

Research Articles

The Reversal of Spearman's Hypothesis in Incarcerated Populations and the Role of Non-Shared Environmentality

Michael A. Woodley of Menie¹, Mateo Peñaherrera-Aguirre², Matthew A. Sarraf³

¹ Independent Researcher, ² School of Animal and Comparative Biomedical Sciences, University of Arizona, ³ Independent Sector

Keywords: Environmental harshness, Jensen Effect, Spearman's Hypothesis, Unshared environmentality, Culture loading

<https://doi.org/10.65550/001c.140843>

Intelligence & Cognitive Abilities

Vol. 1, Issue 1, 2025

White-Black (W-B) cognitive performance differences tend to be more pronounced on more g -loaded subtests, a phenomenon posited by, and sometimes denoted by the term, Spearman's Hypothesis (SH). But SH might not be confirmed in all circumstances. The current study aims to investigate particular conditions under which it might fail. Exploratory factor analysis employing vectors of subtest-level effects as the unit of analysis is used to examine the relationships between non-shared environmentality (e^2), g and culture loadings, and W-B differences in two incarcerated and three normative samples, all evaluated using Wechsler series cognitive tests. Parallel analysis recommended a two-factor solution. Factor 1 was associated with negative clustering for the g and culture loading vectors, but positive clustering for both the incarcerated W-B difference and e^2 vectors. The three normative W-B difference vectors clustered with respect to Factor 2 onto which the g and culture loading vectors also weakly, but positively, cross-loaded. These results suggest that when racialized sources of environmental harshness are high (such as in incarceration settings) SH reverses, with e^2 variance (which captures environmental harshness) and non- g sources of performance variance being generative of these group differences.

Introduction

Spearman's Hypothesis (SH) is a term coined by Arthur Jensen (1923-2012) to describe the finding that tests with greater g saturation (i.e., those that better measure the general factor of intelligence) are more predictive of the magnitude of the mean performance difference between White and Black (W-B) socially identified racial groups on those tests (Jensen, 1985, 1998). This "hypothesis" was so named because this pattern had first been noted qualitatively by Charles Spearman (1863-1943) in his book *The Abilities of Man* (Spearman, 1927, p. 379).

SH has been mostly tested using Jensen's method of correlated vectors (MCV). This involves estimating the g saturation of subtests comprising a sufficiently diverse battery, then estimating the performance disparities across the subtests between the groups being compared (typically scaled in terms of standard deviation units, or d), then finally correlating the two effect size vectors, either using the Pearson or Spearman rank-order approach (Jensen, 1998). Employing 17 different studies examining W-B differences, sourced from Jensen (1985) and elsewhere, Jensen (1998) found strong evidence for SH ($r=0.63$). A more recently published formal meta-analysis of SH in 10 adult US W-B studies also found strong evidence for the effect ($\rho=0.57$, $K=15$ effect sizes) (te Nijenhuis & van den Hoek, 2016).

Figure 1 provides a visual illustration of MCV using data sourced from Jensen (1985). In Table 2 (p. 203) Jensen pre-

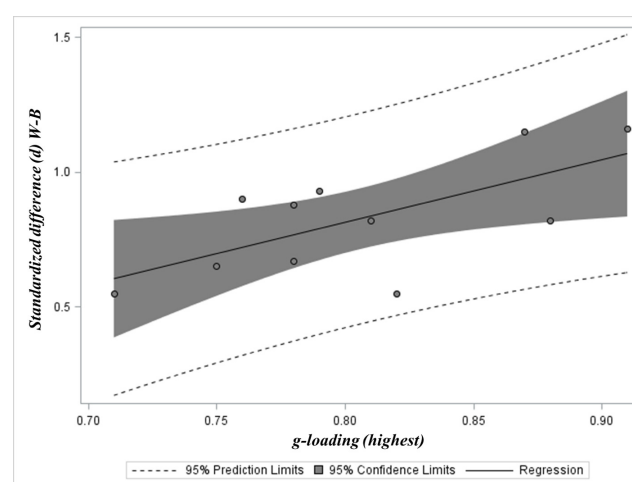


Figure 1. Scatter plot illustrating correlation (with 95% CIs) between vectors of standardized W-B differences (d) and g -loadings for 11 subtests.

sents g loadings for 11 subtests selected from batteries on the basis of their having the highest loadings and presents these along with their corresponding standardized W-B difference scores (difference between White and Black mean performance divided by subtest standard deviation).

The correlation between the two vectors in the above example is 0.667 (95% CI=0.112 to 0.904, $n=11$ subtests). Such positive correlations between vectors of g loadings

and other effects have been termed “Jensen Effects” (Rushton, 1998).

Evidence for SH has also been found in comparisons involving Whites and other socially identified racial and ethnic groups, such as Hispanics (te Nijenhuis et al., 2019), American Indians (te Nijenhuis, van den Hoek, et al., 2015), and European and Oriental Jews (te Nijenhuis, David, et al., 2014). Exceptions to this pattern have been identified in certain other ethnic group comparisons, however (e.g., te Nijenhuis et al., 2017).

