Abstract

There is no consensus within the field of psychology on whether there are sex differences in intelligence. To test this hypothesis, 2,089 effect sizes were compiled, representing 15,976,369 individuals that tested sex differences in ability. Men scored 2.58 IQ points (95% CI [1.93, 3.23], $I^2 = 99.2\%$) above women on general ability tests within adults. Whether this difference is due to general intelligence (*g)* is not clear.

Three of the four methods used to test the developmental theory of sex differences suggested that the male advantage in ability increases with age. There were substantial differences in subtest performance representing more specific abilities, with men scoring 0.71 (95% CI [0.55, 0.87], $p < .001$) standard deviations higher in mechanical reasoning and women scoring 0.29 SD (95% CI [0.43, 0.15], $p < .001$) higher in processing speed.

Keywords

Meta-analysis, sex differences, IQ, intelligence, g-factor

Educational relevance statement

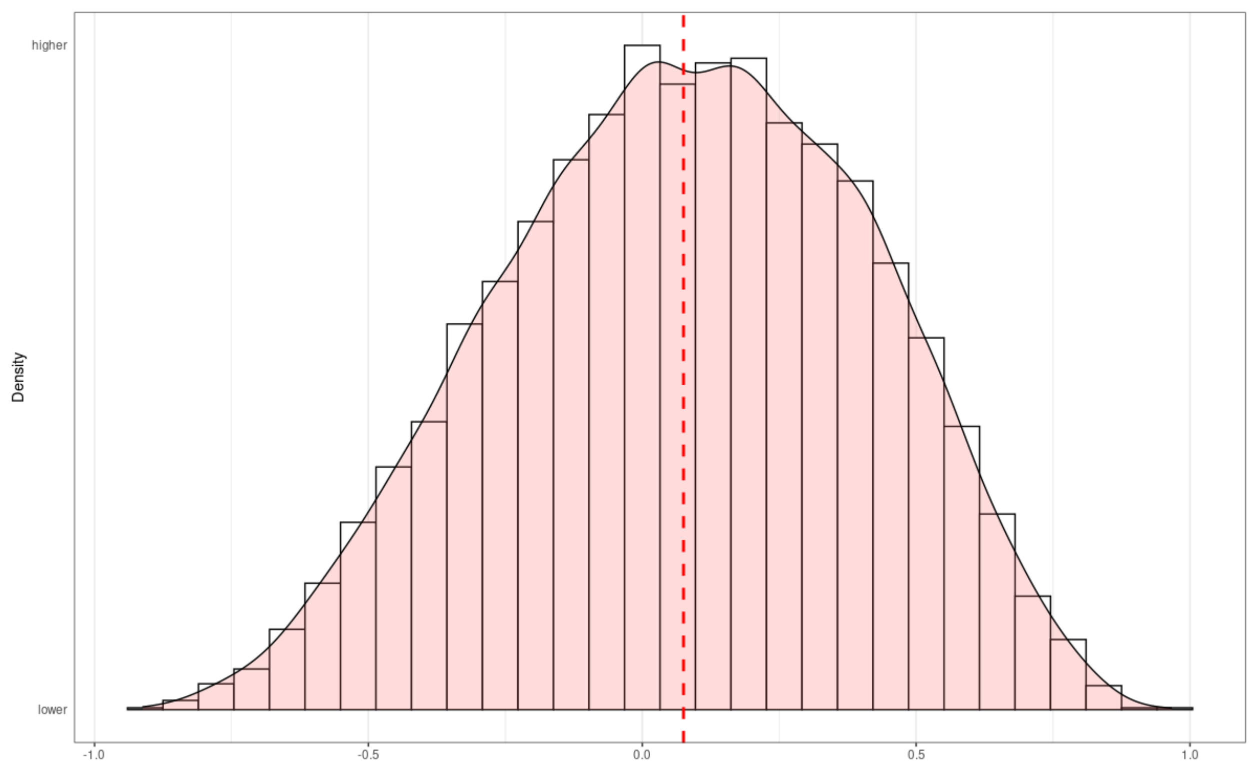
This study finds that, on average, men outperform women on spatial reasoning and mechanical reasoning tasks, while women perform better in tasks related to processing speed. Although the study finds some evidence for a small full-scale IQ difference favoring men, the evidence for it is not that strong. These findings are not intended to suggest superiority or inferiority of any group, but rather to emphasize the importance of recognizing and valuing the diverse cognitive strengths present across individuals. The results invite educators and policymakers to consider how these differences in cognitive abilities can inform educational approaches that foster fair learning environments.

1. Introduction

There is no consensus on whether there is a sex difference in intelligence, with most academics (Jensen, 1998; Nisbett et al., 2012; Cremieux, 2023; Ritchie, 2015; Murray, 2020) arguing that there is no evidence for a difference in general intelligence, and sizable minority arguing that there is a small male advantage that it emerges with age (Lynn, 2021; Nyborg, 2005; Irwing, 2011; Hanania, 2024). Although adult men outscore women on Raven's tests (Lynn & Irwing, 2004) and the full scale tests (Lynn, 2017), the crux of the disagreement is whether the difference is in general intelligence (*g*), that is, intelligence that generalizes to all cognitive tasks. Analysis that employs the method of correlated vectors suggests that the sex difference is not "on g" (Jensen, 1998). Given there are large differences in specific abilities, the method of correlated vectors will not be the best way to assess this.

To demonstrate this, a distribution of subtests was simulated that mimicked the observed distribution in sex differences in ability and g-loadings. The average test was assumed to have a mean g-loading of 0.63 (Blum & Holling, 2017) with a standard deviation of 0.13 (te Nijenhuis & van der Flier, 2013). Within the collected adult samples, the male advantage in ability was 0.13 SD with a standard deviation of 0.29, so these were used as the parameters in the simulation. With 10 subtests, the power of the method of correlated vectors to detect a male advantage in ability of .21 SD was determined to be 3.8% (10,000 iterations). The average correlation between male advantages and g-loadings in each sampling was 0.0754, as shown in Figure 1.

Fig. 1 Simulated distribution of correlations between male advantages and g-loadings in subtests within batteries with 10 subtests



Among the general public, the public consensus appears to be much more tilted towards no differences in comparison to academics, with 80% saying that there is no sex difference in intelligence, and more people saying it favors women than men (Eagly et al., 2020).

Academics have attempted to test whether there is a latent difference in cognitive ability using latent models, and some results suggest that there is an advantage in general ability in men (Arribas-Aguila et al., 2019) while others suggest there is no such advantage (Keith et al., 2008). Conceptually speaking, if the male advantage in full scale ability increases with age, and if the male advantage that does emerge with age is due to brain size, and if the association between brain size and intelligence is a generalized one, then the advantage that men have over women in intelligence should be an advantage in general ability. If latent models do not find male advantages in full scale ability, then that could be argued to be a result of a shortcoming in these methods.

2. Materials

Effect sizes were collected from a variety of sources, including search engines (google scholar, yandex, and google), prior meta-analyses (Lynn, 2017), and widely used datasets: particularly the Programme for International Student Assessment (PISA), National Longitudinal Study of Youth (NLSY), General Social Survey (GSS), National Longitudinal Study of Adolescent to Adult Health (Add health), and the Programme for the International Assessment of Adult Competencies (PIAAC). Within these datasets, differences in group factors and subtests were also calculated. The intention was to calculate the sex difference in intelligence, that is to say, the difference between men and women who were assigned their sex at birth. Intersex and transgender individuals were removed from the analyses when possible. If no sex variable was available, self-reported gender identity was used as a proxy for it.

Included in the analysis were scholastic tests (e.g. PISA), achievement tests (e.g. SweSAT, SAT), and IQ tests (e.g. WAIS). School grades were excluded as they are more reflective of personality traits like conscientiousness, while achievement tests do not exhibit this correlation (Noftle & Robins, 2007; Graetz & Karimi, 2019). This process resulted in the collection of 2,389 effect sizes which represented 51,963,553 individuals.

Studies were removed if they were not representative of the general population (in this case: employees, college graduates, LGBT people, Roma people, college students, gymnasium students, college applicants, twins, convenience samples, job applicants, employees, and high school graduates), tested ability poorly (for example, the WORDSUM), or had unbalanced sex ratios (over 60% female or male). This elimination procedure shrunk the amount of available effect sizes to 2,089, representing 15,976,369 individuals.

While most countries have very large sample sizes, high quality data on sex differences in full scale ability was concentrated in only several countries, as indicated in Figures 2 and 3.

