

Research Articles

A Meta-Analysis of Spearman's Hypothesis in the United States

Michael Woodley of Menie¹ , Mateo Peñaherrera-Aguirre² , Matthew Sarraf¹

¹ Independent Sector, ² University of Arizona

Keywords: Non-shared environmentality, Spearman's Hypothesis, Vector Correlation, White-Black difference

<https://doi.org/10.65550/001c.171909>

Intelligence & Cognitive Abilities

A comprehensive meta-analysis of vector-correlation tests of Spearman's Hypothesis (SH) with respect to US White-Black (W-B) cognitive performance differences is conducted. Thirty-one usable studies yielding 44 independent effect sizes were identified, with a total harmonic N of 1,039,379. A random-effects model, using subtest number to estimate inverse-variance weights, yielded a large-magnitude positive mean vector correlation, $r = .745$, 95% CI [.699, .792], with a prediction interval of .628 to .862, and modest between-study heterogeneity, $I^2 = 25.61\%$, 95% CI [0.00%, 50.5%]. Sampling error accounted for 46.3% to 82.7% of between-study variance, depending on estimator, suggesting that some variance in those data is potentially due to moderators rather than sampling error. A three-moderator meta-regression including harmonic N , sample mean age, and study year yielded evidence of negative moderation by sample age. SH therefore appears to be weaker in older samples. No quadratic effect of age was detected. Funnel-plot diagnostics using Fisher- z -transformed effects indicated significant unadjusted Egger's asymmetry, but this became non-significant after adjustment for the moderators. Trim-and-fill imputed no missing effect sizes. The presence of age-dependent moderation, including both reversals and attenuation of effect sizes in older samples, makes claimed isomorphism between g and W-B differences difficult to sustain. The results instead suggest complex environmental confounding potentiating, attenuating, or even reversing the g -dependence of W-B differences.

Introduction

Spearman's Hypothesis (SH; Jensen, 1980) refers to the empirical tendency for (socially identified) White-Black¹ (W-B) differences in cognitive test performance to be more pronounced on subtests that are more highly saturated with the general factor of intelligence, or g (Spearman, 1927, pp. 379–380). In its standard operational form, the hypothesis has most commonly been tested using Jensen's (1985, 1998) method of correlated vectors (MCV), in which a vector of subtest g -loadings is correlated with a vector of standardized group differences (often Cohen's d statistics) across the same subtests. A positive vector correlation suggests that g can be said to *moderate* the group difference, whereas a weak, null, or negative vector correlation indicates attenuation or reversal of the expected SH pattern and is suggestive of the moderating role of non- g sources of ability variance (such as lower-order factors and test specificities).

SH, especially in relation to the results of studies using MCV, has historically been controversial. A major reason for this is the argument that a positively signed vector cor-

relation is not necessarily indicative of a g moderated difference amongst groups. Differences with respect to non- g latent variables (such as group factors) can generate the appearance of g moderated differences in instances where a battery is sufficiently nomologically unbalanced with respect to the sampling of cognitive domains (Ashton & Lee, 2005; Jensen, 1998). Multi-group confirmatory factor analysis (MG-CFA) has however repeatedly found that both weak and strong forms of measurement invariance obtain in US W-B differences (with g accounting for between 51% and 100% of the variance in the group difference across studies; Lasker et al., 2021, Table 7, p. 24) which indicates that the straightforward interpretation of SH as demonstrating some non-trivial degree of moderation from g on the group-difference vector is reasonable.

The status of SH is theoretically important because it has been argued that it bears on the potential causes of group differences across cognitive abilities (Jensen, 1998; Rush-ton & Jensen, 2010; Warne, 2021). The racial hereditarian hypothesis holds that the reason why group differences

¹ These are census race categories as defined in the US based on the *Office of Management and Budget* (1997) standards on race and ethnicity; it is in this context that the terms "Black" and "White" are used throughout this paper. This is written as White-Black (rather than the other way around) because mathematically, the difference scores are typically estimated by subtracting the latter from the former, which generally (but not always) maintains sign positivity in the resultant difference statistic.

scale with subtest g loading is because genetic influences (as proxied by inbreeding depression effects and heritabilities) tend also to scale with g loading (Rushton, 1999; Warne, 2021). It has also been hypothesised that as more g -loaded subtests exhibit greater cultural loading, it is the biasing effect of these on the performance of differentially enculturated groups which generates SH (Kan, 2012). More g loaded subtests have also been found to be more sensitive to gene-by-environment interactions involving the suppressing effect of perceived racial and ethnic discrimination on the expressivity of cognitive polygenic scores. In Peñaherrera-Aguirre et al. (2022) it was found specifically that the vector of these clustered positively and significantly with a group-differences vector, g loadings, and a shared environmentality vector. If group differences are instead substantially influenced by narrower cognitive abilities, test-specific sources of variance, measurement error, or environmental perturbations that are not strongly aligned with g (such as those stemming from the non-shared environment), then the vector correlation should be weaker. Thus, variation in the magnitude of the SH vector correlation across studies might, in some instances, provide limited information about causal moderation.

Previous meta-analyses and non-systematic reviews have generally supported SH in US W-B comparisons, with positive vector correlations observed across all cognitive batteries and samples (Jensen, 1985, 1998; te Nijenhuis & van den Hoek, 2016). However, there is also evidence that the direction of the effect is not invariant across contexts. Of relevance are findings from incarcerated and institutionalized (carceral) populations, amongst whom SH appears to be reversed (Woodley of Menie et al., 2025). One possible explanation for this finding is that carceral environments are ecologically harsh, increasing the influence of non-shared environmental variance on cognitive performance. Non-shared environmental effects encompass influences that reduce phenotypic resemblance amongst relatives, and may include health insults, violence, ecological stressors, and other person-specific life-course influences (Plomin, 2024). Precisely in line with this, Woodley of Menie et al. (2025) found that carceral W-B difference vectors cluster *positively* with vectors of non-shared environmentality, whilst g and culture-loading vectors show the opposite (*negative*) clustering pattern.

Age is a theoretically important moderator of SH in the context of representative samples. Although the heritability of cognitive ability tends to increase from childhood into early adulthood (this being the Wilson Effect; Bouchard, 2013), evidence from later life suggests that the structure of cognitive variance changes again, as health-related and idiosyncratic environmental influences become more prominent. This manifests as both a decrease in the additive (starting in midlife) and an increase in the non-shared environmental variance components (termed the anti-Wilson Effect; Sarraf et al., 2023, 2025, Fig. 1, p. 2, see also Morris et al., 2025 for similar findings in relation to age-dependent increases in non-shared environmentality amongst younger cohorts). This increase in non-shared environmental variance with age should weaken the orga-

nization of group differences around g , causing the vector correlations indexing SH to become attenuated.