MCV and its limitations

MCV has been criticized on several grounds. Some of these criticisms are flawed, such as the claim that it always yields “Jensen Effects” making confirmations of SH artefactual (as originally claimed by Schönemann, 1989). A range of phenomena do in fact show the opposite pattern, i.e., negative moderation by g , or an “anti-Jensen Effect.” Notable examples include adoption-related IQ gains (te Nijenhuis, Jongeneel-Grimen, et al., 2015), gains from educational interventions (such as Head Start; te Nijenhuis, Jongeneel-Grimen, et al., 2014), and Flynn-Effect gains (te Nijenhuis & van der Flier, 2013).

A more fundamental criticism concerns whether or not the latent variable involved in these correlations can be reliably identified with g (Dolan, 2000). Spearman's (1927) *Principle of the Indifference of the Indicator* suggests that it is generally safe to assume that arbitrarily selected subtests can be reliably ranked in terms of g saturation, owing to the apparent ubiquity of this latent variable across diverse cognitive ability measures and populations (Warne & Burningham, 2019). However, unbalanced sampling of subtests with respect to narrower cognitive domains can yield deceptive results, where non- g group factors might be driving the moderation patterns instead (Ashton & Lee, 2005). Jensen (1998) anticipated this limitation, however, and provided guidelines to avoid misleading results through balanced sampling of subtests so as to ensure adequate nomological breadth. It should be noted also that MCV *cannot* be used to determine the degree to which two populations exhibit measurement invariance with respect to a given latent variable, although the method clearly *assumes* invariance. The finding of a “Jensen Effect” in a test of SH might give the false impression that two populations differ on a common latent factor, when in fact there may be configural and/or other differences between the factors, when estimated for each group separately, making them incommensurable. But alternative methods, such as multi-group confirmatory factor analysis, can be used to directly address this important issue (e.g., Dolan, 2000).

J. Philippe Rushton's method and the causes of SH

A key question is what causes SH. By far the most controversial application of MCV has been in its use as a line of evidence for the view that SH supports a role for genetic or biological influences on W-B, and other socially identified racial and ethnic group, differences (Warne, 2021), the idea

that there are such influences being the racial-hereditary hypothesis (RHH). Rushton (1999), who was a major proponent of RHH, developed a method involving factor analysis where vectors (rather than subjects) were used as the unit of analysis and which attempted to address the topic of causation in SH. Rushton (1999) sourced vectors of various effect sizes reflecting both genetic and environmental influences on subtest performance in addition to g loadings and W-B differences. It was argued that clustering among these vectors might yield evidence for the RHH if a positive manifold emerges among vectors of genetic influences, g loadings, and W-B differences, but excludes environmental ones. In testing this, Rushton (1999) conducted principal component analysis on a correlation matrix generated using a set of nine vectors, reflecting subtest-level W-B differences, two sets of g loadings (from both WISC-III and R), inbreeding depression effects, and four sets of Flynn-Effect vectors, all estimated with respect to the subtests of Wechsler series tests. His analysis yielded two factors, one onto which loaded the vectors of g loadings, W-B differences, and inbreeding depression, and another, onto which loaded the four Flynn-Effect vectors. Rushton argued that this indicated that some portion of the W-B difference is a function of genetic effects (as indexed in this specific instance by their clustering with the inbreeding-depression-effects vector).

Subsequent research has challenged Rushton's interpretation. Most notably Kan (2011) conducted a comprehensive re-analysis of the data in Jensen (1985) and elsewhere, finding that while positive correlations exist among vectors of W-B differences, g loadings, and subtest heritability estimates, much of the variance in SH can be accounted for by the *culture loadings* of these subtests, indicating that performance on them is dependent on culture-bound acquired knowledge, which is of course related to environmental effects, e.g. upbringing and education (Freese & Jao, 2017). A recent application of Rushton's method revealed that the degree to which experiences of racial and ethnic discrimination reduce the expressivity of polygenic scores associated with educational attainment on g is also a “Jensen Effect,” exhibiting positive clustering with vectors of g loadings, SH (estimated using three-way comparisons of Black, White, and Hispanic performance differences), and shared environmental (c^2) effects on IQ (Peñaherrera-Aguirre et al., 2022). These findings are all broadly consistent with Kan's thesis that cultural factors, especially those that act via shared environmentality, can be a major source of SH, weakening hereditary interpretations of these patterns.

The current study

Non-shared environmental effects (e^2) reflect influences on subjects that reduce phenotypic resemblance among relatives (Plomin, 2024). e^2 effects on cognitive performance are generally stronger than ones stemming from the shared environment (Bouchard, 2013), with the latter being more potent early in life, and weakening as heritability increases with age (the increase in heritability with age being the Wilson Effect; Bouchard, 2013).