Fig. 2 Total sample size by country, all abilities

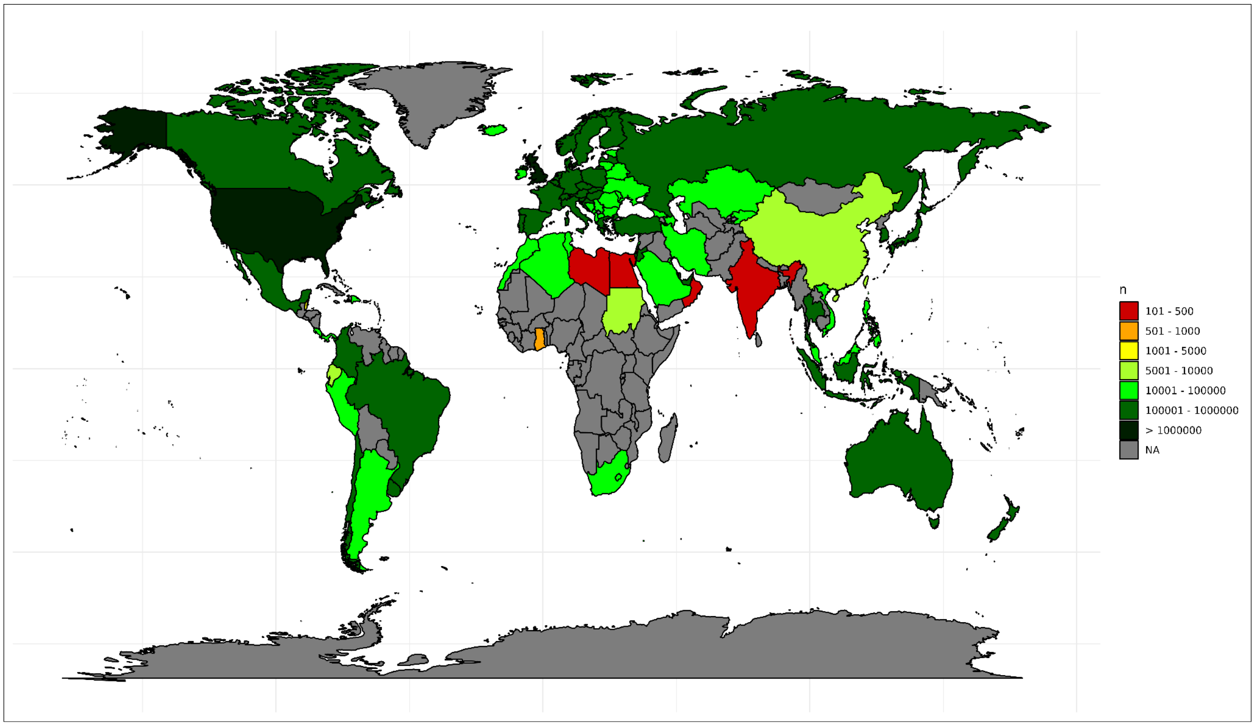
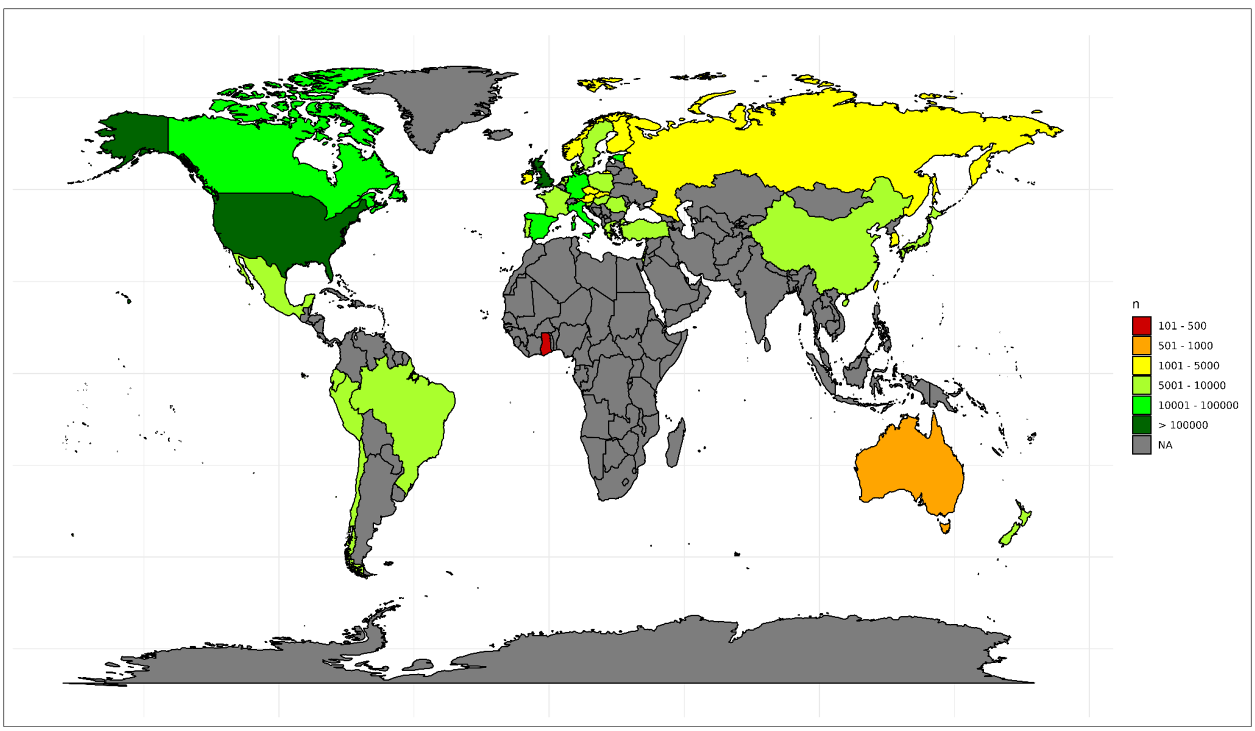


Fig. 3 Total sample size by country, only for full scale ability



These effect sizes were classified based on various moderators. The quantitative ones have been described in Table 1; the nonquantitative ones are the sample type (e.g. college students), country, test type (e.g. WAIS-IV), and ability (e.g. spatial reasoning). Test quality was was classified using seven categories: psychometric tests with over 4 subtests which tested at least 3 broad abilities were assigned a quality of 1, military tests were assigned a quality of 2, academic achievement tests were assigned a quality of 3, ravens tests were assigned a quality of 4, tests testing one subtest exhaustively or a few subtests subtests were assigned a quality of 5, tests comprised of only one subtest were assigned a quality of 6, and tests that tested one ability poorly were assigned a quality of 7. Those with a quality of 7 (e.g. WORDSUM) were excluded.

Table 1. Descriptive statistics of the moderators.

Moderator	Minimum	Maximum	Median	Standard Deviation
Sample size	44	893,000	4,719	48,128
Sample size (weighted)	18.7	893,000	4,670	46,689
Test quality	1	6	5	1.26
% Female	41%	59.90%	50.10%	2.48%
Mean age	2	81	15	11.8

3. Methodology

First, a meta-analysis of sex differences in specific cognitive abilities was made within adults and children separately. This tests both the developmental theory of sex differences in intelligence (Lynn, 2021) and whether there are sex differences in specific abilities. To avoid spurious findings, specific abilities where not enough samples (500) or effect sizes (2) within age subgroups were excluded. In addition, if effect sizes from one source were grouped by age or country, these were combined into one effect size.

A second meta-analysis was conducted to test whether there is a sex difference in full scale ability using only the highest quality samples; these exclusionary criteria are available in Table 2, which reduced the number of effect sizes to 48 and the effective sample size to 390,749. If a study reported multiple effect sizes, these effect sizes were averaged into one effect size. This is to avoid spurious findings of publication bias that could arise from studies with smaller or larger differences reporting more effect sizes than the average study. To calculate the difference between men and women, a random-effects meta-analytic model was used, and a regression test was used to assess whether there was publication bias in the meta-analysis.

Table 2. Exclusion criteria of the meta-analysis of sex differences in full scale ability. “High quality tests” in this case were either psychometric tests that tested at least 4 subtests and 3 broad abilities (e.g. Weschler), or military entrance examinations (e.g. ASVAB).

