E-Risk Study Concept Paper template

Provisional Paper Title: The role of mothers' speech in the intergenerational transmission of educational attainment	
Proposing Author: Sophie von Stumm	
Author's Email: sophie.vonstumm @ york.ac.uk	
Academic supervisor: Click or tap here to enter text. (if the proposing author is a student)	
E-Risk Sponsor: Helen Fisher (if the proposing author is not an E-Risk co-investigator)	
Today's Date: 2/20/2024	
Please indicate if you will require an E-Risk independent reproducibility check: ☐	

Please describe your proposal in 2-3 pages with sufficient detail for helpful review.

Background & objective of the study:

Family background continues to be the strongest predictor of children's educational outcomes, in Britain as well as in other putatively democratic countries (Chmielewski, 2019; von Stumm et al., 2022). The intergenerational transmission of educational attainment occurs through genetic and environmental pathways. Disentangling these pathways is key to identifying suitable targets for interventions that effectively disrupt the perpetuation of family background inequality in education.

Mothers' language – that is, the quality and quantity of her speech -- has been identified as an important factor in the intergenerational transmission of education (Hoff, 2003). A generally accepted model is that mothers' speech defines the language environments that children are exposed to in their family homes, and that this exposure is the foundation upon which children develop their own language skills. Children who start school with better language skills have an easier time to follow the teachers' instructions and meet the demands of the curriculum than children with poorer language skills. Over time, children's' discrepancy in language skills, which are the consequence of being exposed to different language environments in line with their mothers' educational attainment, translates into achieving different levels of education themselves, such as completing vocational training versus a university degree (von Stumm et al., 2020; Hart & Risley, 1995). This model is widely known as the 'language gap' that is a gap in language learning opportunities between children from more resourced as compared to under-resourced family homes (Johnson, 2015; Hart & Risley, 1995).

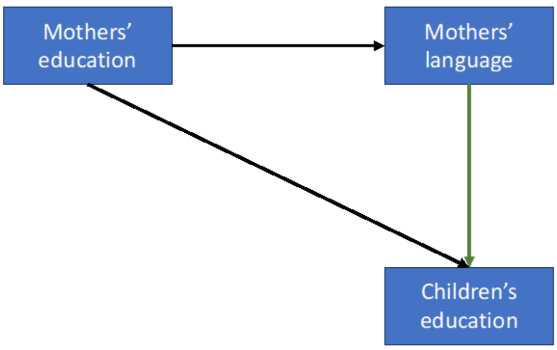

Figure 1

Previous studies in this area have almost exclusively relied on phenotypic measures and tested mediation models, akin to the one displayed in Figure 1. These models test the assumption that part of the association between mothers' and children's education is mediated by the type of language that children experience from their mothers, without discerning genetic and environmental influences. In the proposed study, we will use a polygenic score approach in structural equation models that allow tracing environmental and genetic transmission paths, as well as testing the role of mediating factors on these paths. We will use multi-polygenic scores derived from GWAS for years spent in education (Lee et al., 2018), reading (Eising et al., 2022), and dyslexia (Doust et al., 2022) to capture as many genetic variants as possible that are associated with education and language in both mothers and children. Herein, we use the terms mothers' and children's genotypes to refer to mothers' and children's multi-polygenic scores derived from the three GWAS listed above.

Figure 2 depicts the models that we are proposing to test. Model (a) specifies direct genetic predictions from mothers' genotype to mothers' and children's educational attainment (phenotype), shown by black paths. The green path reflects the extent to which mothers' education phenotype mediates the effect of their genotype onto children's educational attainment (model a). In model (b), children's genotypes have been added; the path from mothers' genotype to that of children's reflects the direct genetic transmission from mothers to children. The blue path from mothers' genotype to children's education reflects genetic nurture effects; that is, genetic influences from the mother on children's education that are environmentally mediated and not directly, genetically transmitted. The extent to which adding children's genotype attenuates the link between mothers' and children's education reflects passive gene-environment correlation between mothers' and children's education (model c). Environmental mediators of the genetic nurture effect in our study include mothers' education (model c/ d, green path) and mothers' language (model d, orange path). Here, we will also test the extent to which mothers' language mediates the association between mothers' and children's education (brown path; see also Figure 1). This modelling step allows drawing parallels between mediators of genetic nurture and of nurture effects.