The present study

The principal goal of this meta-analysis is to identify as many suitable vector correlation tests of SH as possible for the purpose of generating a meta-analytic point estimate of SH. The primary effect size is the (uncorrected) vector correlation between subtest g -loadings and W-B subtest-level differences. As the relevant sampling unit for these correlations is the number of subtests entering each vector, each effect size is weighted using subtest N , with standard errors (SEs) derived from the reported 95% confidence intervals and inverse variance. A Hedges and Olkin-type random-effects framework is used to estimate the overall magnitude of the vector correlation and the extent of between-study heterogeneity.

The more theoretically meaningful goal is to determine whether sample characteristics moderate the magnitude of these vector correlations. Three moderators are examined: harmonic N (the harmonic mean of the pooled Black and White subsamples), mean age (or midpoint when only a range is given), and study year. Harmonic N is included to account for differences in the effective precision of group comparisons; study year is included to capture possible secular or other trends (such as the Decline Effect); and, as noted previously, sample age tests the hypothesis that the g -dependence of W-B differences attenuates with age, potentially due to the effects of rising non-shared environmental influence on cognitive differences. Critically, evidence for the influence of the latter moderator would indicate that SH is not a fixed property of W-B cognitive-performance differences, as has been suggested most notably by Gordon (1985) who argued on the basis of Tucker's congruence coefficients (an alternative to MCV) that "[t]he black-white factor is g " (p. 229, italics for emphasis).

Methods

Meta-analytic search

Previous meta-analyses of SH in relation to the US W-B difference have employed restricted criteria for study inclusion. For example, te Nijenhuis and van den Hoek (2016) used the following rules:

The first requirement for a study to be included in the present meta-analysis was having at least four tests or subtests to which the Method of Correlated Vectors could be applied. Second, the samples in the studies should not be selected on a highly g -loaded variable (e.g., referral or gifted samples...). Third, the average age of the participants had to be 18 or older, which means the sample could include some participants younger than 18. (p. 5)

Aspects of these inclusion criteria were replicated in the present work; however certain constraints were lifted and new ones imposed. Firstly, to test for potential age moderation, the requirement that samples be screened out based

on mean age <18 years was dropped. Secondly, the requirement for a minimum of four subtests was tightened to five, as subtest number is likely related to the nomological breadth of a given battery, and nomological breadth strongly influences the stability of vector correlations (Jensen, 1998). Thirdly, samples were excluded in cases where they included those exhibiting pathology (such as Alzheimer's patients). Carceral samples (specifically those who were tested *in situ* with respect to the prison system) were also excluded from the meta-analysis as incarceration risk is a potentially *g* loaded variable, and they are therefore non-representative of the broader population (Jensen, 1992). Fourthly, studies that examined Spearman's Hypothesis using job complexity criteria (such as Roth et al., 2001) were screened out, as were item-level analyses of SH (such as Kirkegaard & Hu, 2025).

Three meta-analytic studies (te Nijenhuis et al., 2017; te Nijenhuis & van den Hoek, 2016; Woodley of Menie et al., 2025) and three non-systematic reviews (Jensen, 1985, 1998; Lasker et al., 2021) were consulted. As all four studies covered in Woodley of Menie et al. (2025) concerned carceral populations, they were excluded based on being non-representative. In addition, we used a modified form of the Boolean string and reference search criteria employed by te Nijenhuis and van den Hoek (2016), specifically:

[T]he following electronic databases were used: Google Scholar, ProQuest, PsycINFO, and CataloguePlus (Primo). The keywords and phrases used to find further data for the data of Black adults were: "Black adult(s)", "Black and White adults", "Negro adults", "Spearman's hypothesis adults", "Spearman's hypothesis Black", "Spearman's hypothesis Negro", and "Spearman's hypothesis workforce". These keywords were combined with the following keywords: "intelligence", "IQ", "mental ability", "mental capacity", "cognitive ability", "aptitude", "competence", "differences", "WAIS" (Wechsler Adult Intelligence Scale), "KAIT" (Kaufman Adult Intelligence Test), and "Woodcock Johnson". To supplement the data we found, we searched the references of the studies already obtained to find other studies with potential data. (p. 5)

This was supplemented with additional strings corresponding to "Black children", "Black and White children", "Negro children", "Spearman's hypothesis children". As with te Nijenhuis and van den Hoek (2016) studies were only selected for screening if they concerned US-based individuals. The searches were conducted twice by the lead author. Firstly, at the inception of the research phase of the study in February 2025, and again in June 2026 at the closing out of the study. For the second round of searches, assistance was provided by the third author. No formal protocol was registered.

Several studies reported data on overlapping samples (for example, multiple studies have examined SH in relation to the various Woodcock-Johnson normalisation samples, the Armed Services Vocational Aptitude Battery [ASVAB] in the National Longitudinal Survey of Youth [NLSY] '79 cohort, Project Talent etc). In these cases, either the most re-

cent study, or the one with the largest sample size and/or subtest number was retained.

The search was enhanced with three additional measures aimed at further increasing the pool of usable effect sizes. Firstly, studies reporting data based on scholastic attainment criteria (such as the correlations between Advanced Placement examination performance and a measure of *g*) were admitted into the meta-analysis, as were studies employing *g* loadings derived from the common factor amongst elementary cognitive task groupings. This yielded an additional two studies and four effect sizes (Jensen, 1993; Warne, 2016). Secondly, studies were retained in which, via reanalysis of raw data, new tests of SH could be conducted. This netted four additional effect sizes (MIDUS-II and MIDUS Refresher with the Milwaukee African American sample; Woodley of Menie et al., 2015, and two further effect sizes from Osborne's [1980] study). A further nine usable studies were identified based on the search (Blankson & McArdle, 2015; Dalliard, 2013; Fuerst et al., 2023; Hu, 2025; Kane, 2004, 2007; Kane & Brand, 2008; National Academy of Sciences [NAS], 1921; Shibaev & Fuerst, 2026).

These studies were combined with the four studies retained from te Nijenhuis and van den Hoek (2016; Carretta, 1997; Hennessy & Merrifield, 1976; Kaufman et al., 1988; Sternberg & The Rainbow Project Collaborators, 2006), the eight studies retained from Jensen (1985; National Center for Education Statistics, 1972, Kaufman & Kaufman, 1983; Mercer, 1984; Nichols, 1972; Osborne, 1980; Reynolds & Gutkin, 1981; Scarr, 1981; Veroff & McClelland, 1971), the two additional studies retained from Jensen (1998, Naglieri & Jensen, 1987; Peoples et al., 1995), and the five studies retained from Lasker et al. (2021, Frisby & Beaujean, 2015; Hu, 2017; Hu et al., 2019; Lasker et al., 2019). The total yield was of 31 studies comprising 44 independent effect sizes.