Restriction	Number of Remaining Effect Sizes
No restrictions	2,089
Only Full Scale Ability	242
Only Samples With Adults (16 or older)	162
Only High Quality Tests	119
Pooling effect sizes from same studies	48

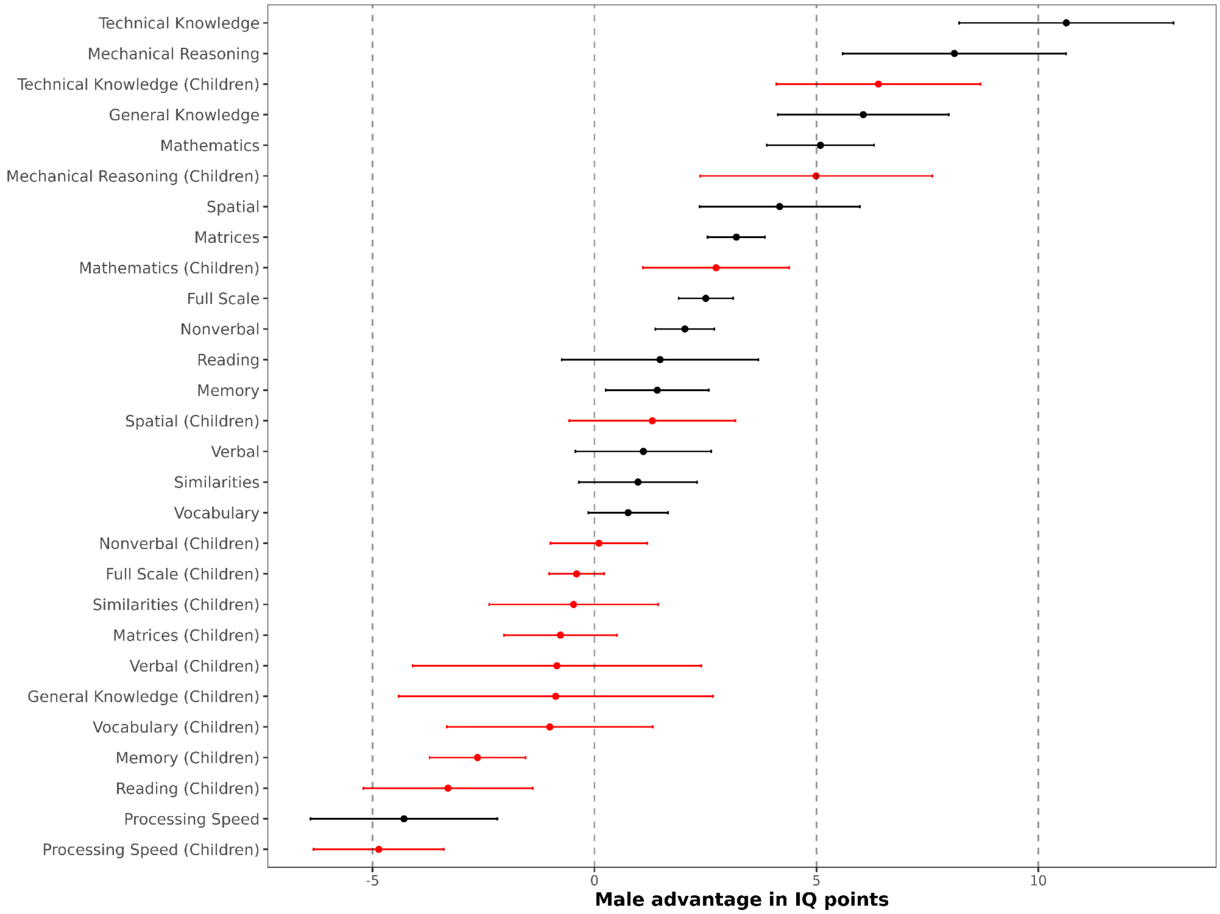
Besides this, the developmental theory of sex differences was formally tested in three ways. First, a meta-analysis was conducted within studies that tested full scale ability and the average age of the samples was examined as a moderator. Then, a second meta-analysis was conducted within all effect sizes that tested the effect of the average age of the sample on male advantages on tests, independent of the sex ratio, type of test, and the year that the study was conducted in. Last, studies that tested the developmental hypothesis were identified by restricting the sample to studies that reported effect sizes corresponding to different ages. Multiple methods of testing the theory were undertaken to observe whether the developmental hypothesis is an artefact.

Of interest was whether some nations have larger sex differences in cognitive ability. Prior research has indicated that nations differ in gender differences in scholastic ability (OECD, 2019). To test this hypothesis, sex differences found in international assessments of student learning (e.g. PISA, PIRLS) were contrasted with those found in psychometric tests.
The sex differences in reading ability (PISA and PIRLS), mathematical literacy (PISA), and scientific literacy (PISA). In the meta-analysis of international reading sex differences, the PISA test was the reference group; and in the meta-analysis of psychometric tests, full scale ability tests on adults were the reference group. Then, the correlation between these four vectors was calculated to test whether international sex differences in cognitive tests generalized to other tests as well.

4. Results

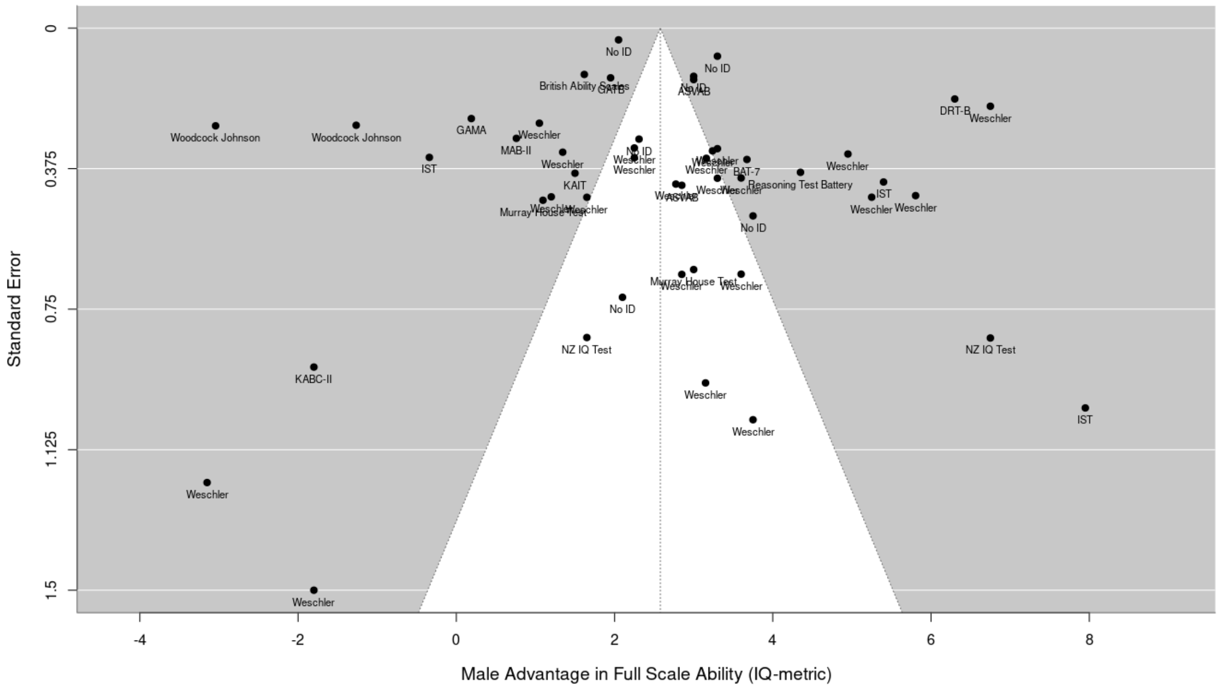
The results suggest that among adults, men score better than women on measurements of technical knowledge, general knowledge, mechanical reasoning, spatial ability, mathematical ability, memory, matrix reasoning, nonverbal tests, and full scale ability. Men and women scored about equally on measurements of reading comprehension, vocabulary and similarities (a type of vocabulary test). Women substantially outscored men on measurements of processing speed. Within children, these differences attenuated, or they flipped in direction to favor women. In all cases where adults and children were tested on the same ability, adults exhibited a larger difference in favor of men, with the only exception being verbal ability. A plot of the standardized sex difference by ability and age group is displayed in Figure 4.

