E-Risk Study Concept Paper template

Provisional Paper Title: The transmission of family background inequality in education across Britain and Germany
Proposing Author: Dr Alexandra Starr, Prof Sophie von Stumm
Author's Email: Alexandra.starr@york.ac.uk; sophie.vonstumm@york.ac.uk
Academic supervisor: NA (if the proposing author is a student)
E-Risk Sponsor: Prof Helen Fisher (if the proposing author is not an E-Risk co-investigator)
Today's Date: 8/6/2023
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:

Children's differences in the ability to learn are evident even before their first day of school and have pervasive, long-term influence on their life chances, health, and well-being. These differences are largely due to family background inequality with influences on learning ability and educational achievement emerging already at an early stage in development (Hair et al., 2015; Nobel et al., 2015; Schwab & Lee-Williams, 2016) and manifesting over the course of children's school career (von Stumm et al., 2015; von Stumm, 2017). By the end of compulsory schooling, children's family background is associated with a two grades difference (von Stumm, 2017) – a gap that translates into a £200,000 disparity in lifetime earnings, among other things (Hodge et al., 2021; Huang, 2015; French et al., 2015). While disrupting the transmission of family background inequality has been identified as a key challenge to address worldwide (OECD, 2018, 2021), our knowledge regarding the mechanisms of how this intergenerational transmission unfolds is limited, which impedes intervention efforts. In addition, little is known about the transmission of family background inequality in education across different countries and school systems.

A large body of research has confirmed strong genetic as well as environmental influences on educational achievement from early primary school until the end of compulsory schooling and regarding educational attainment in young adulthood (Shakeshaft et al., 2013; de Zeeuw et al., 2015, Bartels et al., 2002; Wilding et al., 2023). Thus, whether children struggle or excel in school clusters in families because of both, genetic and environmental transmission pathways. Genetic pathways refer to direct effects of parents' transmitted genotypes on offspring educational achievement and environmental pathways reflect influences of children's rearing environment that they experience over the course of development (e.g., socioeconomic conditions, parenting behaviour, neighbourhood characteristics). These putatively environmental factors, however, are known to also

be genetically influenced through processes of gene-environment correlation (rGE; Plomin et al., 1977). Thus, when measuring genetic effects on educational achievement, for example via children's polygenic score (PGS) of education, the effect does not only capture the 'true' genetic influence. Instead, rGE leads to children also inheriting environments from their parents, including the family socioeconomic background, hence estimates from samples of unrelated individuals might be biased. This issue can be alleviated by using data from biological siblings reared together (preferably dizygotic twins) to estimate the within-family predictiveness of PGS as children in one family share the same parental genotype (Selzam et al., 2019; Belsky et al., 2018). Beyond that, genetically informative family studies allow to not only control for within-family confounding, but to disentangle these pathways more clearly as they enable parsing effects of parents' transmitted alleles from those of non-transmitted alleles. Previous findings have shown that both, transmitted and non-transmitted genotypes are relevant predictors of children's educational outcomes which indicates indirect or environmentally-mediated genetic effects, which are referred to as 'genetic nurture' (Kong et al., 2018).

While first evidence suggests that the predictiveness of PGS is attenuated within families and also parents' non-transmitted genotypes are meaningful for children's education, so far, no study has investigated the role of contextual factors relevant for educational outcomes, such as characteristics of the school system, and potential differences across samples from different countries. One major difference between school systems in different countries lies in their degree of standardisation, which might be a relevant factor to consider for the transmission of family background inequality, because the level of standardization might determine how much of an influence the family background has on children's achievement. One good example here is comparing the educational systems in Britain and Germany. In the German school system, students are streamed in three-tiered secondary schools based on their primary school achievement (i.e., differentiated) and also within tiers, schools do not follow a standard curriculum. Hence, one would assume substantial variation between schools, which might be associated with relatively smaller effects of family background on differences in students' achievement. In contrast, the UK offers comprehensive secondary education for all children following a national curriculum, which potentially magnifies the relative importance of family background influences. While we hypothesise differences between the British and German sample to be in part accounted for by differences in the countries' educational systems, these findings might be confounded with additional sources of differences between sample (e.g., deviations in sampling, measurement and methodology).

Based on these considerations and previous findings, we will answer the following research questions in this study:

(1) To which extent is the prediction of children's educational achievement by PGS and SES attenuated within families?
(2) Does accounting for parents' genotype improve the prediction of children's educational achievement beyond the prediction from children's PGS?