How might e^2 affect SH? e^2 , which is a within-group parameter, and by extension its between-group counterpart (e_b^2), are sources of phenotypic dissimilarity. As these sources of environmentality are random, having effects that are very similar to measurement error (Plomin, 2024), e^2 should therefore attenuate correlation patterns among phenotypes, weakening g saturation in instances where it is most pronounced. In cases where e_b^2 is the primary cause of W-B differences, i.e., where there exist differences in the levels of e^2 between groups, we would expect therefore to find a failure of SH, with more g -loaded subtests being *less* predictive of W-B differences. A plausible source of e_b^2 is environmental harshness, a dimension of environmental risk characterized by high variance in both actual and perceived morbidity and mortality (Ellis et al., 2009). It has been speculated that humans may possess an innate repertoire of harshness-adapted cognitive abilities (so called hidden talents; Ellis et al., 2023) which are ordinarily difficult to measure suggesting that their expression is highly situation-dependent and therefore they do not (generally) participate in the positive manifold. Group differences resulting from the differential expression of these will, contra SH, not manifest as differences in g .

Consistent with this model, a recent study by Sng and colleagues (2024) found that subject perceptions of social race are confounded with perceptions of environmental harshness, such that Black social-racial attribution becomes conflated with harsher environments. This is consistent with biosocial models of race in which social race acquires tangible biological qualities by virtue of, e.g., inequitable niche construction (Henry et al., 2023), and suggests that disparities in environmental harshness might function analogously to a racialized *Factor X*, a hypothesized source of variance that acts between groups to generate differences in the levels of a trait, but has no (or rather no directionally consistent) effect on the levels of that trait within groups (Jensen, 1973).

A line of evidence consonant with the existence of a *Factor X* in group differences comes from meta-analytic studies of SH in incarcerated populations (i.e., those who had been institutionalized for “delinquent behavior” or who were imprisoned), where it has been found that SH is inverted (the W-B difference tends to be an “anti-Jensen Effect”) (te Nijenhuis et al., 2017; van den Hoek, 2015). Incarceration environments, such as those encountered in prisons, are associated with elevated levels of harshness, stemming from adverse interactions among inmates and possibly other personnel also, in the form of violence and victimization patterns that strongly track with socially identified race (Carroll, 1974). Ross and Richards (2002) describe a phenomenon termed the “color line” in prison settings, involving pervasive racialized patterns of both collective (e.g., riots, ganging) and interpersonal violence (see also Bloch & Olivares-Pelayo, 2023). The salience of social race is therefore massively amplified in such populations, creating opportunities for the emergence of disparate and strongly racialized environmental impacts on cognitive functioning.

In the current study the possibility that these reversals of SH in the context of incarcerated populations track with subtest sensitivity to e^2 will be tested using vectors of W-B differences sourced from both incarcerated and normative samples evaluated using Wechsler series tests. Rushton's (1999) factor-analytic method will be employed to identify salient patterns of clustering. The following hypotheses will be tested:

H1: Strong positive clustering should be observed among vectors of e^2 and the incarcerated W-B difference vectors.

H2: Strong positive clustering should be observed among vectors of g and culture loadings, and normative W-B differences (consistent specifically with predictions made in Kan, 2011 concerning the cause of SH).

H3: These clustering patterns will form distinct factors.

Methods

Data

The vectors for each variable were sourced from Wechsler series tests in all cases.

1. *Normative W-B difference data:* For the normative samples, one of the most extensive studies of W-B differences at the subtest level is that of Kaufman, McLean, and Reynolds (1988), who present data on W-B differences by WAIS-R subtest for four different age ranges (16-19, 20-34, 35-54, and 55-74) (see Table 2).
2. *Incarcerated W-B difference data:* SH in the context of incarcerated populations was meta-analyzed by van den Hoek (2015), who identified four relevant effect sizes (Destephens, 1953; Jensen & Faulstich, 1988; Panton, 1960; and one hybrid effect size derived from combining two separate studies with comparable samples: Altus & Clark, 1949, and Franklin, 1945). The effect sizes derived from the oldest set of studies (Altus & Clark, 1949 + Franklin, 1945 and Destephens, 1953) were estimated using 10 subtests of the Wechsler-Bellevue test (without vocabulary), a precursor to WAIS. Altus and Clark (1949) report subtest performance means for a White subsample aged 14 to 18 who had been institutionalized for “juvenile delinquency.” Franklin (1945) provides subtest-level data on a comparable Black institutionalized sample aged from 10 to 20. Panton (1960) used WAIS and Jensen and Faulstich (1988) employed WAIS-R, in both cases on prison populations. In three of the four studies, SH was disconfirmed, with the W-B difference being an “anti-Jensen Effect” (range $r=-0.03$ to -0.46). In Panton (1960) the effect was positive ($r=0.73$), although van den Hoek (2015) noted that this study was of “low quality” because the participants had prior exposure to the test. On the basis of this, the effect size was excluded from subsequent analysis. Relevant subtest-level W-B effect sizes were

recomputed for the remaining studies using the original data in these papers.

