: Phenotypic variance in literacy- and language-related abilities explained by polygenic ADHD risk



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.

* - 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: ADHD-specific and EA-ADHD pleiotropic effects of polygenic ADHD risk on literacy- and language-related abilities



Abbreviations: ADHD-assoc., Attention-Deficit/Hyperactivity Disorder-associated; EA-assoc., educational attainment-associated; LRAs, Literacy and literacy- and language-related oral language abilities

(a) Hypothesized biological model of genetic relationships between ADHD, EA, and LRAs, including complex, pleiotropic and reciprocal genetic links (b) multivariable regression model assessing trait-specific and EA-ADHD pleiotropic influences on literacy- and language-related abilities, (c) multivariable regression analyses of LRAs based on 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. To compare the magnitude of ADHD and EA-related multivariable regression estimates, analyses were conducted using standardised genetic effects (Supplementary Methods). Analyses using raw estimates are shown in Table 3. Trait-specific (β_{ADHD}) and shared EA-ADHD ($\beta_{EA-ADHD}$) genetic effects on literacy- and language-related abilities were estimated with multivariable regressions. Pooled estimates for reading, spelling and global LRA measures (Table 1) were obtained through random-effects meta-regression. Pleiotropic EA-ADHD effects are presented with respect to missing school years. Only effects passing the experiment-wide significance threshold ($P<0.007$) are shown with corresponding 95% confidence intervals.

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[75]), 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[76]), words	7.55 (2.44)	9.87 (0.32)	5,738 (49.3)		
Reading speed[a] (NARA II[77]), passages	105.50 (12.47)	9.88 (0.32)	5,189 (49.1)		
Reading accuracy[a] (NARA II[77]), passages	104.11 (13.62)	9.88 (0.32)	5,201 (49.1)		
Reading speed (TOWRE[78]), words	82.58 (10.28)	13.83 (0.20)	4,247 (48.4)		
Non-word reading accuracy (ALSPAC specific: NBO[76])	5.24 (2.48)	9.87 (0.32)	5,731 (49.2)		
Non-word reading speed (TOWRE[78])	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[79])	20.23 (9.51)	7.53 (0.31)	5,919 (50.6)		
Listening comprehension (WOLD[80])	7.50 (1.96)	8.63 (0.30)	5,473 (49.9)		
Non-word repetition (NWR[81])	7.26 (2.51)	8.63 (0.30)	5,464 (49.9)		
Verbal intelligence[a] (WISC-III[82])	107.85 (16.74)	8.64 (0.31)	5,456 (49.7)		

a - 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 assessments 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 assessments 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 of reading- and spelling-related abilities, as well as phonemic awareness, listening comprehension, non-word repetition and verbal intelligence scores 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[20]

Abbreviations: LRAs, literacy- and language-related oral language 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 analyses using ADHD instruments (raw estimates)

LRA	Trait-specific (ADHD)						EA-ADHD pleiotropic (EA)[b]					
	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

b - EA genetic effects as captured by 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}, *P*-value indicating heterogeneity among individual traits; 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

Sets of conservative ($P<5x10^{-8}$) and subthreshold ($P<0.0015$) ADHD instruments were extracted from ADHD (PGC+iPSYCH), EA (SSGAC) and literacy- and language-related (ALSPAC) GWAS summary statistics. Trait-specific (β_{ADHD}) and shared EA-ADHD ($\beta_{EA\text{-}ADHD}$) genetic effects on LRAs were estimated with multivariable regressions. To compare the magnitude of ADHD and EA-related multivariable regression estimates, analyses were conducted using standardised genetic effects (Figure 2c, Table S10). Pooled estimates were obtained through random-effects meta-regression, monitoring evidence for heterogeneity through Cochran's Q-test. Pleiotropic EA-ADHD effects are presented with respect to missing school years. Effects passing the experiment-wide significance threshold ($P<0.007$) are depicted in grey.