(3) Is the effect of non-transmitted parental alleles on children's educational achievement confounded by family socio-economic status?
(4) Does the prediction and confounding differ between samples from Britain and Germany, thus across school systems?

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

By capitalizing on genetically informative data from family trios and quads in a German and a British sample, our findings will elucidate potential contextual effects on genetic and environmental transmission of family background inequality in education. The focus of this paper meets the current OECD call for identifying global strategies to reduce the transmission of family background inequality, which has been identified globally as an urgent challenge to address. Large-scale, comparative educational assessments such as PISA persistently show that growing up in socioeconomically deprived environments is a major risk for poor learning and educational outcomes. In the UK, the Social Mobility Commission (2022) highlighted in their latest communication plan that childhood poverty and impoverished home learning environments were key barriers for children to reach their educational potential and experience upward social mobility. Overcoming these and other barriers requires "interventions which work", calling evidence-based approaches in education that tackle causes, not confounders. With this paper we will contribute to building this evidence base.

Data analysis methods:

We will capitalise on data from British and German twin families (trios or quads, i.e., at least one parent and their twin children), drawn from E-Risk and TwinLife, two population-representative cohort studies from the UK and Germany, respectively. Both cohorts collected data on twins' school performance and their family socioeconomic background, and they both include genotypes from twins and their parents (either both parents or the twins' mother only). For the British sample, we will also include information on school quality (OFSTED ratings). TwinLife data in general contains information on the parental genotype from both parents, this might not be the case for all families due to missing values. We will handle this issue by using mendelian imputation of parental genotypes (Young et al., 2022), while also acknowledging that these families might differ from those with complete information on parental genotypes in part for systematic reasons (e.g., lower polygenic scores; see Wertz et al., 2023).

Using polygenic scores (PGS) for years of education (EA3; Lee et al., 2018), we will first compare the prediction of educational achievement (teacher ratings and report card data) over development from late primary throughout secondary school from measures of family socioeconomic status (SES) and their inherited DNA differences across Britain and Germany.

Second, we will leverage on DZ twin pairs from both cohorts to test the predictiveness of PGS within families (RQ1), which we expect to be attenuated due to stratification and assortative mating compared to predictions in samples of unrelated individuals (i.e., between families).

Third, we will investigate if accounting for the parental PGS improves the strengths of the prediction of offspring educational achievement over and above the children's PGS (RQ2), which would indicate genetic nurture. Additionally, we will control for SES to account for possible attenuation of the effect across countries due to family-level heterogeneity (RQ3). All analyses will be performed separately for the German and British sample and results will be compared to investigate potential differences that we would in part attribute to differences in the school systems in both countries (RQ4). Here we will also include OFSTED ratings to gain more insight into possible contextual effects in genetic nurture. Given that this information is only available for the British sample, we include the measure for exploratory analyses, however the data basis does not allow us to draw implications regarding the cross-country comparison.

Variables needed and at which ages:

Phenotypic data:

Age 5:

Variable name	Description
familyid	Family ID
atwinid	Twin1 ID
btwinid	Twin2 ID
rorderp5	Random order
sampsex	Sex
zygosity	Zygosity
hiedm5	Highest educational qualification mother
SESWQ35	Social Class Composite
ofsted99e5	OFSTED 1999 information Available - Phase 5
tyschl99e5	Type of school (OFSTED data) - Elder
year99e5	Inspection Year - Elder
peng_499e5	Percentage of KS2 English tests achieving level 4 and above,
pmath_499e5	Percentage of KS2 Mathematics tests achieving level 4 and above,
psci_499e5	Percentage of KS2 Science tests achieving level 4 and above,
qualgood99e5	Quality of teaching: % Very good or better (inspection data),
qualless99e5	Quality of teaching: % less than satisfactory (inspection data),
nchildren99e5	Total Number of Children - Elder
classsize99e5	Average class size - Elder
exppupil99e5	Expenditure per pupil (inspection data) - Elder
unabsence99e5	Unauthorised absence from school - Elder
schmeals99e5	Percentage of pupils eligible for free school meals - Elder
english99e5	Percentage of pupils with whom English is an additional language - Elder
ofsted00e5	OFSTED 2000 information Available - Phase 5
tyschl00e5	Type of school (OFSTED data) - Elder
year00e5	Inspection Year - Elder
peng_400e5	Percentage of KS2 English tests achieving level 4 and above,
pmath_400e5	Percentage of KS2 Mathematics tests achieving level 4 and above,
psci_400e5	Percentage of KS2 Science tests achieving level 4 and above,
qualgood00e5	Quality of teaching: % Very good or better (inspection data),
qualless00e5	Quality of teaching: % less than satisfactory (inspection data),
nchildren00e5	Total Number of Children - Elder
classsize00e5	Average class size - Elder

exppupil00e5	Expenditure per pupil (inspection data) - Elder
unabsence00e5	Unauthorised absence from school - Elder
schmeals00e5	Percentage of pupils eligible for free school meals - Elder
english00e5	Percentage of pupils with whom English is an additional language - Elder

