ENVIRONMENTAL-RISK (E-RISK) LONGITUDINAL TWIN STUDY
CONCEPT PAPER FORM

Proposing Author: DW Belsky

Author's affiliation, phone, and e-mail address: Duke U Dept of Pop Health Sciences; +1 919 613 4534; dbelsky@duke.edu

Sponsoring Investigator (if the proposing author is a student, a post-doc or a colleague): TE Moffitt & A Caspi

Proposed co-authors: TE Moffitt, A Caspi, L Arseneault, K Sugden, D Corcoran, J Prinz, B Williams, J Mill, and J Wertz as part of a team of investigators also including B Domingue, M Nivard, KP Harden, E Tucker-Drob, and others.

Provisional Paper Title: Development and evaluation of a polygenic score for non-cognitive skills

Date: August 15, 2018

Objective of the study and its significance:

OBJECTIVE: To develop and evaluate a novel genetic measurement of an inherited tendency to develop non-cognitive skills influencing life course socioeconomic attainment.

SIGNIFICANCE: A genetic measurement of such skills could be used to test (a) how non-cognitive skills develop, including potential "genetic nurture" effects; (b) the extent to which non-cognitive skills influence performance on measurements of academic skills/abilities and cognitive tests; (c) the importance of non-cognitive skills to life attainments, health behaviors, and healthy lifespan.

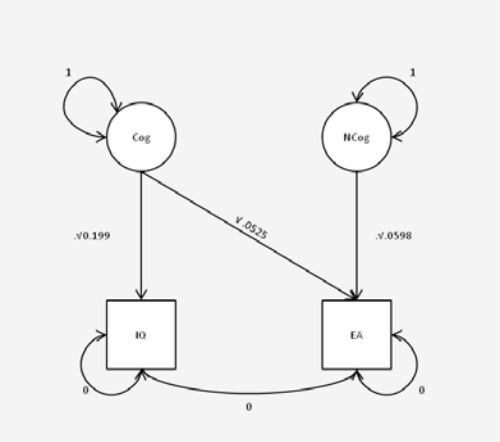
Expanded Significance. Success in school and socioeconomic attainment in adulthood are important determinants of healthy lifespan (1). Understanding the etiology of this pattern of achievement is therefore of public health significance (2). In life-course longitudinal studies, individual differences in cognitive skills and other skills loosely referred to under the umbrella of "non-cognitive skills" influence success in school and beyond (3–5). Individual differences in cognitive and so-called non-cognitive skills are heritable (6, 7). Recently, genome-wide association studies of educational attainment and cognitive test performance have begun to outline the molecular basis of this heritability (8). Polygenic scores, summary measures of tiny-effect genetic influences from across the whole genome (9), can be derived from these GWAS and, in independent samples, predict educational and economic outcomes as well as individual differences in cognitive and non-cognitive skills (10–14). These data suggest education GWAS discoveries reflect genetics influencing both cognitive and non-cognitive skill development. At present, it is not possible to study these genetics separately; the educational attainment polygenic score measures their combined influence on success in school. Having available measurements of genetics that uniquely affect development of either cognitive or non-cognitive skills would enhance opportunities to investigate independent etiologies and sequelae of cognitive and non-cognitive skills.

STATISTICAL ANALYSIS

Stage 1. Development

We propose to use the new Genomic Structural Equation Modeling (GSEM) suite of methods (15) to derive a GWAS of non-cognitive skills from published GWAS results for the phenotypes of educational attainment and cognitive test performance. We will use GSEM to approximate an analysis which effectively runs a GWAS of educational attainment with a covariate for cognitive test performance. The residual genetic coefficients from this model, by construction, reflect genetic influences on non-cognitive skills influencing educational attainment. GSEM will actually produce two novel sets of GWAS summary statistics. One set will reflect genetic influences shared between education and cognitive test performance. The second set will reflect genetic influences on education that are not shared with cognitive test performance, i.e. "non-cognitive" summary statistics.