Fig. 4 Sex difference in mental abilities by age group and ability type. 95% confidence intervals are displayed. Effect sizes calculated within children are displayed in red, while effect sizes displayed in black are calculated within adults. Positive effects indicate advantages for men



Adult men scored slightly higher in full scale ability (d = .17, p < .001) when all of the adult samples were pooled together, this difference remained within higher quality tests as well (d = .17, 95% CI [.13, .22], I^2 = 99.2%, p < .00001). Publication bias in favor of either sex was not statistically significant (p = .51) according to the regression test, and the funnel plot in Figure 5 shows no visual signs of publication bias.

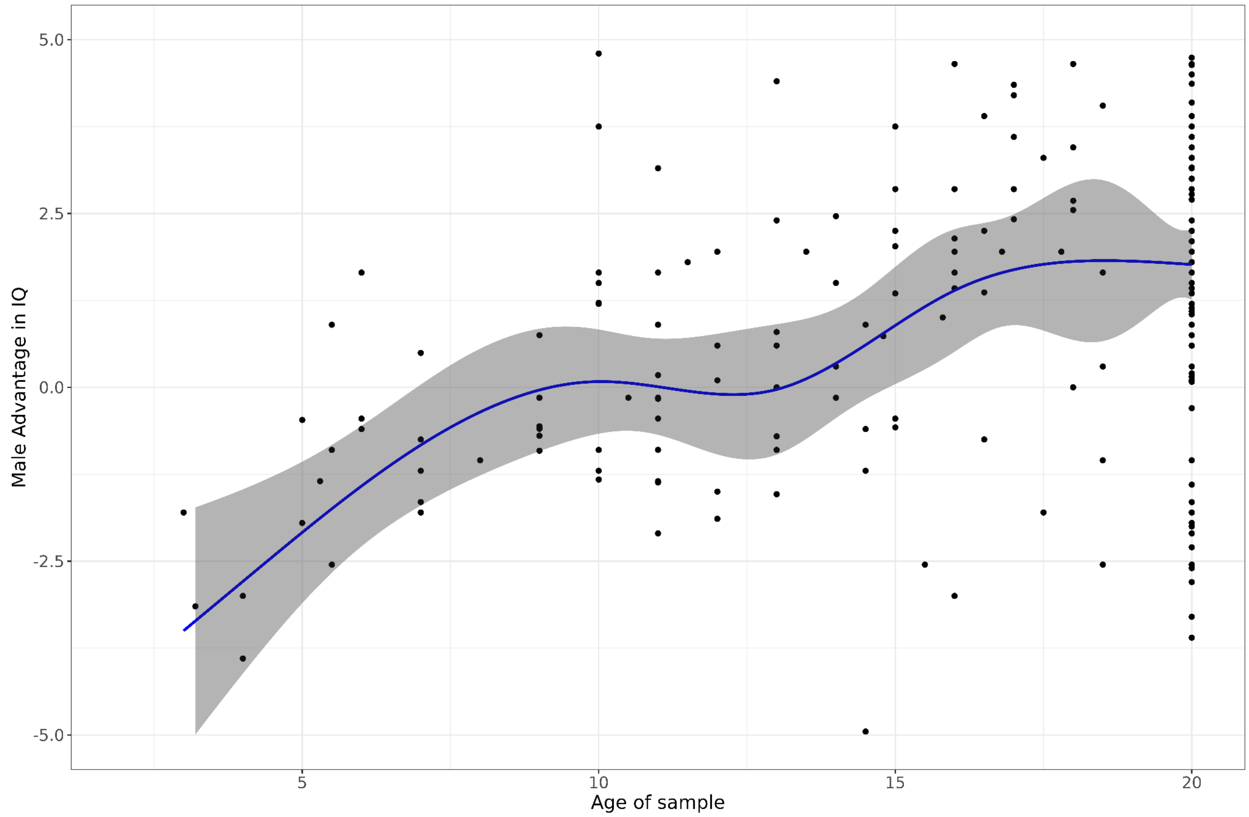
Fig. 5 Funnel plot of the difference in full scale ability between adult men and women



The developmental theory of sex differences in intelligence held in two of the three methods used to test it. An age effect was found within the samples that tested full scale ability ($b = .0023$, $p < .0001$), and when all tests were analyzed, an age effect was found even when the ability tested was controlled for ($b = .0031$, $p < .0001$). Studies that explicitly tested the developmental theory by comparing sex differences within age groups also had an age effect, though it did not pass significance testing ($b = .0010$, $p = .053$).

To visualize and quantify the difference, the dataset was restricted to representative samples with balanced sex ratios that tested matrix reasoning, full scale ability, or scholastic ability; which were chosen because they have similar differences in terms of magnitude. Then, restricted cubic splines were used to calculate the non-linear relationship between the two variables. Based on this analysis, male advantages in intelligence increased from $d = -0.15$ (95% CI [-0.20, -0.090]) at 5 years old, to $d = -0.001$ at 12 years old (95% CI [-0.049, 0.047]), and finally to $d = .13$ (95% CI [0.078, 0.18]) at 17 years old, as shown in Figure 6.

Fig. 6 Male advantage in mental ability by age group. Samples with ages of above 20 were set to 20. The 95% CI is shaded in grey



Prior literature which found variance in sex differences in international scholastic test scores by country were replicated in this study. Sex differences in cognitive ability (not assessed with these scholastic test scores) between nations correlated with the differences in international scholastic test scores, regardless of the indicator, as shown in Table 3.

Table 3. Correlation between gender differences between nations in various abilities.

Ability	Scientific Literacy	Reading	Mathematical Literacy
Scientific Literacy			
Reading	.83***		
Mathematical Literacy	.87***	.72***	
Cognitive Ability	.40**	.46**	.45**

*** →p < .001, ** →p < .01, * →p < .05.

5. Discussion

The results from the subgroup analysis suggested that adult men score higher than women on tests of technical, mathematical, spatial, general, and nonverbal ability. Men and women scored about equally on tests of vocabulary, verbal ability, and reading comprehension. Adult women only surpassed adult men on tests of processing speed. Most of these findings are uncontroversial and in line with prior literature on the topic (Hyde & Linn, 1988; Voyer et al., 1995). In a few cases, the ability differences that are found in this study are not consistent with the rest of the literature. Notably, this meta-analysis suggested that males score higher than females in mathematical ability regardless of age, while others suggest they do not differ in representative samples (Hyde et al., 1990).

The results were supportive of the existence of a male advantage in full scale ability. Male brains are about 10-12% ($d = 1.1$ to 1.6) larger than female ones (Jensen, 1998; DeCarli et al., 2023; Eliot et al., 2021; Ritchie et al., 2018), the difference survives controls for height and weight (Williams et al., 2021), and brain size and intelligence correlate at about 0.28 (Cox et al., 2019), so men and women would be expected to differ in intelligence by 4.6 to 6.7 points. While this is larger than the meta-analytic full scale difference, sex differences in intelligence could also be influenced by other variables which causes the difference to go in different directions.

Intracranial brain volume correlates with performance IQ and verbal IQ by about the same magnitude (Pietschnig et al., 2022), so the relationship between brain size and intelligence is almost certainly a generalized one. If men and women differ in brain size, then they will differ in a factor that is causal for g (Lee et al., 2019), and if they differ in overall intelligence, then it is likely that this difference is on g as well.

Concerns about sampling bias have been brought up when assessing differences in ability between men and women, particularly about whether low IQ men are poorly sampled. Men make up a larger fraction of criminals (Federal Bureau of Prisons, 2023) and homeless people (HUD Exchange, 2017), who are unlikely to be sampled accurately in scientific literature. Homeless people have IQs of about 85 (Pluck et al., 2012) and the average criminal has an IQ of about 90 (Jensen, 1998; Black & Hornblow, 1973). There are about 1.7M million prisoners (World Prison Brief, 2021) and 500k homeless people (U.S. Department of Housing and Urban Development, 2022) in the United States. If it is assumed that all homeless people and prisoners are excluded from scientific data, then the expected male advantage due to non-representative sampling error is only 0.1 IQ points. In addition, samples that were not representative of the general population (e.g. college students and gymnasium students) were labeled accordingly and not included in this study.

It is unlikely that the sex difference in intelligence is due to the fact that unrepresentative studies were flagged as representative ones in the literature search. Nationally representative datasets such as the NLSY find a difference of 3 points between men and women within adults. Most of the Weschler samples, which have an average difference of 3 points, come from norming samples. In a similar vein, it’s unlikely that the difference is due to publication bias or search bias, as no publication bias could be statistically detected, and the differences consistently showed up in the larger datasets such as the UK Biobank. The largest source of heterogeneity appears to be the test in question, which has traditionally been the strongest argument against sex differences in intelligence.