3. *Non-shared environmentality*: Estimates of subtest e^2 variance were derived from two studies. Rijdsdijk, Vernon, and Boomsma (2002) generated estimates of behavior-genetic variance components for each of the 11 WAIS subtests (Table 3). e^2 variances were recomputed directly from the monozygotic and dizygotic twin correlations using Falconer's equation (after first estimating h^2 and c^2) as follows:

$$e^2 = 1 - (h^2 + c^2)$$
 WAIS subtest-level estimates of e^2 are also provided in the study of Tambs, Sundet, and Magnus (1984, Table 2).
4. *g loadings*: Reliability-corrected meta-analytic WAIS g loadings were obtained from Kan's (2011) doctoral thesis (Table 3.7).
5. *Culture loadings*: Culture loading estimates were obtained for each of the WISC-III subtests from Table 3.1 in Kan (2011). Most of these values came from the study of Georgas, van de Vijver, Weiss, and Saklofske (2003, Table 18.1), with the value for Coding Speed having been obtained by Kan via personal communication with one of the authors of this paper.

The vocabulary subtest was dropped in all cases where effect sizes had been estimated using WAIS, WAIS-R, or WISC-III bringing the subtests into alignment with those available for the samples using the Wechsler-Bellevue test, yielding an n of 10 subtests per vector.

Statistical analyses

The youngest age group (16-19) from Kaufman et al. (1988) and the prison sample from Jensen and Faulstich (1988) had to be excluded as the dataset that included them was non-factorizable. Bartlett's test, which is a more robust indicator of factorizability than the Kaiser-Meyer-Olkin test when case number is low (Johnson & Wichern, 2007), was applied to determine the suitability of the remaining nine vectors for exploratory factor analysis (EFA). This analysis was conducted using the *cortest.bartlett* function (*psych* package; Revelle, 2015). Horn's parallel analysis was also computed with the *fa.parallel* function to identify the number of latent dimensions (*psych* package; Revelle, 2015). Lastly, a Principal Axis EFA was estimated, fixing the number of latent dimensions based on the results of the aforementioned parallel analysis. This model was estimated using the *fa* function (*psych* package; Revelle, 2015). All statistical analyses were performed using the R program (v.4.3.1). All code used in these analyses is made available in the Appendix. All subtest-level data on all vectors used in the current analysis, along with a correlation matrix, are presented in [Table 1](#).

Results

Measurement models

Bartlett's test reached statistical significance ($\chi^2 = 116.167$, $p < .0001$), with Horn's parallel analysis recommending the extraction of two dimensions. A Principal Axis EFA was estimated by fixing the model to a two-factor solution. *Factor 1* exhibited negative loadings for the g -loading vector sourced from Kan (2011) and also the culture loading vector sourced from Georgas et al. (2003). In contrast, *Factor 1* exhibited positive loadings for the two subtest non-shared environmentality vectors sourced from Tambs et al. (1984) and Rijdsdijk et al. (2002). Similarly, it was associated with positive loadings for the prison sample W-B difference vectors obtained from DeStephens (1953), and Franklin (1945) + Altus and Clark (1949). *Factor 2* exhibited positive loadings for the three normative sample W-B vectors sourced from Kaufman et al. (1988). All loadings are reported in [Table 2](#). [Figure 2](#) illustrates the relevant factor pattern.

Discussion

Consistent with H1, positive clustering with respect to *Factor 1* was observed among the two e^2 and the two incarcerated W-B difference vectors, which suggests that e^2 may be causal of these differences in this context. Consistent in part with H2, positive clustering was also observed among the three vectors of normative W-B differences, forming a distinct *Factor 2*. However, the vectors of g and culture loadings negatively co-loaded onto *Factor 1*, but cross-loaded only weakly (but positively) ($\lambda = 0.259$ and 0.160 respectively) onto *Factor 2*, which is not consistent with H2. Moreover, the W-B difference vectors also cross-loaded onto *Factor 2* ($\lambda = 0.636$ and 0.380 for De Stephens and Franklin + Altus & Clark respectively), which suggests that there is some shared etiology between the normative and incarcerated W-B vectors, but only with respect to influences that are independent of g and culture loadings (by contrast the three normative W-B vectors all cross-load negatively onto *Factor 1*). The patterns of associations among the vectors comprising *Factor 1*, suggest that this factor might accurately be characterized as an *anti-Spearman Hypothesis* factor, with g and culture loading playing a central role in negatively co-moderating the magnitude of the incarceration group differences and associated (non-shared) environmental influences. Consistent with H3 the two factors were effectively orthogonal, exhibiting only a very weak negative inter-factor correlation ($r = -0.032$).

Within an incarceration context, sources of environmental harshness likely become highly racialized (as per the "color line" model; Ross & Richards, 2002). This could translate into group differences via amplification of the effects of broader synchronic socio-structural sources of inequality on incarcerated populations that generates acute Factor X-type effects (specifically, strongly discontinuous racialized between-group variation in the *intensity* of harshness patterns). Consistent with this, Sonderlund et al. (2022) have theorized that there are "pathways that fo-

Table 1. Subtest-reliability-corrected g loadings, culture loadings, W-B differences (White total score minus Black total score divided by the pooled standard deviation [d]), and two estimates of subtest-level e^2 , along with a correlation matrix of all vectors.