Figure 1 presents a PRISMA-style flow chart illustrating how the initial pool of studies was reduced to the final sample.

Meta-analytic methods

Two meta-analytic strategies were employed. Firstly, the Hedges and Olkin approach was used to estimate the meta-analytic effect sizes and between-study heterogeneity (I^2). Studies were weighted based on inverse variance using SEs derived from their subtest numbers, and a random effects model was estimated. Three moderators were employed; harmonic mean of sample size, defined as $n_h = \frac{2}{\frac{1}{n_w} + \frac{1}{n_b}}$,

where n_w and n_b correspond to the individual White and Black group sizes respectively (Klockars & Sax, 1987). This allowed for any additional precision-related influences associated with differences in sample size, net of variation in subtest number to be examined. Mean (or midpoint) sample ages were sourced from the studies or were inferred based on additional criteria. As noted in the Introduction, age is a prospectively important moderator because it potentially captures rising non-shared environmentality which should weaken SH. Consequently, a very broad range of ages were considered in this study, with the youngest sample having

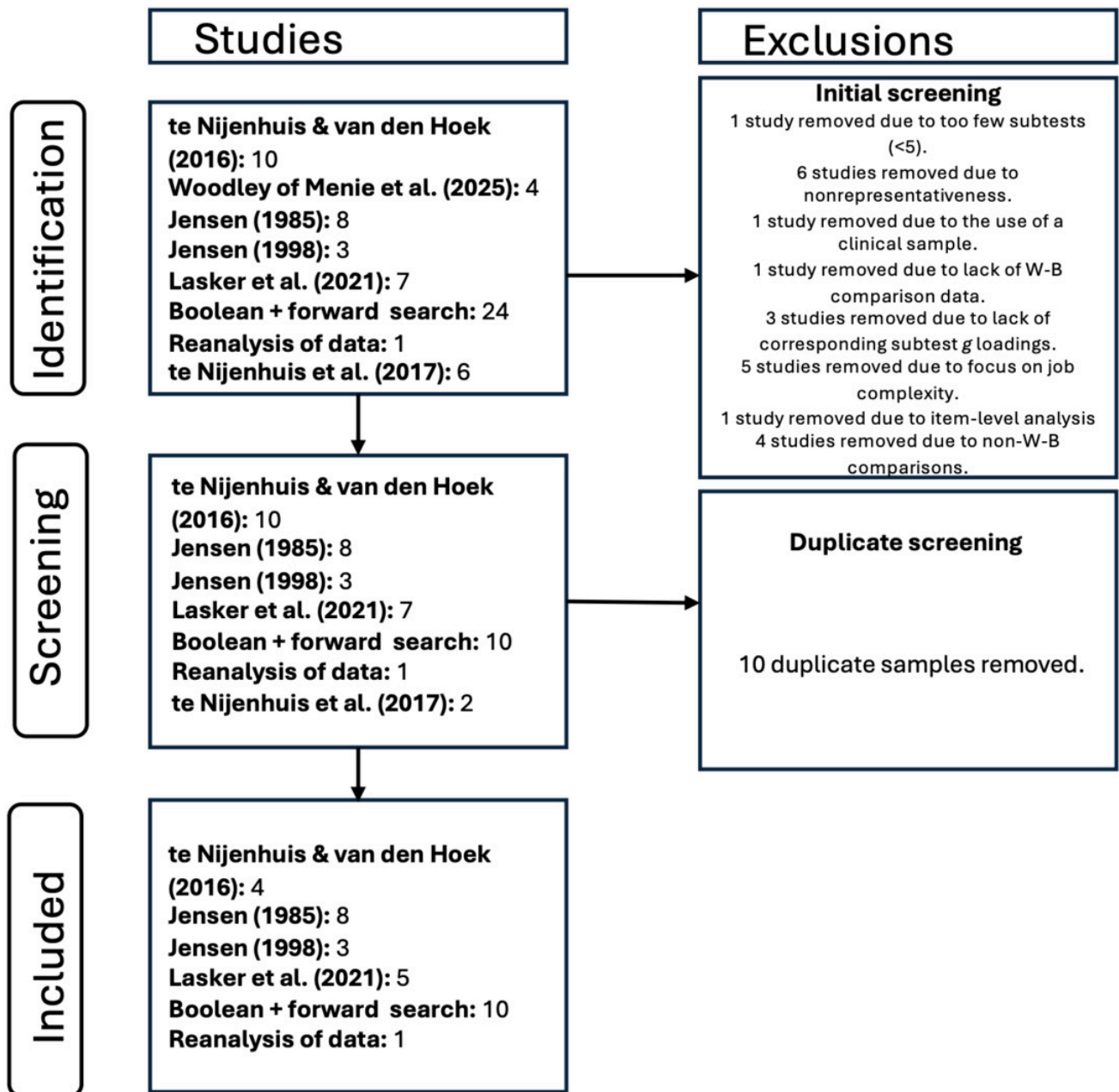


Figure 1. PRISMA-style flow chart illustrating the progression from the initial pool of 63 studies to the final usable sample of 31 studies, with 44 independent effect sizes.

a mean age of just three (Peoples et al., 1995) and the oldest 76 years (Shibaev & Fuerst, 2026), yielding a span of 73 years. Finally, study year was also included to test for the possibility of secular trends or so-called Decline Effects (decreases in effect sizes over time) perhaps reflecting selective reporting early in the research paradigm (e.g., Schooler, 2011). The oldest study from which an estimate of SH could be obtained was Robert Yerkes' (1876-1956) Army Alpha and Beta validation (NAS, 1921); the most recent was from 2026 (Shibaev & Fuerst, 2026), yielding a span of 105 publication years. Funnel asymmetry was examined using Fisher's *r*-to-*z* transformed vector correlations. Egger's regressions were estimated before and after adjustment for the three moderators. Duval and Tweedie's (2000) trim-

and-fill procedure was used to evaluate whether asymmetry suggested missing effect sizes.

Secondly, sampling error was also estimated using a bare-bones meta-analytic approach. Sampling error estimates allow for the degree to which between-study variance might be substantive, i.e. due to the action of potential moderators (Schmidt & Hunter, 2015).

Software

All analyses were conducted in *Python 3.13.5* using *NumPy* (Harris et al., 2020), *pandas* (McKinney, 2010), *SciPy* (Virtanen et al., 2020), and *Matplotlib* (Hunter, 2007). Excel input/output was handled using *pandas* with *openpyxl/XlsxWriter*. Custom *Python* functions, generated with GPT 5.5, were used to estimate restricted maximum likelihood ran-

dom-effects meta-regression models, Egger's regressions, and trim-and-fill diagnostics. All code used in this analysis is included in the first Supplement (S1). All data are included as an Excel spreadsheet in the second Supplement (S2).