The tests where men obtain higher scores are often missing from the batteries, perhaps due to social concerns. Chiefly, this concerns 3-D mental rotation (large male advantage) and reaction time/elementary cognitive tests. This lack of representativity of the batteries would tend to slightly decrease the male advantage. Similarly, since test constructors are concerned with political opposition to testing, they may tend to systematically bias their test items so as to minimize sex differences. There is evidence that this happened historically, but it is unknown if it still occurs.

Whether the sex difference in full-scale IQ is due to a difference in generalized intelligence is unclear; based on priors it is likely, but the results from studies that use latent methods to estimate the *g* gap are contradictory (Keith et al., 2008; Arribas-Aguila et al., 2019): in the first study there is weak evidence for a female advantage in *g* within adults, and in the second study there is a difference in both abilities that favours men by about 4-5 points in older teenagers. A complicating factor here is that full-scale IQ scores are assigned based on subtest performance, and if there are sex differences in group factors of intelligence independent of *g*, then the total score will be confounded with these group factors. Whether these differences would remain after using more sophisticated methods, such as correcting for violations of measurement invariance using latent models, has yet to be seen.

One should also question whether a test can be created that does not have a sex bias due to the large group-factor differences. Even if latent methods are used, it's not clear whether it's even possible to adjust for the effect of the group factors on the estimation of the overall difference. Ideally, research in this area should try to test for whether eliminating or adding subtests from the battery affects the estimation of latent differences.

Intelligence is only moderately predictive of most social outcomes, for example, IQ and job performance as measured by work samples only correlate at about 0.38 (Strenze, 2014). Based on our meta-analysis, the difference in full scale ability should only cause a difference in 0.065 standard deviations in job performance between men and women, which is not practically significant. In comparison to other sex differences, such as sexual orientation (d = 6.5), height (d = 2), and physical aggression (d = 1) (Hines, 2019), the sex difference in intelligence would be relatively small in magnitude assuming it exists.

6. Conclusion

The available evidence is suggestive of a small male advantage in intelligence, but the quality of the evidence is too low to make a definitive judgment, as the sex differences in group factors of intelligence confound the estimation of the latent difference in general intelligence. Sex differences in specific abilities (notably mathematical ability, spatial ability, processing speed) exist and a few are large in magnitude.

7. Glossary

g: the latent trait that underlies performance on all mental tasks. See Jensen (1998) for a detailed review of the term and objections to its use.

Full-Scale IQ: the score that is generated from a composite of all the subtests of an IQ test.

g-loading: the loading of a subtest on the first general factor of mental ability tests in an IQ battery. Can also be used in isolation to denote the degree to which a test correlates with the *g* in general.

Method of correlated vectors: a method first developed by Arthur Jensen to estimate whether a correlation between IQ and one variable was due to the general factor of intelligence. Involves computing the g-loading of each subtest and the degree to which each subtest correlates with the other trait, and then computing the correlation between those two vectors.

8. Ethics

No formal ethical approval was required for this study as it did not involve the collection of human participants, personal data, or sensitive material. All procedures and analyses were conducted in accordance with the Declaration of Helsinki.

9. References

Arribas-Aguila, D., Abad, F. J., & Colom, R. (2019). Testing the developmental theory of sex differences in intelligence using latent modeling: Evidence from the TEA Ability Battery (BAT-7). *Personality and Individual Differences*, *138*, 212–218. https://doi.org/10.1016/j.paid.2018.09.043

Black, W. A. M., & Hornblow, A. R. (1973). Intelligence and criminality. *Australian & New Zealand Journal of Criminology*, *6*(2), 83–92. https://doi.org/10.1177/000486587300600203

Blum, D., & Holling, H. (2017). Spearman's law of diminishing returns. A meta-analysis. *Intelligence*, *65*, 60–66. https://doi.org/10.1016/j.intell.2017.07.004

Cox, S. R., Ritchie, S. J., Fawns-Ritchie, C., Tucker-Drob, E. M., & Deary, I. J. (2019). Structural brain imaging correlates of general intelligence in UK Biobank. *Intelligence*, *76*, 101376. https://doi.org/10.1016/j.intell.2019.101376

Cremieux, R. (2023). Male and female brains are very different sizes... *X (Formerly Twitter)*. https://twitter.com/cremieuxrecueil/status/1660737812050706433

DeCarli, C., Pase, M., Beiser, A., Kojis, D., Satizabal, C., Himali, J., Aparicio, H., Flether, E., Maillard, P., & Seshadri, S. (2023). *Secular trends in head size and cerebral volumes in the Framingham heart study for birth years 1902-1985*. Research Square Platform LLC. http://dx.doi.org/10.21203/rs.3.rs-2524684/v1

Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2020). Gender stereotypes have changed: A cross-temporal meta-analysis of U.S. public opinion polls from 1946 to 2018. *American Psychologist*, *75*(3), 301–315. https://doi.org/10.1037/amp0000494

Eliot, L., Ahmed, A., Khan, H., & Patel, J. (2021). Dump the “dimorphism”: Comprehensive synthesis of human brain studies reveals few male-female differences beyond size. *Neuroscience & Biobehavioral Reviews*, *125*, 667–697. https://doi.org/10.1016/j.neubiorev.2021.02.026

Federal Bureau of Prisons. (2023). *BOP statistics: Inmate gender*. https://www.bop.gov/about/statistics/statistics_inmate_gender.jsp

Graetz, G., & Karimi, A. (2019). *Explaining gender gap variation across assessment forms*. https://www.econstor.eu/handle/10419/201472

Haier, R. J. (2023). *The Neuroscience of Intelligence*. Cambridge University Press.

Hanania, R. (2024, January 2). Are men smarter than women? *Richard Hanania’s Newsletter*. https://www.richardhanania.com/p/are-men-smarter-than-women

Hines, M. (2019). Neuroscience and sex/gender: Looking back and forward. *The Journal of Neuroscience*, *40*(1), 37–43. https://doi.org/10.1523/jneurosci.0750-19.2019

HUD Exchange. (2017). *2017 AHAR: Part 1 - PIT estimates of homelessness in the U.S.* HUD Exchange.

https://www.hudexchange.info/resource/5639/2017-ahar-part-1-pit-estimates-of-homelessness-in-the-us/

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, *107*(2), 139–155. https://doi.org/10.1037/0033-2909.107.2.139

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, *104*(1), 53–69. https://doi.org/10.1037/0033-2909.104.1.53

Irwing, P. (2011). Sex Differences in g: An analysis of the US standardization sample of the WAIS-III. *PsycEXTRA Dataset*. https://doi.org/10.1037/e676392012-034

Jensen, A. R. (1998). *The G factor: The science of mental ability*. Praeger.

Keith, T. Z., Reynolds, M. R., Patel, P. G., & Ridley, K. P. (2008). Sex differences in latent cognitive abilities ages 6 to 59: Evidence from the Woodcock–Johnson III tests of cognitive abilities. *Intelligence*, *36*(6), 502–525. https://doi.org/10.1016/j.intell.2007.11.001

Lee, J. J., McGue, M., Iacono, W. G., Michael, A. M., & Chabris, C. F. (2019). The causal influence of brain size on human intelligence: Evidence from within-family phenotypic associations and GWAS modeling. *Intelligence*, *75*, 48–58. https://doi.org/10.1016/j.intell.2019.01.011

Lokeshwar, S. D., Patel, P., Fantus, R. J., Halpern, J., Chang, C., Kargi, A. Y., & Ramasamy, R. (2021). Decline in serum testosterone levels among adolescent and young adult men in the USA. *European Urology Focus*, *7*(4), 886–889. https://doi.org/10.1016/j.euf.2020.02.006

Lynn, R. (2017). Sex differences in intelligence: The developmental theory. *Mankind Quarterly*, *58*(1), 9–42. https://doi.org/10.46469/mq.2017.58.1.2

Lynn, R. (2021). *Sex differences in intelligence: The developmental theory*. Arktos Media Limited.