Age 7:

Variable name	Description
TRFPEE7	English - School performance
TRFPEY7	English - School performance
TRFPME7	Maths – School performance
TRFPMY7	Maths – School performance

Age: 10

Variable name	Description
TRFPEE10	English - School performance
TRFPEY10	English - School performance
TRFPME10	Maths – School performance
TRFPMY10	Maths – School performance

Age 12:

Variable name	Description
TRFPEE12	English - School performance
TRFPEY12	English - School performance
TRFPME12	Maths – School performance
TRFPMY12	Maths – School performance

Age 18:

Variable name	Description
EDUCACHVE18	Highest educational achievement (child)
EDUCACHVY18	Highest educational achievement (child)

Genetic data:

Variable name	Description
zrpgsEA3	Residualized, standardized education polygenic score – both elder and younger needed
mom_zrpgsEA3	Residualized, standardized education polygenic score - mothers

References cited:

Bartels, M., Rietveld, M. J. H., Van Baal, G. C. M., & Boomsma, D. I. (2002). Heritability of educational achievement in 12-year-olds and the overlap with cognitive ability. *Twin Research: The Official Journal of the International Society for Twin Studies*, *5*(6), 544–553. https://doi.org/10.1375/136905202762342017

Belsky, D. W., Domingue, B. W., Wedow, R., Arseneault, L., Boardman, J. D., Caspi, A., Conley, D., Fletcher, J. M., Freese, J., Herd, P., Moffitt, T. E., Poulton, R., Sicinski, K., Wertz, J., & Harris, K. M. (2018). Genetic analysis of social-class mobility in five longitudinal studies. *Proceedings of the National Academy of Sciences*, *115*(31), E7275–E7284. https://doi.org/10.1073/pnas.1801238115

de Zeeuw, E. L., de Geus, E. J. C., & Boomsma, D. I. (2015). Meta-analysis of twin studies highlights the importance of genetic variation in primary school educational achievement. *Trends in Neuroscience and Education*, *4*(3), 69–76. https://doi.org/10.1016/j.tine.2015.06.001

French, M. T., Homer, J., Popovici, I., & Robins, P. (2015). What You Do in High School Matters: High School GPA, Educational Attainment, and Labor Market Earnings as a Young Adult. Eastern Economic Journal, 41(3), 386. https://doi.org/10.1057/eej.2014.22

Hair, N. L., Hanson, J. L., Wolfe, B. L., & Pollak, S. D. (2015). Association of Child Poverty, Brain Development, and Academic Achievement. *JAMA Pediatrics*, *169*(9), 822–829. https://doi.org/10.1001/jamapediatrics.2015.1475

Hodge, L., Little, A., & Weldon, M. (2021). GCSE attainment and lifetime earnings. Research report. Department for Education. https://www.gov.uk/government/publications/gcse-attainment-and-lifetime-earnings

Huang, C. (2015). Academic Achievement and Subsequent Depression: A Meta-analysis of Longitudinal Studies. *Journal of Child and Family Studies*, *24*(2), 434–442. https://doi.org/10.1007/S10826-013-9855-6

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsson, B. J., Young, A. I., Thorgeirsson, T. E., ... & Gudbjartsson, D. F. (2018). The nature of nurture: Effects of parental genotypes. *Science*, *359*(6374), 424-428.

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., Nguyen-Viet, T. A., Bowers, P., Sidorenko, J., Karlsson Linnér, R., Fontana, M. A., Kundu, T., Lee, C., Li, H., Li, R., Royer, R., Timshel, P. N., Walters, R. K., Willoughby, E. A., … 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*(8), Article 8. https://doi.org/10.1038/s41588-018-0147-3

Noble, K. G., Engelhardt, L. E., Brito, N. H., Mack, L. J., Nail, E. J., Angal, J., ... & Pass Network. (2015). Socioeconomic disparities in neurocognitive development in the first two years of life. *Developmental psychobiology*, *57*(5), 535-551.