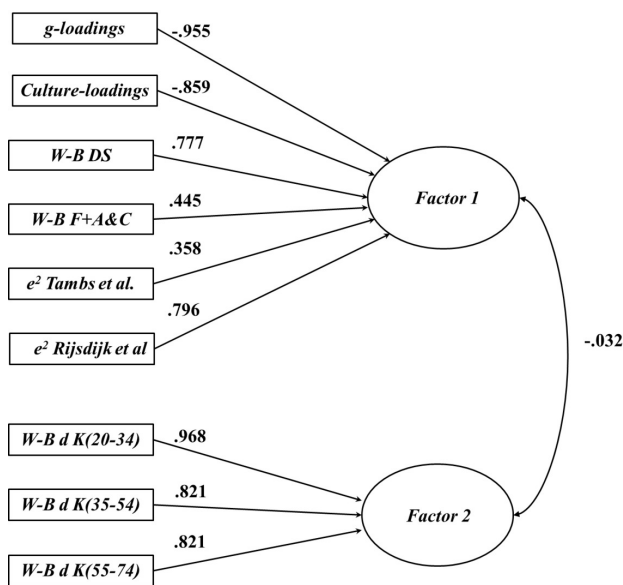
WAIS Subtest	g - loadings (1)	Culture- loadings (2)	W-B d K (20-34) (3)	W-B d K (35-54) (4)	W-B d K (55-74) (5)	e^2 Rijdsdijk et al. (6)	e^2 Tambs al. (7)	W-B F+A&C (8)	W-B DS (9)
Information	0.89	0.22	0.83	0.80	0.83	0.20	0.26	1.36	0.18
Comprehension	0.89	0.15	0.72	0.86	0.84	0.14	0.21	0.61	0.09
Coding	0.61	0.00	0.44	0.67	0.71	0.56	0.23	0.91	0.26
Arithmetic	0.80	0.08	0.77	0.89	0.89	0.26	0.33	0.57	0.28
Similarities	0.89	0.09	0.69	0.81	0.68	0.00	0.27	0.49	0.12
Digit Span	0.63	0.00	0.45	0.58	0.50	0.36	0.39	0.89	0.22
Picture Completion	0.70	0.03	0.89	0.71	0.70	0.72	0.56	1.22	0.33
Block Design	0.72	0.01	1.16	0.95	1.00	0.33	0.31	1.17	0.53
Picture Arrangement	0.75	0.02	0.55	0.62	0.51	0.18	0.42	0.71	0.21
Object Assembly	0.65	0.00	0.71	0.82	0.75	0.50	0.29	1.13	0.38
1	1.000								
2	0.872	1.000							
3	0.303	0.213	1.000						
4	0.482	0.363	0.772	1.000					
5	0.332	0.352	0.779	0.929	1.000				
6	-0.777	-0.550	0.023	-0.269	-0.048	1.000			
7	-0.337	-0.401	0.105	-0.433	-0.412	0.486	1.000		
8	-0.324	-0.003	0.422	0.006	0.214	0.585	0.235	1.000	
9	-0.564	-0.573	0.577	0.315	0.416	0.558	0.272	0.551	1.000

Note. K: Kaufman et al.; F+A&C: Franklin + Altus & Clark; DS: DeStephens. $p \leq .05$ for correlation values ≥ 0.631 for an n of 10.

Table 2. Factor structure with two latent dimensions. Preferential loading of indicators is indicated in bold.

Variable	Factor 1	Factor 2	h^2	u^2
<i>g-loadings</i>	-0.955	0.259	0.995	0.005
<i>Culture-loadings</i>	-0.859	0.160	0.772	0.228
e^2 Rijdsdijk et al.	0.796	-0.007	0.634	0.366
W-B DS	0.777	0.636	0.975	0.025
W-B F+A&C	0.445	0.380	0.331	0.669
e^2 Tambs et al.	0.358	0.007	0.128	0.872
W-B d K (20-34)	-0.021	0.968	0.939	0.061
W-B d K (35-54)	-0.102	0.821	0.690	0.310
W-B d K (55-74)	-0.255	0.821	0.752	0.248

Note. h^2 : Commonality; u^2 : Unique variance. K: Kaufman et al.; F+A&C: Franklin + Altus & Clark; DS: DeStephens. $p \leq .05$ for λ values ≥ 0.631 for an n of 10.

**Figure 2. Two-factor EFA solution.**

K: Kaufman et al.; F+A&C: Franklin + Altus & Clark; DS: DeStephens. $p \leq .05$ for λ values ≥ 0.631 for an n of 10 in each case. Cross-loadings not shown.

cus on social cohesion and community-level chronic stress, community crime and police victimization, as well as infra-structural community disinvestment” that link “spatial and racialized clustering of incarceration ... and disparities in adverse birth outcomes” (p. 1). Incarceration mechanisms, by virtue of their intergenerational persistence as part of the ecological inheritance system, are also clear sources of *inequitable niche construction*, whereby a population’s experience of ecological harshness and concomitant adverse health (and cognitive) disparities are diachronically conditioned by ecological legacy effects stemming from the persistence of these inherited institutions (Henry et al., 2023).