Lynn, R., & Irwing, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence*, *32*(5), 481–498. https://doi.org/10.1016/j.intell.2004.06.008

Murray, C. (2020). *Human diversity: The biology of gender, race, and class*. Twelve.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist*, *67*(2), 130–159. https://doi.org/10.1037/a0026699

Noftle, E. E., & Robins, R. W. (2007). Personality predictors of academic outcomes: Big five correlates of GPA and SAT scores. *Journal of Personality and Social Psychology*, *93*(1), 116–130. https://doi.org/10.1037/0022-3514.93.1.116

Nyborg, H. (2005). Sex-related differences in general intelligence g, brain size, and social status. *Personality and Individual Differences*, *39*(3), 497–509. https://doi.org/10.1016/j.paid.2004.12.011

OECD. (2019). *PISA 2018 results (volume II) where all students can succeed: Where all students can succeed*. OECD Publishing.

Pietschnig, J., Gerdesmann, D., Zeiler, M., & Voracek, M. (2022). Of differing methods, disputed estimates and discordant interpretations: The meta-analytical multiverse of brain volume and IQ associations. *Royal Society Open Science*, *9*(5). https://doi.org/10.1098/rsos.211621

Pluck, G., Lee, K.-H., David, R., Spence, S. A., & Parks, R. W. (2012). Neuropsychological and cognitive performance of homeless adults. *Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement*, *44*(1), 9–15. https://doi.org/10.1037/a0025382

Ritchie, S. (2015). *Intelligence: All that matters*. Hodder & Stoughton.

Ritchie, S. J., Cox, S. R., Shen, X., Lombardo, M. V., Reus, L. M., Alloza, C., Harris, M. A., Alderson, H. L., Hunter, S., Neilson, E., Liewald, D. C. M., Auyeung, B., Whalley, H. C., Lawrie, S. M., Gale, C. R., Bastin, M. E., McIntosh, A. M., & Deary, I. J. (2018). Sex differences in the adult human brain: Evidence from 5216 UK biobank participants. *Cerebral Cortex*, *28*(8), 2959–2975. https://doi.org/10.1093/cercor/bhy109

Sowers, Mf., Beebe, J. L., McConnell, D., Randolph, J., & Jannausch, M. (2001). Testosterone concentrations in women aged 25–50 years: Associations with lifestyle, body composition, and ovarian status. *American Journal of Epidemiology*, *153*(3), 256–264. https://doi.org/10.1093/aje/153.3.256

Strenze, T. (2014). Intelligence and success. In *Handbook of Intelligence* (pp. 405–413). Springer New York. http://dx.doi.org/10.1007/978-1-4939-1562-0_25

te Nijenhuis, J., & van der Flier, H. (2013). Is the Flynn effect on g?: A meta-analysis. *Intelligence*, *41*(6), 802–807. https://doi.org/10.1016/j.intell.2013.03.001

U.S. Department of Housing and Urban Development. (2022, December 19). *HUD releases 2022 annual homeless assessment report*. HUD.Gov / U.S. Department of Housing and Urban Development (HUD). https://www.hud.gov/press/press_releases_media_advisories/hud_no_22_253

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*(2), 250–270. https://doi.org/10.1037/0033-2909.117.2.250

Williams, C. M., Peyre, H., Toro, R., & Ramus, F. (2021). Sex differences in the brain are not reduced to differences in body size. *Neuroscience & Biobehavioral Reviews*, *130*, 509–511. https://doi.org/10.1016/j.neubiorev.2021.09.015

World Prison Brief. (2021). *United States of America*. World Prison Brief. https://www.prisonstudies.org/country/united-states-america

10. Appendix

Fig. A1 Relationship between sex differences in PISA tests and national IQs

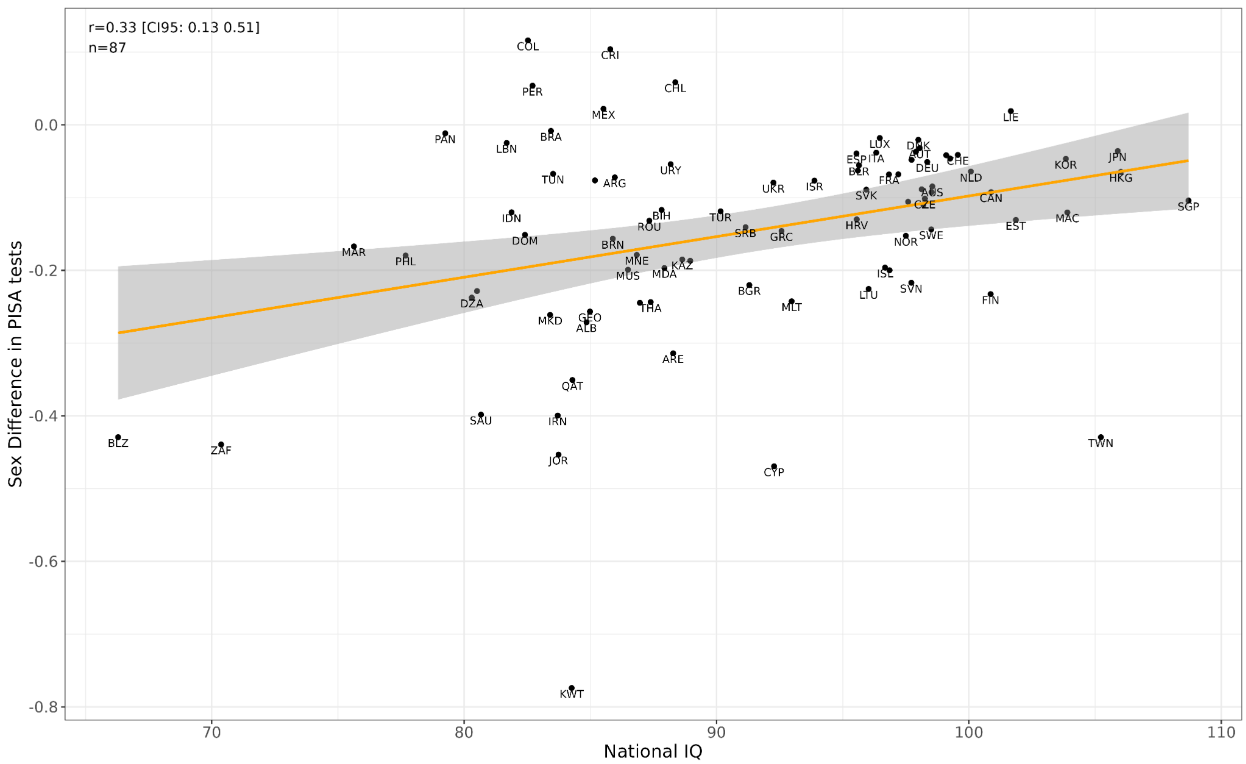


Table A1. Regression model predicting sex differences in PISA scores between countries.

Parameter	Estimate
National IQ	0.0038 (0.0017)*
Spatial Prediction	0.50 (0.11)***

*** →p < .001, ** →p < .01, * →p < .05.

Fig. A2 Intelligence distribution by sex in the NLSY97 dataset, within those tested above 15. Dotted lines denote the means for each group. Standard deviation ratio is 1.15, in favour of men