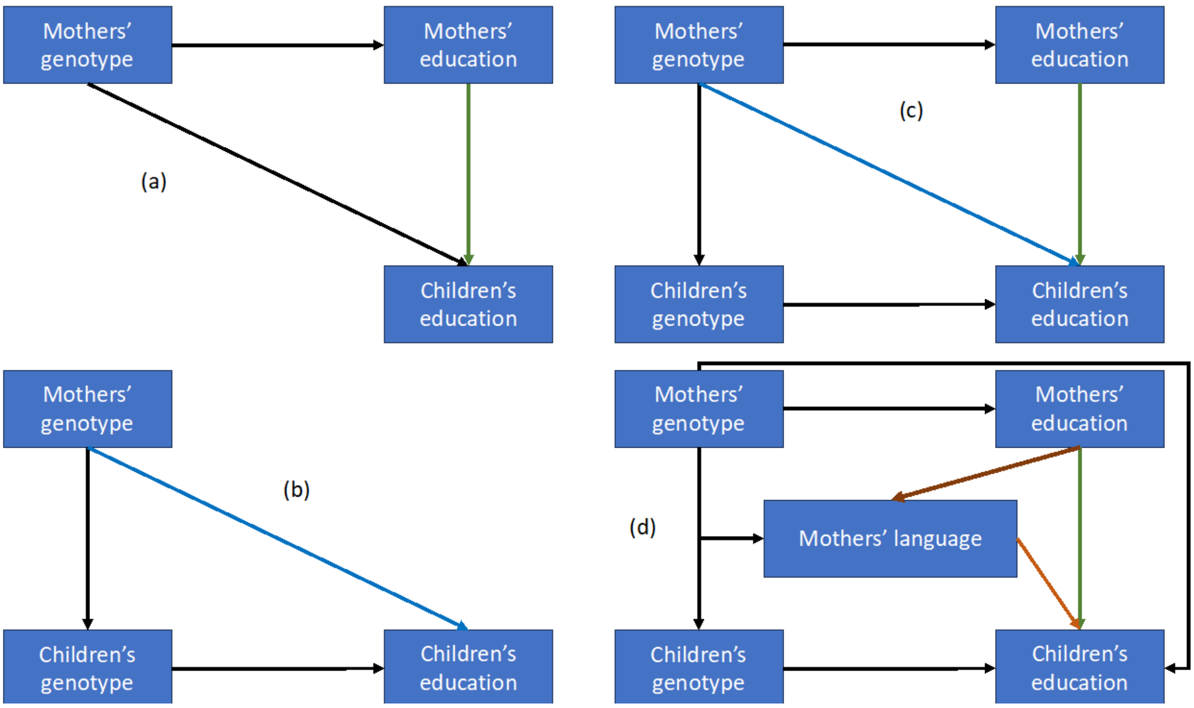

Figure 2: Modelling approach

We predict that mothers' genotype will positively predict their phenotype at medium effect size. We specified this association in the direction from genotype –> phenotype because the reverse is not plausible (one's phenotypic education cannot change inherited DNA differences that are fixed from conception). We also predict that mothers' geno- and phenotypes will positively predict children's educational attainment, with greater effect sizes for the pheno- than the genotype association. Again, path direction was specified because reverse effects are implausible, as well as because of the temporal order of data (i.e., mothers' education follows children's). We predict significant genetic nurture (i.e., significant blue path in model b) and genetic confounding (i.e., attenuated effect of mothers' education after adding children's genotypes, model c). We expect that mothers' language mediates both the genetic nurture effect and the nurture path from mothers' to children's education. For the latter, the path direction was specified to allow comparisons between mediation effects, but the direction of influence is speculative. Depending on these last model paths' effect sizes, we can draw conclusions about the likelihood that contributions of mothers' language to the transmission of family background inequality are causal, and about the extent to which these contributions may be genetic in origin.

Significance of the study (for theory, research methods or clinical practice):

This paper tackles a topic of ongoing public and policy debate and concern: Why do children from families where parents achieved fewer educational qualifications attain less education themselves compared to children whose parents are better educated, and can anything be done to help these children overcome the educational disadvantages associated with their family background? Furthermore, the paper expands on previous works on (a) using genomic and phenotypic measures of education to predict children's educational outcomes (von Stumm et al., 2020), and (b) genetic nurture effects in educational attainment (Wertz et al., 2019), applying a structural equation modelling approach. This is novel in the way that previous studies either tested direct predictions of phenotypic outcomes from polygenic scores and demographic measures, or tested for genetic nurture effects, but not both at the same time.

Data analysis methods:

Models are described in detail above. We will fit them using lavaan in R, clustering the Standard Error at the family level. All variables are directly observed. We will use FIML under the assumption that data are missing at random. We will build composites from the polygenic scores of the 3 GWAS (see below) to reflect genotypes.