Pathways linking broader socio-cultural factors to disparate outcomes among incarcerated populations might account for the observation that, among the incarcerated W-B populations, there still are aggregate cognitive performance disparities in the commonly found direction (i.e., White > Black). Although the magnitude of these disparities is smaller in the case of the incarcerated populations (av-

erage d across subtests = .58 vs. .74 in the normative samples), perhaps reflecting a range-restriction effect stemming from the general negative association between IQ and contact with the criminal justice system (Hirschi & Hindelang, 1977).

In what ways might incarcerated W-B populations differ cognitively? As mentioned in the Introduction, the capacity to acquire certain sets of harshness-adapted cognitive skills, which are poorly indexed by conventional cognitive tests, such as those comprising the hidden talent framework, may play a critical role. Hidden talents in this context describe “stress-adapted skills [that] represent a form of adaptive intelligence enabling individuals to function within the constraints of harsh environments” (Ellis et al., 2023). Chronic exposure to harshness may elicit these narrower skills among incarcerated populations as *evoked adaptations*, which essentially “blossom” at the expense of the integrity of the g factor. Ability tilts (difference scores resulting from tradeoffs between two abilities) might be yet another manifestation of non- g hidden talents, differing in magnitude in White relative to Black populations (Coyle, 2016). A long-established theory maintains that incarcerated populations (relative to non-incarcerated ones) may exhibit a tilt toward spatial abilities (Matarazzo, 1972), so perhaps this might be a source of W-B differences in the incarceration context. In testing Matarazzo’s model, Jensen and Faulstich (1988) noted, however, that the effect was extremely weak once the difference in g between incarcerated and normative populations had been controlled. Another possibility is that the performance disparity is a function of differences in error rates, which would also be consistent with an effect of e^2 on this disparity, as error variance is folded into measures of e^2 (Plomin, 2024). Error in this context could also be an adaptive function of factors such as reduced attentional capacity in response to sources of heightened environmental harshness.

Thus, there are potentially many axes of non- g variance along which W-B populations could diverge when exposed to different levels of racialized environmental harshness. Being multivariate, such extra cognitive “degrees of freedom” could easily lead to non-invariant differences between groups, whereby individuals within such groups might employ a large range of effectively idiosyncratic cog-

nitive strategies in solving test items. Thus, the W-B difference in the incarceration context is likely to be one of diverse *kinds* of cognitive strategies, rather than one of *degrees* of difference with respect to some common factor variance.

It should be noted that the studies reporting data on the two incarceration samples used currently are seven to eight decades old. Studies of such populations *in situ* are extremely rare. It would be ideal for more data of this sort to be obtained using larger, contemporary incarcerated populations to replicate the finding of an anti-SH effect in them. Populations of W-B individuals having to cope with other sources of elevated harshness could also be considered in future research. One potentially relevant set of environments are those with high levels of pollution, where it has been found that social race, rather than socio-economic status, is the more robust predictor of risk of living in such an environment (Lancet, 2018), suggesting that such settings may constitute another regime of racialized environmental harshness in which SH will fail. Potentially consistent with this is the observation that pollution affects subtest scores in a way that is typically associated with weak “Jensen” or “anti-Jensen Effects”. Woodley of Menie, Sarraf, Peñaherrera-Aguirre, Fernandes, and Becker (2018) conducted a small meta-analysis of the results of MCV applied to the impacts of various toxic substances on the

pattern of subtest performance. These included lead, mercury, fetal alcohol and cocaine, and atmospheric pollution. The overall effect was found to be consistent with a weak “Jensen Effect” ($\rho=0.11$, 95% CI=0.08 to 0.14, $n=3841$). This indicates that much of the effects of pollutants and other toxic substances likely occur on non-*g* sources of subtest variance which is consistent with the speculation that polluted environments may at the very least significantly attenuate, or even reverse SH among Black and White populations exposed to them.

The current findings fit with the existence of environmental effects on the W-B difference, but only insofar as the patterns of co-moderation are compatible with an inference of causation. This finding highlights the importance of making *clear testable predictions* concerning potential causal influences on these differences. Proponents of racial-environmental hypotheses (REH) frequently fail to make predictions of a sort that would substantiate or falsify their claims, and in some cases even suggest that such predictions cannot be made (e.g. Kaplan, 2015).

Submitted: April 11, 2025 CST. Accepted: June 10, 2025 CST.

Published: June 30, 2025 CST.



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at <http://creativecommons.org/licenses/by/4.0> and legal code at <http://creativecommons.org/licenses/by/4.0/legalcode> for more information.