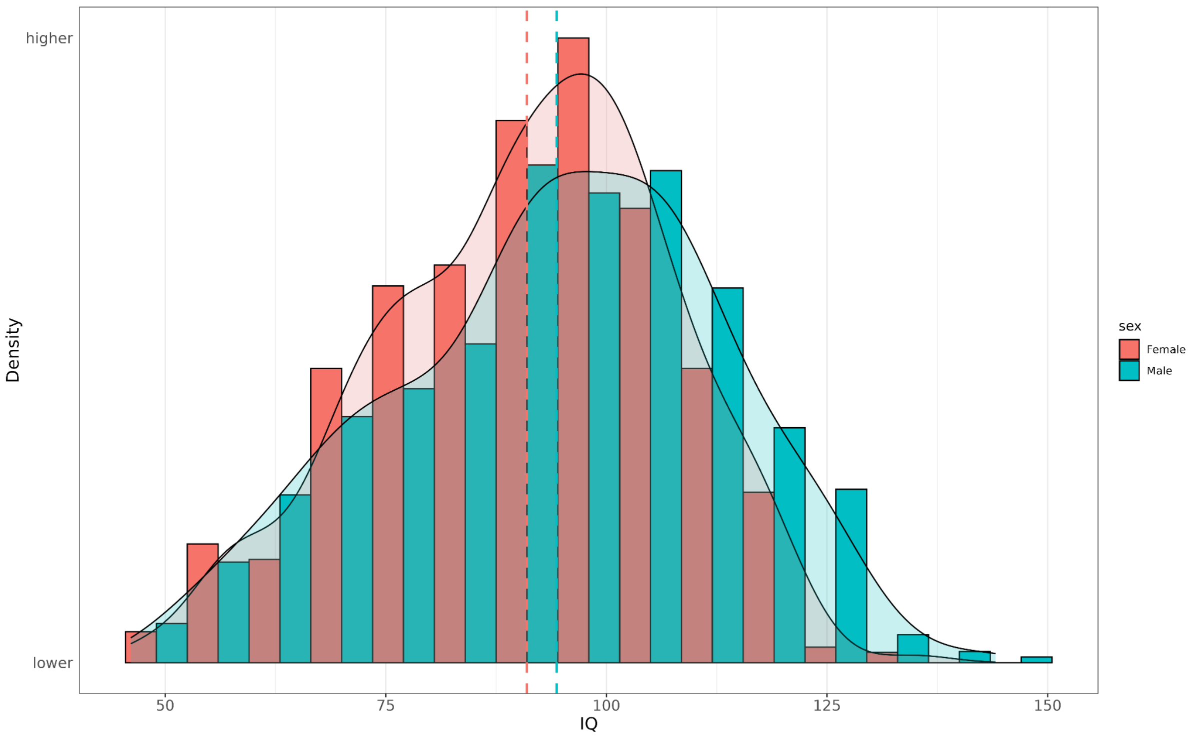


Fig. A3 Intelligence distribution by sex in the NLSY79 dataset. Dotted lines denote the means for each group. Standard deviation ratio is 1.16, in favour of men

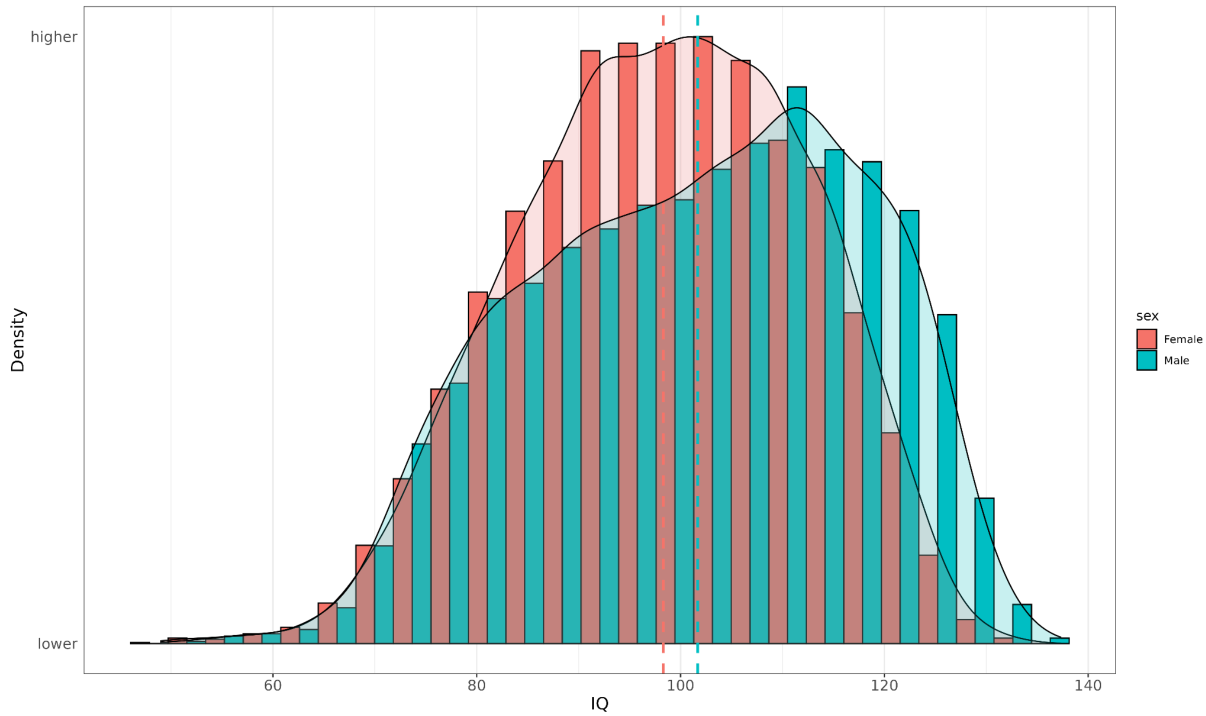


Fig. A4 Intelligence distribution by sex in the Project Talent dataset. Dotted lines denote the means for each group. Standard deviation ratio is 1.13, in favour of men

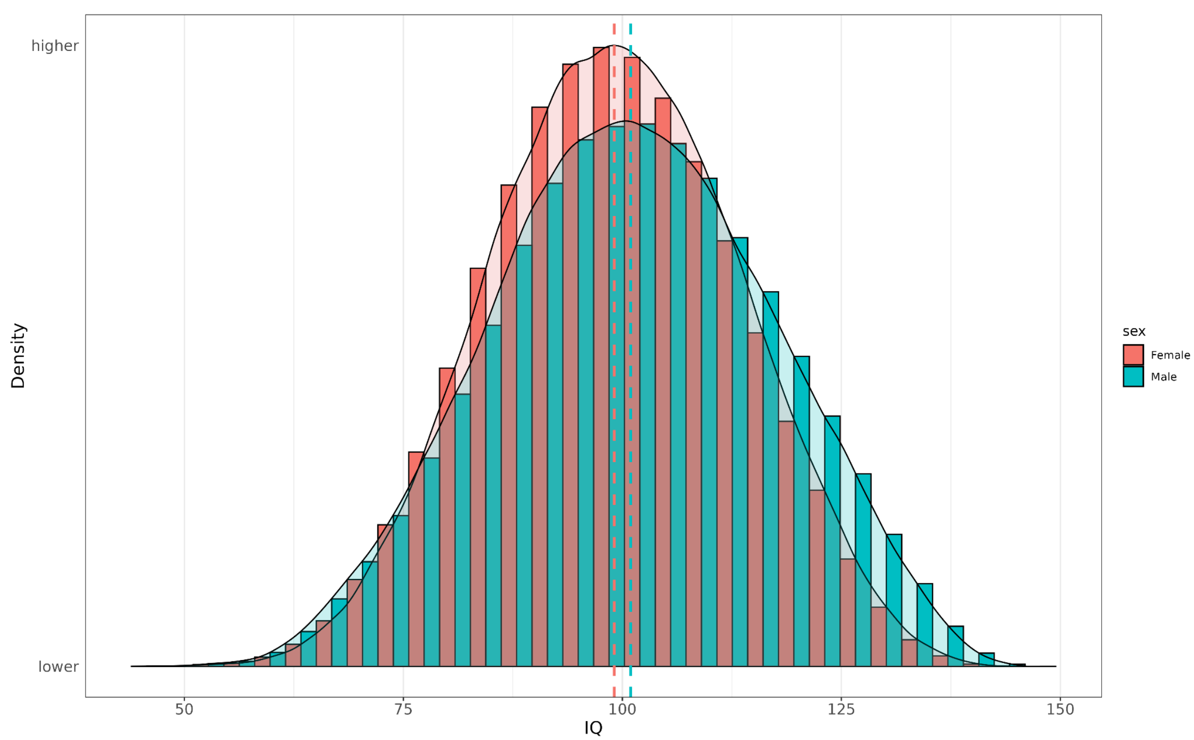


Fig. A5 Intelligence distribution by sex in the Philadelphia Neurodevelopmental Cohort dataset, within individuals tested at the age of above 15. Dotted lines denote the means for each group. Standard deviation ratio is 1.054, in favour of men