The structural equation path diagram can be visualized below (path coefficients from preliminary analysis)



Development analyses will be conducted by the laboratories of M Nivard & Elliott Tucker-Drob at the Vrije University Amsterdam and the University of Texas at Austin. *Development analyses involve only GWAS summary data and will not involve any E-Risk Data.*

Stage 2. Validation

We will conduct analyses of 4 polygenic scores:
EA3 – based on the original education GWAS summary statistics (8)
IQ – based on original GWAS of cognitive test performance (16)
Cog – based on the overlap between EA3 and IQ GWAS
NonCog – based on genetics unique to EA3 GWAS and that do not overlap with IQ GWAS

We will conduct 3 sets of analyses using E-Risk and Data:

1) Test correlations among derived polygenic scores
2) Test correlations of derived polygenic scores with
    a. Educational attainment - measured as in (12)
    b. IQ – measured in childhood as in (3) and at age 18y by WAIS (it may also be of interest to examine WAIS sub-tests and CANTAB scores as a second stage analysis)
        i. We propose analysis of WAIS subtests and potential follow-up in CANTAB tests because the GWAS of cognitive test performance consists mostly of samples with cognitive tests of "fluid" intellectual functions (such as are measured by the WAIS Matrix Reasoning and Digit-Symbol Coding sub-tests). Therefore, a possible outcome of our GSEM analysis is that we will parse the EA3 GWAS results into a component reflecting primarily fluid intellectual functions and a component that comprises some combination of so-called "crystalized" cognitive functions (such as are measured by the WAIS Information subtest) and non-cognitive skills.
    c. Self Control – measured as in (3)

d. Big 5 Personality Traits – measured at age 18y (BFI variables)

3) Test multivariate associations of derived polygenic scores with the outcomes in (2)

We will conduct parallel analyses of similar phenotypes in the Add Health Study, the Texas Twin Study, and, pending approval from the Study PIs, the Dunedin Study and other studies.

The primary objective of validation analysis is to test if derived NonCog PGS associations with non-cognitive phenotypes (personality, self-control) are stronger than NonCog PGS associations with cognitive phenotypes (IQ, potentially WAIS subtests/ CANTAB scales). In parallel, we will test of Cog PGS associations with cognitive phenotypes are stronger than Cog PGS associations with non-cognitive phenotypes. A secondary objective is to compare effect-sizes for the two polygenic scores for cognitive and non-cognitive phenotypes and for educational attainment. A caveat to this secondary objective is that statistical power is not equivalent for derivation for the cognitive and non-cognitive PGSs, so results will not permit strong inference about the genetic architecture of the phenotypes. However, these analyses will shed light on how these polygenic scores should be interpreted in future analyses.

Validation analyses using E-Risk data will be conducted by DW Belsky with collaboration from TE Moffitt, A Caspi, K Sugden, D Corcoran, and J Prinz.

Stage 3. Phenotypic Annotation

Phenotypic annotation analysis will test associations of newly derived polygenic scores with outcomes in school and beyond using data from studies with follow-up into adulthood. *E-Risk data are not requested for this stage of analysis.*

Variables Needed at Which Ages (names and labels):

Polygenic scores computed from summary statistics to be provided to K Sugden at Duke.

Study:

Age 5:
IQE5

Age 7:

Age 10:
LOWSC510

Age 12:

Age 18:

Big 5 Personality (BFI variables – requesting guidance on whether we should focus on interviewer, informant, or composite of both)
WAIS (FSIQ variables + MRE, INFE, INMRE, DSCE variables)
CANTAB (No variables needed at present.

References cited:

1. Adler NE, Rehkopf DH (2008) U.S. Disparities in Health: Descriptions, Causes, and Mechanisms. *Annu Rev Public Health* 29(1):235–252.

2. Berkman LF, Kawachi I, Glymour MM (2014) *Social epidemiology* (Oxford University Press).

3. Moffitt TE, et al. (2011) A gradient of childhood self-control predicts health, wealth, and public safety. *Proc Natl Acad Sci U S A* 108(7):2693–2698.

4. Strenze T (2007) Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence* 35:401–426.

5. Deary I J, Strand S, Smith P, Fernandes C (2007) Intelligence and educational achievement. *Intelligence* 35(1):13–21.

6. Polderman TJC, et al. (2015) Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet* 47(7):702–709.

7. Tucker-Drob EM, Briley DA, Engelhardt LE, Mann FD, Paige K (2016) Genetically-mediated associations between measures of childhood character and academic achievement. *J Pers Soc Psychol* 111(5):790–815.

8. Lee JJ, et al. (2018) Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* 50(8):1112–1121.

9. Dudbridge F (2013) Power and Predictive Accuracy of Polygenic Risk Scores. *PLOS Genet* 9(3):e1003348.

10. Belsky DW, et al. (2016) The Genetics of Success: How Single-Nucleotide Polymorphisms Associated With Educational Attainment Relate to Life-Course Development. *Psychol Sci* 27(7):957–972.