Results

Meta-analysis

The study effect sizes are presented in the forest plot depicted in [Figure 2](#).

The results of the random effects meta-analysis are reported in [Table 1](#).

The aggregate effect size is .745, 95% CI [.699, .792]. This is of large magnitude ($r \geq .50$, Cohen, 1988). Heterogeneity fell into the moderate range based on conventional guidelines, $I^2 = 25.61\%$, 95% CI [0.00%, 50.5%], although the confidence interval spans values consistent with relatively low to substantial heterogeneity (Deeks et al., 2024). The prediction interval ranged from .628 to .862. The results of the moderator analysis are presented in [Table 2](#).

Moderator collinearity was low, with variance inflation factors ranging from 1.043 to 1.079 (values < 5 are considered indicative of negligible collinearity; Menard, 1995). A leave-one-out (LOO) analysis was conducted on the age moderator. The coefficient remained negative after every iteration (involving the removal of one of the 31 studies) and was significant 21 out of 31 iterations, with B values ranging from -.006 to -.002 per year, and an average LOO $p \approx .051$.

A quadratic analysis was also conducted on age to check for non-linear moderation; however, the results were non-significant ($F(1,41) = .002$, $p = .962$, $\Delta AIC = +1.98$).

The age moderation effect is graphed in [Figure 3](#).

Next, Schmidt and Hunter's (2015) bare-bones meta-analytic sampling error was estimated. The results are presented in [Table 3](#).

Only one of these estimates exceeds the 75% rule-of-thumb threshold (Schmidt et al., 1988) indicating that variability amongst the vector correlations likely reflects meaningful residual variance and is therefore more than just sampling error, consistent with the action of potential age moderation.

Publication bias

A funnel plot was generated using r -to- z transformed vector correlations to investigate whether the effect sizes were symmetrically distributed relative to the mid-line. Two sets of Egger's regression coefficients were also estimated to determine the effects of moderators on any asymmetry in the distribution amongst effect sizes. The plot is presented in [Figure 4](#).

The results of the Egger's regressions (using delta-method Fisher- z SEs) are presented in [Table 4](#).

The funnel asymmetry appears to be due, in large part, to the effects of the moderators. Once these are controlled the asymmetry becomes statistically non-significant. Trim-and-fill revealed no potentially missing effect sizes.

As the primary model is presented using a raw- r scale, a sensitivity analysis was conducted in which models that were also estimated on the Fisher- z scale were back-transformed; these yielded substantively similar results.

Discussion

The meta-analysis of SH in US W-B comparisons yields evidence of a large-magnitude effect size ($r = .745$, 95% CI [.699, .792]), exhibiting modest between-effect size heterogeneity. Sampling error accounts for between 46.3% and 82.7% of the variance amongst effect sizes (depending on the estimation protocol), leaving some room for moderator-related heterogeneity. The age moderator was found to exhibit a statistically significant effect. Amongst the unadjusted (z -transformed) effect sizes, funnel asymmetry plots yielded indications of significant asymmetry, however when moderator-controlled values were used, the asymmetry became non-significant. Trim-and-fill yielded no indications of missing effect sizes.

These findings greatly generalise those of more restricted meta-analyses of SH in adult US W-B data (te Nijenhuis & van den Hoek, 2016) extending their scope with respect to moderation. If increasing age is associated with a larger contribution of idiosyncratic, non-shared, health-related, and stochastic influences to cognitive-performance variance, then the alignment between g -loadings and between-group subtest differences should weaken. This is compatible with anti-Wilson-effect accounts of late-life cognition, in which the genetic structuring of g becomes increasingly disrupted by cumulative environmental insults (Sarraf et al., 2023, 2025). Age-related attenuation of SH therefore implies that such differences become less cleanly organized with respect to g , as non- g and environmentally contingent sources of variance accumulate.

Another potential mechanism is selective survival or differential morbidity in older cohorts. If selective mortality disproportionately removes lower-performing individuals from one group, the remaining sample may exhibit restricted variance on some subtests. Such a process would weaken the observed association between g -loadings and group-difference vectors, thereby attenuating SH, without necessitating a change in the latent source of the original group difference. These possibilities remain speculative, but they suggest a potentially fertile future research avenue in which cross-sectional age and mortality moderation effects on the contribution of test-specificities and error variance to the W-B difference could be examined in studies employing MGCFA. MGCFA conducted on environmentally stressed groups of Blacks and Whites (such as those from carceral or other environments exhibiting racialised i.e., racially, as opposed to socioeconomically correlated, sources of harshness, such as pollution contexts) could also decisively address the matter of the involvement (and types) of non- g variance in the generation of group differences under these conditions.

The finding that the magnitude, and even the sign of vector correlations vary across age-related contexts is inconsistent with the idea that SH is isomorphic to g (as proposed in its strongest form by Gordon, 1985).