References

- Altus, W. D., & Clark, J. H. (1949). Subtest variation on the Wechsler-Bellevue for two institutionalized behavior problem groups. *Journal of Consulting Psychology*, 13, 444–447. <https://doi.org/10.1037/h0058312>
- 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>
- Bloch, S., & Olivares-Pelayo, E. A. (2023). Carceral racialization, prison segregation, and the Integrated Housing Program in Arizona. *Theoretical Criminology*, 28, 212–231. <https://doi.org/10.1177/13624806231179127>
- 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>
- Carroll, L. (1974). *Hacks, Blacks, and Cons: Race relations in a maximum security prison*. Lexington Books.
- Coyle, T. (2016). Ability tilt for whites and blacks: Support for differentiation and investment theories. *Intelligence*, 56, 28–34. <https://doi.org/10.1016/j.intell.2016.02.002>
- Destephens, W. P. (1953). Are criminals morons? *The Journal of Social Psychology*, 38, 187–199. <https://doi.org/10.1080/00224545.1953.9711454>
- Dolan, C. V. (2000). Investigating Spearman's Hypothesis by Means of Multi-Group Confirmatory Factor Analysis. *Multivariate Behavioral Research*, 35, 21–50. https://doi.org/10.1207/S15327906MBR3501_2
- Ellis, B. J., Abrams, L. S., Masten, A. S., Sternberg, R. J., Tottenham, N., & Frankenhuys, W. E. (2023). *The hidden talents framework*. Cambridge Elements. <https://doi.org/10.1017/9781009350051>
- Ellis, B. J., Figueredo, A. J., Brumbach, B. H., & Schlomer, G. L. (2009). Fundamental dimensions of environmental risk. *Human Nature*, 20, 204–268. <https://doi.org/10.1007/s12110-009-9063-7>
- Franklin, J. C. (1945). Discriminative value and patterns of the Wechsler-Bellevue scales in the examination of delinquent Negro boys. *Educational & Psychological Measurement*, 5, 71–85. <https://doi.org/10.1177/001316444500500106>
- Freese, J., & Jao, Y. H. (2017). Shared environment estimates for educational attainment: A puzzle and possible solutions. *Journal of Personality*, 85, 79–89. <https://doi.org/10.1111/jopy.12226>
- Georgas, J., van de Vijver, F. J. R., Weiss, L. G., & Saklofske, D. H. (2003). A cross-cultural analysis of the WISC-III. In J. Georgas, L. G. Weiss, F. J. R. van de Vijver, & D. H. Saklofske (Eds.), *Culture and children's intelligence: Cross-cultural analysis of the WISC-III* (pp. 277–313). Academic Press.
- Henry, P. I., Beaulieu, M. R. S., Bradford, A., & Graves, J. L., Jr. (2023). Embedded racism: Inequitable niche construction as a neglected evolutionary process affecting health. *Evolution, Medicine, & Public Health*, 11, 112–125. <https://doi.org/10.1093/emph/eoad007>
- Hirschi, T., & Hindelang, M. J. (1977). Intelligence and delinquency: a revisionist review. *American Sociological Review*, 42, 571–587. <https://doi.org/10.2307/2094556>
- Jensen, A. R. (1973). *Educability and Group Differences*. Harper & Row.
- 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. (1998). *The g factor: The science of mental ability*. Praeger.
- Jensen, A. R., & Faulstich, M. E. (1988). Difference between prisoners and the general population in psychometric g. *Personality & Individual Differences*, 9, 925–928. [https://doi.org/10.1016/0191-8869\(88\)90013-X](https://doi.org/10.1016/0191-8869(88)90013-X)
- Johnson, R. A., & Wichern, D. W. (2007). *Applied multivariate statistical analysis* (6th ed.). Pearson.
- Kan, J.-K. (2011). *The Nature of Nurture: The role of gene-environment interplay in the development of intelligence* [Doctoral thesis]. University of Amsterdam.
- Kaplan, J. M. (2015). Race, IQ, and the search for statistical signals associated with so-called “X”-factors: environments, racism, and the “hereditarian hypothesis.” *Biology & Philosophy*, 30, 1–17. <https://doi.org/10.1007/s10539-014-9428-0>
- 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](https://doi.org/10.1002/1097-4679(198803)44:2)
- Lancet. (2018). Environmental racism: Time to tackle social injustice. *The Lancet: Planetary Health*, 2, e462.
- Matarazzo, J. D. (1972). *Wechsler's measurement and appraisal of adult intelligence* (5th ed.). Williams & Wilkins.
- Panton, J. H. (1960). Beta-WAIS comparisons and WAIS subtest configurations within a State prison population. *Journal of Clinical Psychology*, 16, 312–317. [https://doi.org/10.1002/1097-4679\(196007\)16:3](https://doi.org/10.1002/1097-4679(196007)16:3)
- 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>
- Plomin, R. (2024). Nonshared environment: Real but random. *JCPP Advances*, 4, e12229. <https://doi.org/10.1002/jcv2.12229>
- Revelle, W. (2015). *psych: Procedures for psychological, psychometric, and personality research*. Northwestern University.