Variables needed and at which ages:

Age 5:

Variable	Description
FAMILYID	Unique family identifier
ATWINID	Twin A ID
BTWINID	Twin B ID
SAMPSEX	Sex of Twins: In sample
ZYGOSITY	Zygosity
TORDER	True Twin Order
RORDERP5	Random Twin Order – P5
MAINLANG	Main Language Spoken to Twins
SESW	Social Class Composite (Continuous Version)
MAGEM5	Mother's age at P5 assessment
Obtained from FMSS recordings	Mother's Vocabulary sophistication
HIEDM57	Highest educational qualification (mother) P5-P7 Combined

Age 18:

Variable	Description
TAGEE18	Age at Interview – P18
EDUCACHVE18 EDUCACHVY18	Educational Attainment children (elder/younger) – P18

Polygenic Scores:

Variable	Description
EA3PGS_twins_Jan2019e EA3PGS_twins_Jan2019y	Educ years Lee et al 2018 Nature genetics – Elder/Younger + this PGS in mothers
	Polygenic score for reading (Eising et al., 2022) For: Elder twin, younger twin, and mother
	Polygenic score for dyslexia (Doust et al., 2022) For: Elder twin, younger twin, and mother
	Covariates for PGS (e.g., first 10 PCs, chiptype, batch number + additional control variables if applicable)

Note, for the polygenic scores from GWAS for years spent in education (Lee et al., 2018), reading (Eising et al., 2022), and dyslexia (Doust et al., 2022), the ones for years spent in education are already available in E-Risk. For the Eising one, summary statistics are publicly available. For the Doust one, I applied for data access to 23&me on 02/02/2024 (application form surprisingly short though). If acceptance of this application does not come through, we can use the sum stats for the 10,000 top hits, which are publicly available.

References cited:

Chmielewski, A. K. (2019). The Global Increase in the Socioeconomic Achievement Gap, 1964 to 2015. American Sociological Review, 84(3), 517-544. https://doi.org/10.1177/0003122419847165

Doust, C., Fontanillas, P., Eising, E. et al. Discovery of 42 genome-wide significant loci associated with dyslexia. Nat Genet 54, 1621–1629 (2022). https://doi.org/10.1038/s41588-022-01192-y

Eising, E., Mirza-Schreiber, N., de Zeeuw, E. L., Wang, C. A., Truong, D. T., Allegrini, A. G., Shapland, C. Y., Zhu, G., Wigg, K. G., Gerritse, M. L., Molz, B., Alagöz, G., Gialluisi, A., Abbondanza, F., Rimfeld, K., van Donkelaar, M., Liao, Z., Jansen, P. R., Andlauer, T. F. M., Bates, T. C., … Fisher, S. E. (2022). Genome-wide analyses of individual differences in quantitatively assessed reading- and language-related skills in up to 34,000 people. *Proceedings of the National Academy of Sciences of the United States of America*, *119*(35), e2202764119. https://doi.org/10.1073/pnas.2202764119

Hart, B., & Risley, T. R. (1995). Meaningful differences in the everyday experience of young American children. Baltimore, MD: Paul H. Brookes.

Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects early vocabulary development via maternal speech. Child Development, 74, 1368–1378.

Johnson, E.J. (2015), "Debunking the “language gap”", *Journal for Multicultural Education*, Vol. 9 No. 1, pp. 42-50. https://doi.org/10.1108/JME-12-2014-0044

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., … Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nature Genetics, 50, 1112–1121. https://doi.org/10.1038/s41588-018-0147-3

von Stumm, S., Cave, S. N., & Wakeling, P. (2022). Persistent association between family socioeconomic status and primary school performance in Britain over 95 years. *NPJ science of learning*, *7*(1), 4. https://doi.org/10.1038/s41539-022-00120-3 on Stumm, S., Rimfeld, K., Dale, P. S., & Plomin, R. (2020). Preschool verbal and nonverbal ability mediate the association between socioeconomic status and school performance. Child Development, 91(3), 705–714. https://doi.org/10.1111/cdev.13364

von Stumm, S., Smith-Woolley, E., Ayorech, Z., McMillan, A., Rimfeld, K., Dale, P. S., & Plomin, R. (2020). Predicting educational achievement from genomic measures and socioeconomic status. *Developmental science*, *23*(3), e12925. https://doi.org/10.1111/desc.12925

Wertz, J., Moffitt, T. E., Agnew-Blais, J., Arseneault, L., Belsky, D. W., Corcoran, D. L., Houts, R., Matthews, T., Prinz, J. A., Richmond-Rakerd, L. S., Sugden, K., Williams, B., & Caspi, A. (2020). Using DNA From Mothers and Children to Study Parental Investment in Children's Educational Attainment. Child development, 91(5), 1745–1761. https://doi.org/10.1111/cdev.13329