OECD (2018), *Equity in Education: Breaking Down Barriers to Social Mobility*, PISA, OECD Publishing. https://doi.org/10.1787/9789264073234-en

OECD (2021), *Education at a Glance 2021: OECD Indicators*, OECD Publishing. *https://doi.org/10.1787/b35a14e5-en*

Plomin, R., DeFries, J. C., & Loehlin, J. C. (1977). Genotype-environment interaction and correlation in the analysis of human behavior. *Psychological Bulletin*, *84*, 309–322. https://doi.org/10.1037/0033-2909.84.2.309

Schwab, J. F., & Lew-Williams, C. (2016). Language learning, socioeconomic status, and child-directed speech. *Wiley Interdisciplinary Reviews: Cognitive Science*, *7*(4), 264-275.

Selzam, S., Ritchie, S. J., Pingault, J.-B., Reynolds, C. A., O'Reilly, P. F., & Plomin, R. (2019). Comparing Within- and Between-Family Polygenic Score Prediction. *The American Journal of Human Genetics*, *105*(2), 351–363. https://doi.org/10.1016/j.ajhg.2019.06.006

Shakeshaft, N. G., Trzaskowski, M., McMillan, A., Rimfeld, K., Krapohl, E., Haworth, C. M., ... & Plomin, R. (2013). Strong genetic influence on a UK nationwide test of educational achievement at the end of compulsory education at age 16. *PloS one*, *8*(12), e80341.

Social mobility commission (2022). *Social Mobility Commission Business Plan 2022 to 2023. A fresh approach to social mobility*. Report. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1112611/Social_Mobility_Commission_Business_Plan_2022_to_2023.docx.pdf

von Stumm, S. & Plomin, R. (2015). Socioeconomic status and the growth of intelligence from infancy through adolescence. Intelligence, 48, 30-36.

von Stumm, S. (2017). Socioeconomic status amplifies the achievement gap throughout compulsory education independent of intelligence. Intelligence, 60, 57–62.

von Stumm, S., Smith-Woolley, E., Ayorech, Z., McMillan, A., Rimfeld, K., Dale, P., & Plomin R., (2020). Predicting educational achievement from genomic measures and socioeconomic status. Developmental Science, 23, 1-8.

Wertz, J., Moffitt, T. E., Arseneault, L., Barnes, J.C., Boivin, M., Corcoran, D. L., …, & Caspi, A. (2023). Genetic associations with parental investment from conception to wealth inheritance in six cohorts. Preprint.

Wilding, K., Wright, M., & von Stumm, S. (2023). *Using DNA to predict Educational Success: A Meta-Analytic review*. PsyArXiv. https://doi.org/10.31234/osf.io/f7vt9

Young, A. I., Nehzati, S. M., Benonisdottir, S., Okbay, A., Jayashankar, H., Lee, C., Cesarini, D., Benjamin, D. J., Turley, P., & Kong, A. (2022). Mendelian imputation of parental genotypes improves estimates of direct genetic effects. *Nature Genetics*, *54*(6), Article 6. https://doi.org/10.1038/s41588-022-01085-0

DATA SECURITY AGREEMENT

Provisional Paper Title: The transmission of family background inequality in education across Britain and Germany
Proposing Author: Dr Alexandra Starr, Prof Sophie von Stumm
Today's Date: 8/6/2023

☒	I am current on Human Subjects Training (CITI (www.citiprogram.org) or equivalent)
☒	My project is covered by the Duke or King's ethics committee OR I have /will obtain ethical approval from my home institution.
☒	I will treat all data as "restricted" and store in a secure fashion. My computer or laptop is: a) encrypted (recommended programs are FileVault2 for Macs, and Bitlocker for Windows machines) b) password-protected c) configured to lock-out after 15 minutes of inactivity AND d) has an antivirus client installed as well as being patched regularly.
☒	I will not "sync" the data to a mobile device.
☒	In the event that my laptop with data on it is lost, stolen or hacked, I will immediately contact Prof Helen Fisher (helen.2.fisher@kcl.ac.uk).
☒	I will not share the data with anyone, including my students or other collaborators not specifically listed on this concept paper.
☒	I will not post data online or submit the data file to a journal for them to post. *Some journals are now requesting the data file as part of the manuscript submission process. Study participants have not given informed consent for unrestricted open access, so we have a managed-access process. Speak to* Prof *Helen Fisher (helen.2.fisher@kcl.ac.uk) for strategies for achieving compliance with data-sharing policies of journals.*
☒	I will delete all data files from my computer after the project is complete. Collaborators and trainees may not take a data file away from the office. This data remains the property of the Study and cannot be used for further analyses without an approved concept paper for new analyses.
☒	I have read the Data Use Guidelines and agree to follow the instructions.

Signature: Alexandra Starr