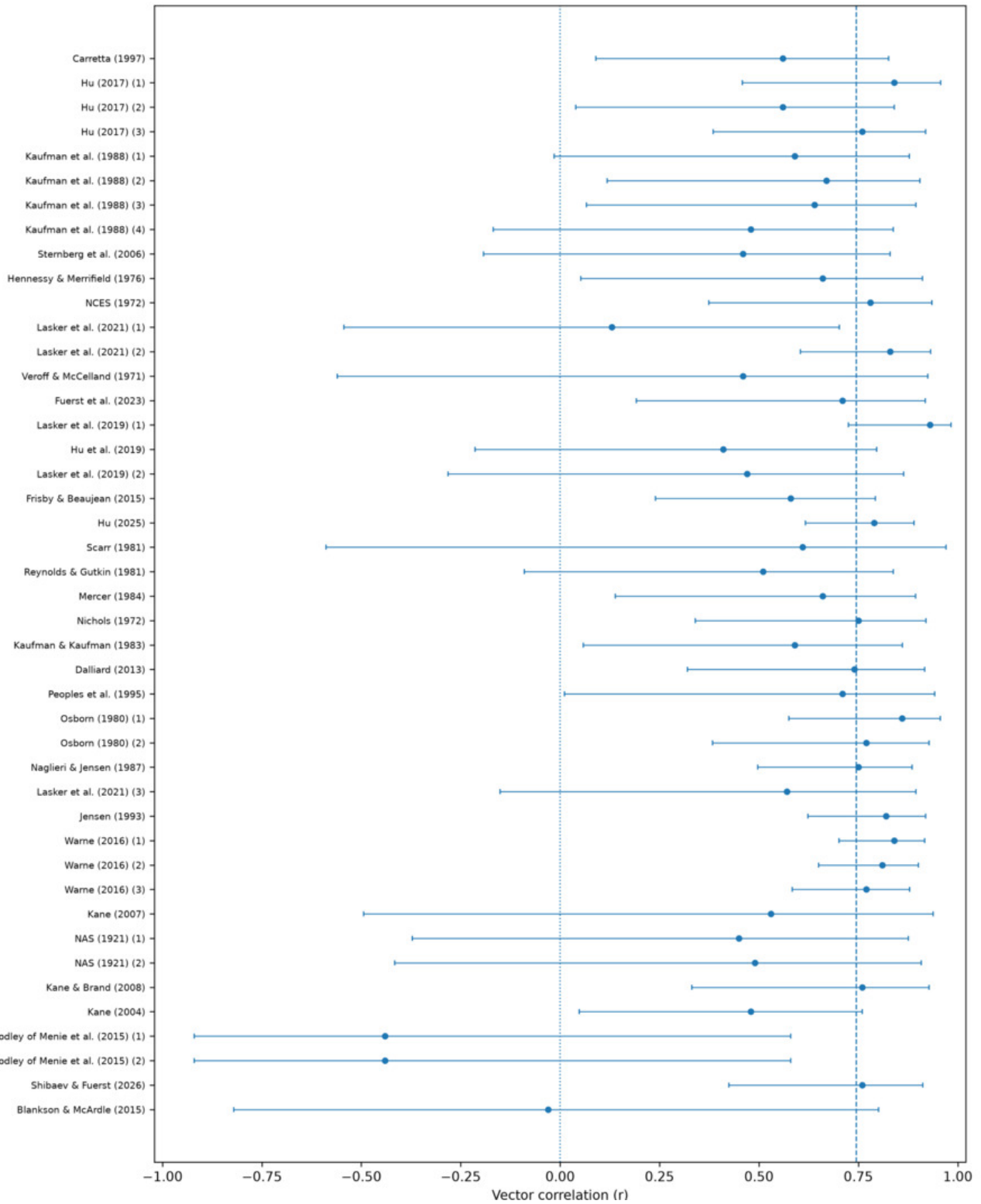


Figure 2. Forest plot of 44 independent SH effect sizes (scaled as vector correlations, r).

Table 1. Random-effects meta-analysis of 44 SH vector correlations.

Estimate	95% CI	p	τ^2	I^2	95% CI*
.745	.699 to .792	<.001	.003	25.61%	0.00% to 50.5%

*Estimated using a non-central χ^2 test.

Table 2. Mutually adjusted three-moderator analysis.

Moderator	Coding / scaling	B	SE	95% CI	p
Harmonic N	Per 100,000	.013	.017	-.021 to .048	.466
Mean age	Per year	-.003	.002	-.007 to -.000	.042
Publication year	Per decade (centred at 2000)	.019	.013	-.007 to .045	.149

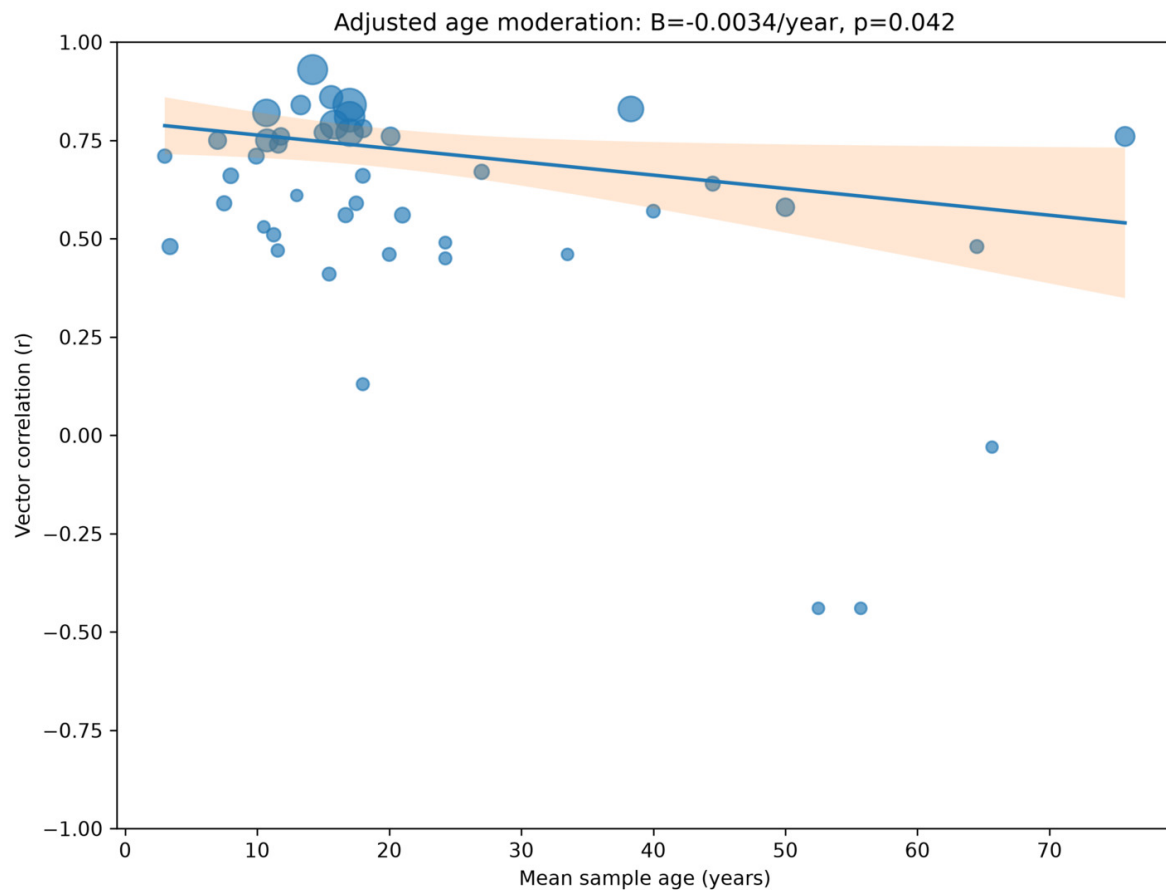


Figure 3. Bubble plot illustrating the negative effect of sample mean age on vector correlation magnitude. Bubble area is proportional to random-effects inverse-variance weight. The trend line and its associated confidence interval are from the mutually adjusted model.

Table 3. Three estimates of sampling error based on classic, CI-derived SE^2 , and Fisher-z scale transformed approaches.