- Rijsdijk, F. V., Vernon, P., & Boomsma, D. I. (2002). Application of hierarchical genetic models to Raven and WAIS subtests: A Dutch twin study. *Behavior Genetics*, 32, 199–210. <https://doi.org/10.1023/A:1016021128949>
- Ross, J. I., & Richards, S. C. (2002). *Behind bars: Surviving prison*. Alpha Books.
- Rushton, J. P. (1998). The “Jensen Effect” and the “Spearman-Jensen Hypothesis” of Black-White IQ differences. *Intelligence*, 26, 217–225. [https://doi.org/10.1016/S0160-2896\(99\)80004-X](https://doi.org/10.1016/S0160-2896(99)80004-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.
- Schönemann, P. H. (1989). Some new results on the Spearman hypothesis artifact. *Bulletin of the Psychonomic Society*, 27, 462–464. <https://doi.org/10.3758/BF03334656>
- Sng, O., Boyd-Frenkel, K. A., & Williams, K. E. G. (2024). Can race be replaced? Ecology and race categorization. *Evolution & Human Behavior*, 45, 106630. <https://doi.org/10.1016/j.evolhumbehav.2024.106630>
- Sonderlund, A. L., Charifson, M., Oritz, R., Khan, M., Schoenthaler, A., & Williams, N. J. (2022). A comprehensive framework for operationalising structural racism in health research: The association between mass incarceration of Black people in the U.S. and adverse birth outcomes. *SSM – Population Health*, 18, 101225.
- Spearman, C. (1927). *The abilities of man*. Macmillan.
- Tambs, K., Sundet, J. M., & Magnus, P. (1984). Heritability analysis of the WAIS subtests: A study of twins. *Intelligence*, 8, 283–293. [https://doi.org/10.1016/0160-2896\(84\)90012-6](https://doi.org/10.1016/0160-2896(84)90012-6)
- te Nijenhuis, J., David, H., Metzen, D., & Armstrong, E. L. (2014). Spearman's hypothesis tested on European Jews vs non-Kewish Whites and vs Oriental Jews: Two meta-analyses. *Personality & Individual Differences*, 44, 15–18.
- te Nijenhuis, J., Jongeneel-Grimen, B., & Armstrong, E. L. (2015). Are adoption gains on the g factor? A meta-analysis. *Personality & Individual Differences*, 73, 56–60. <https://doi.org/10.1016/j.paid.2014.09.022>
- te Nijenhuis, J., Jongeneel-Grimen, B., & Kirkegaard, E. O. W. (2014). Are Headstart gains on the g factor? A meta-analysis. *Intelligence*, 46, 209–215. <https://doi.org/10.1016/j.intell.2014.07.001>
- te Nijenhuis, J., & van den Hoek, M. (2016). Spearman's Hypothesis tested on black adults: a meta-analysis. *Journal of Intelligence*, 4, 6. <https://doi.org/10.3390/jintelligence4020006>
- te Nijenhuis, J., van den Hoek, M., & Armstrong, E. (2015). Spearman's Hypothesis and Amerindians: A meta-analysis. *Intelligence*, 50, 87–92. <https://doi.org/10.1016/j.intell.2015.02.006>
- te Nijenhuis, J., van den Hoek, M., & Dragt, J. (2019). A meta-analysis of Spearman's hypothesis tested on Latin-American Hispanics, including a new way to correct for imperfectly measuring the construct of g. *Psych*, 1, 101–122. <https://doi.org/10.3390/psych1010008>
- te Nijenhuis, J., van den Hoek, M., Metzen, D., & David, H. (2017). Spearman's hypothesis not supported? Three meta-analyses of Black and White prisoners, Northeast Asians, and Arabs and Jews. *Personality & Individual Differences*, 117, 52–59. <https://doi.org/10.1016/j.paid.2017.05.032>
- te Nijenhuis, J., & van der Flier, H. (2013). Is the Flynn effect on g?: A metaanalysis. *Intelligence*, 41, 802–807. <https://doi.org/10.1016/j.intell.2013.03.001>
- van den Hoek, M. (2015). *Group differences in intelligence: Spearman's hypothesis or Spearman's Law* [Masters Thesis]. University of Amsterdam.
- 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>
- Warne, R. T., & Burningham, C. (2019). Spearman's g found in 31 non-Western nations: Strong evidence that g is a universal phenomenon. *Psychological Bulletin*, 145, 237–272. <https://doi.org/10.1037/bul0000184>
- Woodley of Menie, M. A., Sarraf, M. A., Peñaherrera-Aguirre, M., Fernandes, H. B. F., & Becker, D. (2018). What caused over a century of decline in general intelligence? Testing predictions from the genetic selection and neurotoxin hypotheses. *Evolutionary Psychological Science*, 4, 272–284. <https://doi.org/10.1007/s40806-017-0131-7>

Appendix

R code

```
library(psych)
library(psychTools)
library(Hmisc)
DATA3<-data.frame(
  Data$g.loading,
  Data$Culture.loadings,
  Data$W.B.d.Kaufman..2.,
  Data$W.B.d.Kaufman..3.,
  Data$W.B.d.Kaufman..4.,
```

```
Data$e.2.Rijsdijk.et.al,
Data$e.2..Tambs.,
Data$W.B.Franklin...Altus...Clark,
Data$W.B.DeStephenes)
```

```
cortest.bartlett(DATA3)
fa.parallel(DATA3, n.iter = 500, fm="ml")
FA1=fa(DATA3,nfactors=2, residuals=T, SMC=T, fm="ml")
print(FA1,sort=TRUE,digits = 3)
diagram(FA1,cut=-0,digits = 3)
cor_2 <- rcorr(as.matrix(DATA3))
cor_2
rvalues<-cor_2$r
```