11. Wertz J, et al. (In Press) Genetics and crime: Integrating new genomic discoveries into psychological research about antisocial behavior. *Psychol Sci*.

12. Belsky DW, et al. (2018) Genetic analysis of social-class mobility in five longitudinal studies. *Proc Natl Acad Sci*:201801238.

13. de Zeeuw EL, et al. (2014) Polygenic scores associated with educational attainment in adults predict educational achievement and ADHD symptoms in children. *Am J Med Genet Part B Neuropsychiatr Genet Off Publ Int Soc Psychiatr Genet* 165B(6):510–520.

14. Plomin R, von Stumm S (2018) The new genetics of intelligence. *Nat Rev Genet*.

15. Grotzinger AD, et al. (2018) Genomic SEM Provides Insights into the Multivariate Genetic Architecture of Complex Traits. *bioRxiv*:305029.

16. Savage JE, et al. (2018) Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nat Genet* 50(7):912–919.

Data Security Agreement

Provisional Paper Title	Development & Evaluation of a Polygenic Score for Non-Cognitive Skills
Proposing Author	DW Belsky
Today's Date	August 15, 2018

Please keep one copy for your records
(Please initial your agreement)

DWB I am familiar with the King's College London research ethics guidelines (https://www.kcl.ac.uk/innovation/research/support/ethics/about/index.aspx) and the MRC good research practice guidelines (https://www.mrc.ac.uk/research/policies-and-guidance-for-researchers/good-research-practice/).

DWB My project has ethical approval from my institution. [NEW APPROVAL WILL BE OBTAINED FROM COLUMBIA AFTER DEC 2018]

DWB I am familiar with the EU General Data Protection Regulation (https://mrc.ukri.org/documents/pdf/gdpr-guidance-note-3-consent-in-research-and-confidentiality/), and will use the data in a manner compliant with its requirements.

DWB My computer is (a) encrypted at the hard drive level, (b) password-protected, (c) configured to lock after 15 minutes of inactivity, AND (d) has an antivirus client which is updated regularly.

DWB I will treat all data as "restricted" and store in a secure fashion.

DWB I will not share the data with anyone, including students or other collaborators not specifically listed on this concept paper.

DWB I will not merge data from different files or sources, except where approval has been given by the PI.

DWB 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. The E-Risk Study cannot be shared because the Study Members have not given informed consent for unrestricted open access. Speak to the study PI for strategies for dealing with data sharing requests from Journals.

DWB Before submitting my paper to a journal, I will submit my draft manuscript and scripts for data checking, and my draft manuscript for co-author mock review, allowing three weeks.

DWB I will submit analysis scripts and new variable documentation to project data manager after the manuscript gets accepted for publication.

DWB I will delete the data after the project is complete.

____ **For projects using location data:** I will ensure geographical location information, including postcodes or geographical coordinates for the E-Risk study member's homes or schools, is never combined or stored with any other E-Risk data (family or twin-level data)

DWB **For projects using genomic data:** I will only use the SNP and/or 450K data in conjunction with the phenotypes that have been approved for use in this project at the concept paper stage.

Signature: DWB

CONCEPT PAPER RESPONSE FORM

A. To be completed by the proposing author

Proposing Author: DW Belsky

X I have read the E-Risk data-sharing policy guidelines and agree to follow them

Provisional Paper Title: Development & Evaluation of a Polygenic Score for Non-Cognitive Skills

Potential co-authors: TE Moffitt, A Caspi, L Arseneault, K Sugden, D Corcoran, J Prinz, B Williams, J Mill, and J Wertz as part of a team of investigators also including B Domingue, M Nivard, KP Harden, E Tucker-Drob, and others.

Potential Journals:

Intended Submission Date (month/year): June 1, 2019

Please keep one copy for your records and return one to Louise (louise.arseneault@kcl.ac.uk)

B. To be completed by potential co-authors:

☐ Approved ☐ Not Approved ☐ Let's discuss, I have concerns

Comments:

Please check your contribution(s) for authorship:

☐ Conceptualizing and designing the longitudinal study

☐ Conceptualizing and collecting one or more variables

☐ Data collection

☐ Conceptualizing and designing this specific paper project

☐ Statistical analyses

☐ Writing

☐ Reviewing manuscript drafts

☐ Final approval before submission for publication

☐ Acknowledgment only, I will not be a co-author

Signature: .....