Method	% variance due to sampling error
Classic Schmidt–Hunter using subtest N	46.3%
CI-derived SE^2	61.7%
Fisher-z scale, using $\frac{1}{N-3}$	82.7%

Limitations

The primary limitation is that the effect sizes were not disattenuated for error. Factors such as range-restriction with respect to the nomological breadth of g loadings, in addition to the reliabilities of the g and W - B vectors, and even the validity of the g vector will contribute to between-study heterogeneity independently of sampling error

(Jensen, 1998). When such corrections are applied, the magnitude of vector correlations will invariably increase (Jensen, 1998); however, overcorrection (where the vector correlation becomes >1.00) is a commonly encountered problem when these sorts of adjustments are applied (e.g., Woodley of Menie et al., 2018). This indicates that vector-level reliabilities have not yet been estimated with sufficient precision for full psychometric meta-analysis involv-

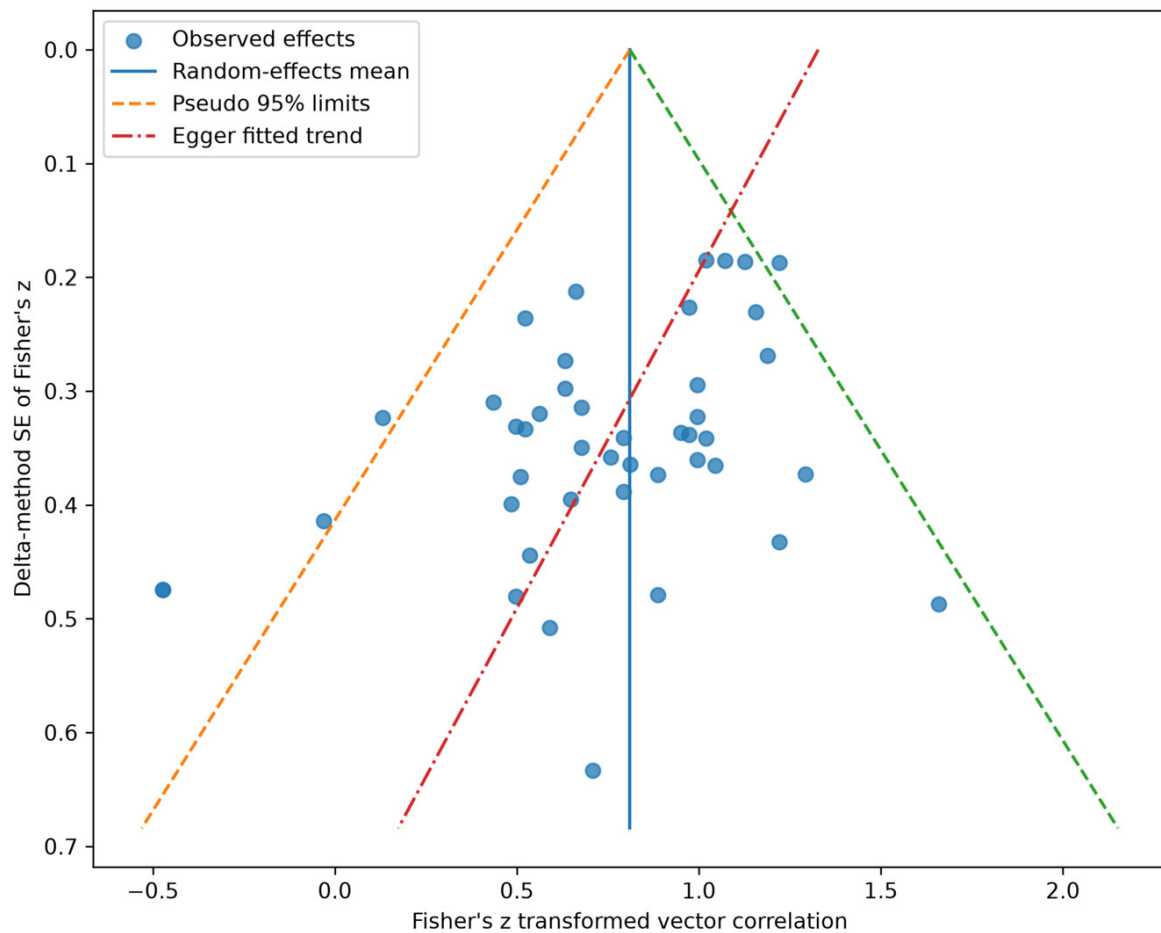


Figure 4. Funnel asymmetry plot of vector correlations with Egger's fitted trend line.

Table 4. Egger's regression terms estimated using unadjusted Fisher-z effects and moderator-adjusted Fisher-z meta-regression.

Egger asymmetry estimate	Estimate	SE	95% CI	<i>p</i>
Intercept	-1.687	.524	-2.744 to -.629	.002
Moderator adjusted SEz term	-1.195	.651	-2.511 to .121	.074

ing the results of studies applying MCV. Future research could attempt to meta-analyse the convergence amongst the relevant vectors to derive higher precision estimates of their associated reliabilities, making disattenuation less risky.

A secondary limitation stems from the fact that, as was noted in the Introduction, vector correlations can be distorted by subtest sampling in such a way as to yield misleading indications of *g* moderated differences (Ashton & Lee, 2005). It would therefore be useful to develop a heuris-

tic scale that could approximate the degree to which a battery was potentially unbalanced with respect to sampling across domains for use as an additional moderator in future meta-analytic investigations of SH.

Submitted: June 20, 2026 CDT. Accepted: September 18, 2026 CDT. Published: October 13, 2026 CDT.



References

- Ashton, M. C., & Lee, K. (2005). Problems with the method of correlated vectors. *Intelligence*, 33, 431–444. <https://doi.org/10.1016/j.intell.2004.12.004>
- Blankson, A. N., & McArdle, J. J. (2015). Measurement invariance of cognitive abilities across ethnicity, gender, and time among older Americans. *The Journals of Gerontology. Series B, Psychological Sciences & Social Sciences*, 70, 386–397. <https://doi.org/10.1093/geronb/gbt106>
- Bouchard, T. J., Jr. (2013). The Wilson effect: The increase in heritability of IQ with age. *Twin Research & Human Genetics*, 16, 923–930. <https://doi.org/10.1017/thg.2013.54>
- Carretta, T. R. (1997). Group differences on US Air Force pilot selection tests. *International Journal of Selection & Assessment*, 5, 115–127. <https://doi.org/10.1111/1468-2389.00051>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.
- Dalliard. (2013). *Spearman's Hypothesis and racial differences on the DAS-II*. Human Varieties. <https://humanvarieties.org/2013/12/08/spearmans-hypothesis-and-racial-differences-on-the-das-ii/>
- Deeks, J. J., Higgins, J. P. T., Altman, D. G., McKenzie, J. E., ... Veroniki, A. A. (2024). Chapter 10: Analysing data and undertaking meta-analyses. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, ... M. J. Page (Eds.), *Cochrane handbook for systematic reviews of interventions* (Version 6.5). Cochrane.
- Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56, 455–463. <https://doi.org/10.1111/j.0006-341X.2000.00455.x>
- Frisby, C. L., & Beaujean, A. A. (2015). Testing Spearman's hypotheses using a bi-factor model with WAIS-IV/WMS-IV standardization data. *Intelligence*, 51, 79–97. <https://doi.org/10.1016/j.intell.2015.04.007>
- Fuerst, G. R., Shibaev, V., & Kirkegaard, E. O. W. (2023). A genetic hypothesis for the American race/ethnic differences in mean g: A reply to Warne (2021) with fifteen new empirical tests using the ABCD dataset. *Mankind Quarterly*, 63, 527–600. <https://doi.org/10.46469/mq.2023.63.4.2>
- Gordon, R. A. (1985). The black-white factor is g. *Behavioral & Brain Sciences*, 8, 229–231. <https://doi.org/10.1017/S0140525X00020501>
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585, 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Hennessy, J. J., & Merrifield, P. R. (1976). A comparison of the factor structures of mental abilities in four ethnic groups. *Journal of Educational Psychology*, 68, 754–759. <https://doi.org/10.1037/0022-0663.68.6.754>
- Hu, M. (2017). *Modeling the Spearman's hypothesis using MGCEA: The Woodcock-Johnson data* [Unpublished manuscript].
- Hu, M. (2025). Spearman's g explains Black-White but not sex differences in cognitive abilities in the Project Talent. *OpenPsych*. <https://doi.org/10.26775/OP.2025.07.18>
- Hu, M., Lasker, J., Kirkegaard, E. O. W., & Fuerst, J. G. R. (2019). Filling in the gaps: The association between intelligence and both color and parent-reported ancestry in the National Longitudinal Survey of Youth 1997. *Psych*, 1, 240–261. <https://doi.org/10.3390/psych1010017>
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9, 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- Jensen, A. R. (1980). *Bias in mental testing*. Free Press.
- Jensen, A. R. (1985). The nature of the Black-White difference on various psychometric tests: Spearman's hypothesis. *Behavioral & Brain Sciences*, 8, 193–263. <https://doi.org/10.1017/S0140525X00020392>
- Jensen, A. R. (1992). Spearman's hypothesis: Methodology and evidence. *Multivariate Behavioral Research*, 27, 225–233. https://doi.org/10.1207/s15327906mbr2702_5
- Jensen, A. R. (1993). Spearman's Hypothesis tested with chronometric information-processing tasks. *Intelligence*, 17, 47–77. [https://doi.org/10.1016/0160-2896\(93\)90039-8](https://doi.org/10.1016/0160-2896(93)90039-8)
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Praeger.
- Kan, J.-K. (2012). *The Nature of Nurture: The role of gene-environment interplay in the development of intelligence* [Doctoral thesis]. University of Amsterdam.
- Kane, H. (2004). *Spearman's hypothesis: Implications for practice in a diverse society* [Oral presentation]. 5th Annual Meeting of the International Society for Intelligence Research, New Orleans.
- Kane, H. (2007). Group differences in nonverbal intelligence: Support for the influence of Spearman's g. *Mankind Quarterly*, 48, 65–82. <https://doi.org/10.46469/mq.2007.48.1.2>
- Kane, H., & Brand, C. (2008). Spearman's Hypothesis: Support from the Wechsler Intelligence Scale for Children, Third Edition. *Mankind Quarterly*, 49, 3–22. <https://doi.org/10.46469/mq.2008.49.1.1>
- Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman assessment battery for children*. American Guidance Service.
- Kaufman, A. S., McLean, J. E., & Reynolds, C. R. (1988). Sex, race, residence, region, and education differences on the 11 WAIS-R subtests. *Journal of Clinical Psychology*, 44, 231–248. [https://doi.org/10.1002/1097-4679\(198803\)44:2<231::AID-ICLP2270440224>3.0.CO;2-I](https://doi.org/10.1002/1097-4679(198803)44:2<231::AID-ICLP2270440224>3.0.CO;2-I)

- Kirkegaard, E. O. W., & Hu, M. (2025). Black-White differences in an English vocabulary test using an online Prolific sample. *OpenPsych*. <https://doi.org/10.26775/OP.2025.10.10>
- Klockars, A. J., & Sax, G. (1987). *Multiple Comparisons*. Sage.
- Lasker, J., Nyborg, H., & Kirkegaard, E. O. W. (2021). Spearman's Hypothesis in the Vietnam Experience Study and National Longitudinal Survey of Youth '79. *PsyArXiv*. <https://doi.org/10.31234/osf.io/m4yn9>
- Lasker, J., Pesta, B. J., Fuerst, J. G. R., & Kirkegaard, E. O. W. (2019). Global ancestry and cognitive ability. *Psych*, 1, 431–459. <https://doi.org/10.3390/psych1010034>
- McKinney, W. (2010). Data structures for statistical computing in Python. *Proceedings of the 9th Python in Science Conference*, 56–61. <https://doi.org/10.25080/Majora-92bf1922-00a>
- Menard, S. (1995). *Applied Logistic Regression Analysis*. SAGE Publications Inc.
- Mercer, J. R. (1984). What is a racially and culturally nondiscriminatory test? In C. R. Reynolds & R. T. Brown (Eds.), *Perspectives on Bias in Mental Testing. Perspectives on Individual Differences* (pp. 293–356). Springer. https://doi.org/10.1007/978-1-4684-4658-6_9
- Morris, D., Ritchie, S. J., & Plomin, R. (2025). Mechanisms of developmental change in genetic and environmental influences on intelligence. *PsyArXiv*. <https://doi.org/10.31234/osf.io/uqybg>
- Naglieri, J. A., & Jensen, A. R. (1987). Comparison of black-white differences on the WISC-R and the KABC: Spearman's hypothesis. *Intelligence*, 11, 21–43. [https://doi.org/10.1016/0160-2896\(87\)90024-9](https://doi.org/10.1016/0160-2896(87)90024-9)
- National Academy of Sciences. (1921). Distribution of scores on the tests in examinations alpha and beta. In R. M. Yerkes (Ed.), *Memoirs of the National Academy of Sciences, Vol. xv: Psychological examining in the United States Army* (pp. 873–875). Washington Government Printing Office.
- National Center for Education Statistics (NCES). (1972). *National Longitudinal Study of the High School Class of 1972 (NLS-72)*. U.S. Department of Education.
- Nichols, P. L. (1972). *The effects of heredity and environment on intelligence test performance in 4- and 7-year-old White and Negro sibling pairs* [Ph.D. dissertation]. University of Minnesota.
- Office of Management and Budget. (1997). Revisions to the standards for the classification of federal data on race and ethnicity. *Federal Register*, 62, 58782–58790.
- Osborne, R. T. (1980). *Twins: Black and White*. Foundation for Human Understanding.
- Peñaherrera-Aguirre, M., Woodley of Menie, M. A., Sarraf, M. A., & Beaver, K. M. (2022). Social adversity reduces polygenic score expressivity for general cognitive ability, but not height. *Twin Research & Human Genetics*, 25, 10–23. <https://doi.org/10.1017/thg.2022.3>
- Peoples, C. E., Fagan, J. F., III, & Drotar, D. (1995). The influence of race on 3-year-old children's performance on the Stanford-Binet: Fourth edition. *Intelligence*, 21, 69–82. [https://doi.org/10.1016/0160-2896\(95\)90039-X](https://doi.org/10.1016/0160-2896(95)90039-X)
- Plomin, R. (2024). Nonshared environment: Real but random. *JCPP Advances*, 4, e12229. <https://doi.org/10.1002/jcv2.12229>
- Reynolds, C. R., & Gutkin, T. B. (1981). A multivariate comparison of the intellectual performance of black and white children matched on four demographic variables. *Personality & Individual Differences*, 2, 175–180. [https://doi.org/10.1016/0191-8869\(81\)90021-0](https://doi.org/10.1016/0191-8869(81)90021-0)
- Roth, P. L., Bevier, C. G., Bobko, P., Switzer, F. S., III, & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology*, 54, 297–330. <https://doi.org/10.1111/j.1744-6570.2001.tb00094.x>
- Rushton, J. P. (1999). Secular gains in IQ not related to the g factor and inbreeding depression – unlike Black-White differences: A reply to Flynn. *Personality & Individual Differences*, 26, 381–389.
- Rushton, J. P., & Jensen, A. R. (2010). The rise and fall of the Flynn Effect as a reason to expect a narrowing of the Black-White IQ gap. *Intelligence*, 38, 213–219. <https://doi.org/10.1016/j.intell.2009.12.002>
- Sarraf, M. A., Woodley of Menie, M. A., & Peñaherrera-Aguirre, M. (2023). The anti-Wilson effect: The decrease in heritability of IQ, as proxied by polygenic score expressivity, with advanced age. *Personality & Individual Differences*, 202, 111969. <https://doi.org/10.1016/j.paid.2022.111969>
- Sarraf, M. A., Woodley of Menie, M. A., Peñaherrera-Aguirre, & Ujma, P. (2025). Gene expressivity on educational attainment increases with age, and then slows. *Personality & Individual Differences*, 232, 112850. <https://doi.org/10.1016/j.paid.2024.112850>
- Scarr, S. (1981). *Race, social class, and individual differences in I.Q.: New studies of old issues*. Lawrence Erlbaum Associates.
- Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.). SAGE Publications. <https://doi.org/10.4135/9781483398105>
- Schmidt, F. L., Hunter, J. E., & Raju, N. S. (1988). Validity generalization and situational specificity: A second look at the 75% rule and Fisher's z transformation. *Journal of Applied Psychology*, 73, 665–672. <https://doi.org/10.1037/0021-9010.73.4.665>
- Schooler, J. (2011). Unpublished results hide the decline effect. *Nature*, 470, 437. <https://doi.org/10.1038/470437a>
- Shibaev, V., & Fuerst, G. R. (2026). Spearman's Hypothesis tested between and within genetic ancestry groups in the HRS-HCAP sample. *Qeios*. <https://doi.org/10.23288/os5GAA>
- Spearman, C. (1927). *The abilities of man: Their nature and measurement*. Macmillan.

- Sternberg, R. J. & The Rainbow Project Collaborators. (2006). The Rainbow Project: Enhancing the SAT through assessment of analytical, practical, and creative skills. *Intelligence*, 34, 321–350. <https://doi.org/10.1016/j.intell.2006.01.002>
- te Nijenhuis, J., & van den Hoek, M. (2016). Spearman's hypothesis tested on Black adults: A meta-analysis. *Journal of Intelligence*, 4, 6.
- te Nijenhuis, J., van den Hoek, M., & Willigers, D. (2017). Testing Spearman's hypothesis with alternative intelligence tests: A meta-analysis. *Mankind Quarterly*, 57, 687–705. <https://doi.org/10.46469/mq.2017.57.4.12>
- Veroff, J., & McClelland, L. (1971). *Measuring intelligence and achievement motivation in surveys. Final report to U.S. Department of Health, Education and Welfare, Office of Economic Opportunity: Contract No. OEO-4180*. Survey Research Center, Institute for Social Research, University of Michigan.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., & SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17, 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
- Warne, R. T. (2016). Testing Spearman's hypothesis with advanced placement examination data. *Intelligence*, 57, 87–95. <https://doi.org/10.1016/j.intell.2016.05.002>
- Warne, R. T. (2021). Between-group mean differences in intelligence in the United States are >0% genetically caused: Five converging lines of evidence. *The American Journal of Psychology*, 134, 480–501. <https://doi.org/10.5406/amerjpsyc.134.4.0479>
- Woodley of Menie, M. A., Fernandes, H. B. F., Kanazawa, S., & Dutton, E. (2018). Sinistrality is associated with (slightly) lower general intelligence: A data synthesis and consideration of the secular trend in handedness. *HOMO: Journal of Comparative Human Biology*, 69, 118–126. <https://doi.org/10.1016/j.jchb.2018.06.003>
- Woodley of Menie, M. A., Figueredo, A. J., Madison, G., & Dunkel, C. S. (2015). Estimating the strength of genetic selection against g in a sample of 3520 Americans, sourced from MIDUS II. *Personality & Individual Differences*, 86, 266–270. <https://doi.org/10.1016/j.paid.2015.05.032>
- Woodley of Menie, M. A., Peñaherrera-Aguirre, M., & Sarraf, M. A. (2025). The reversal of Spearman's Hypothesis in incarcerated populations and the role of non-shared environmentality. *Intelligence & Cognitive Abilities*, 1, 86–95. <https://doi.org/10.65550/001c.140843>

Supplementary Materials

S1 - Meta-Analysis - Code (Zip File)

Download: <https://icajournal.scholasticahq.com/article/171909-a-meta-analysis-of-spearman-s-hypothesis-in-the-united-states/attachment/368967.zip>

S2 - Meta-Analysis - Data

Download: <https://icajournal.scholasticahq.com/article/171909-a-meta-analysis-of-spearman-s-hypothesis-in-the-united-states/attachment/368968.xlsx>