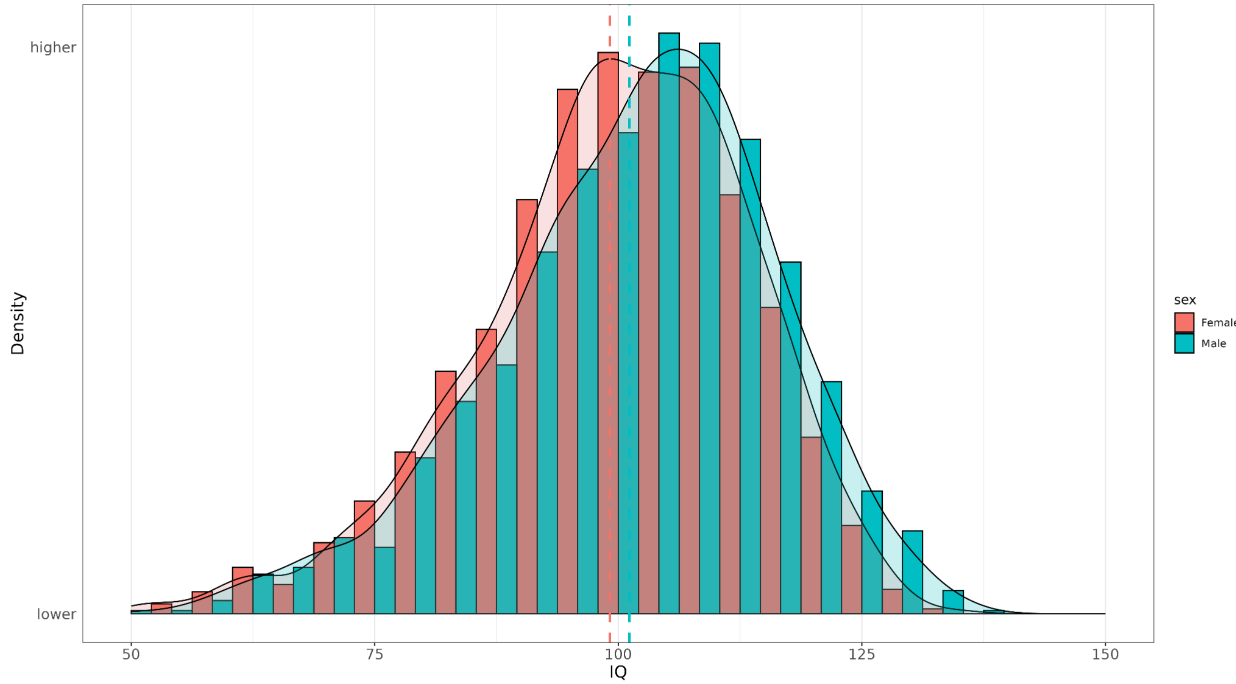


Fig. A6 Violin plot of sex difference in full scale ability within adults by test. Positive values indicate a male advantage

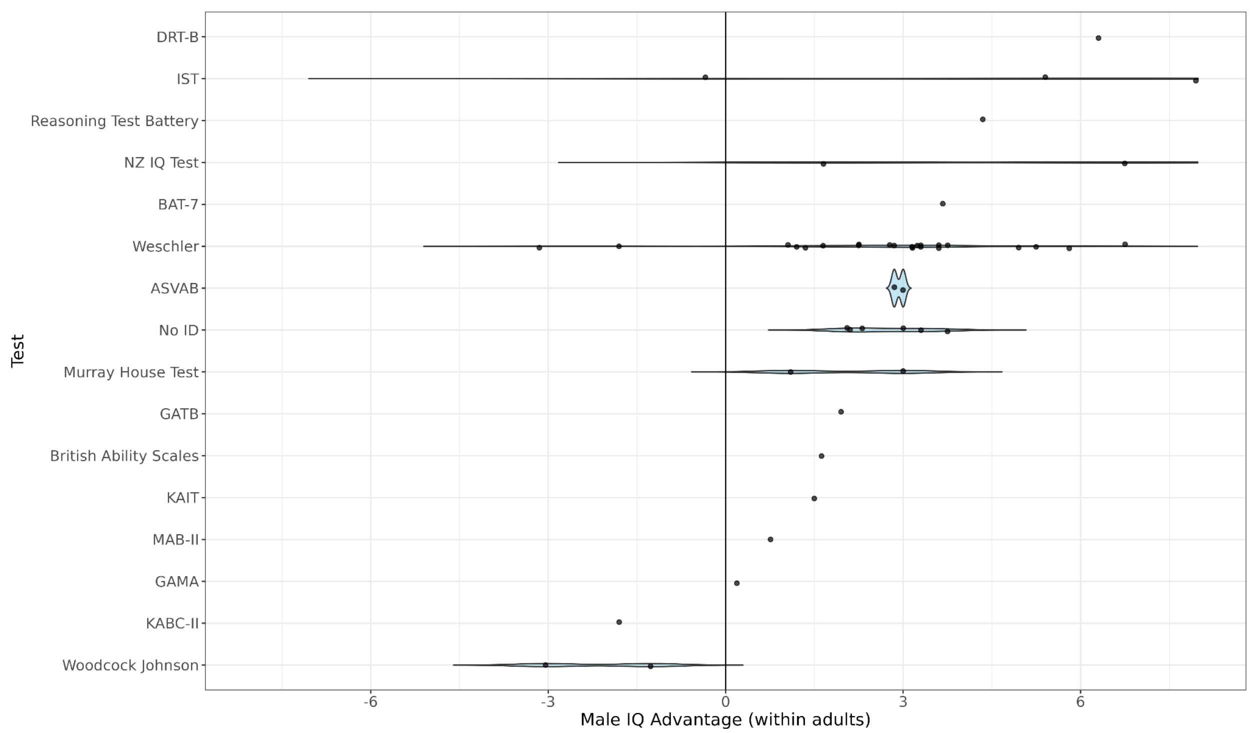


Fig. A7 Relationship between male advantages and the growth in the male advantage that occurs with age

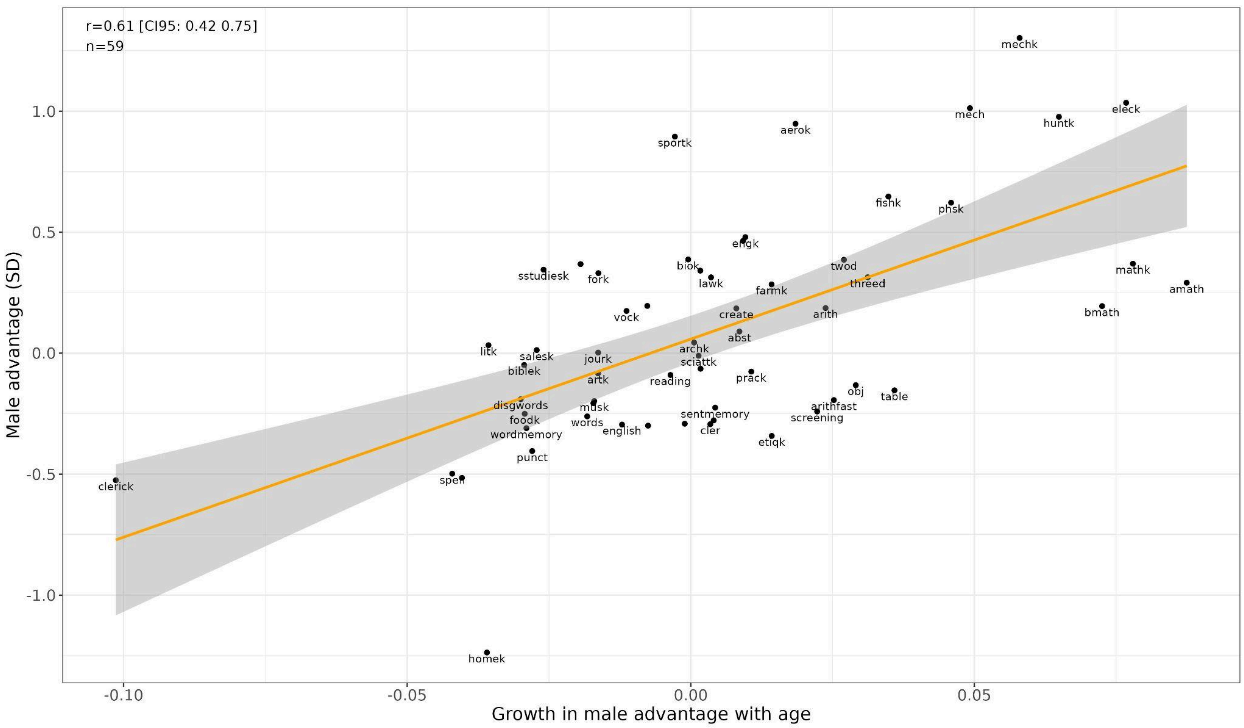


Table A2. Regression model predicting male advantages, with the independent variables being the subtest g-loadings and the rate of growth of the male advantage. Standard error in parenthesis

Parameter	Estimate
intercept	-0.095 (0.19)
growth in difference	8.29 (1.41)***
g-loading	0.27 (0.32)

Fig. A8 Male advantage by ability and age in the Project Talent. 95% CIs plotted around the dots

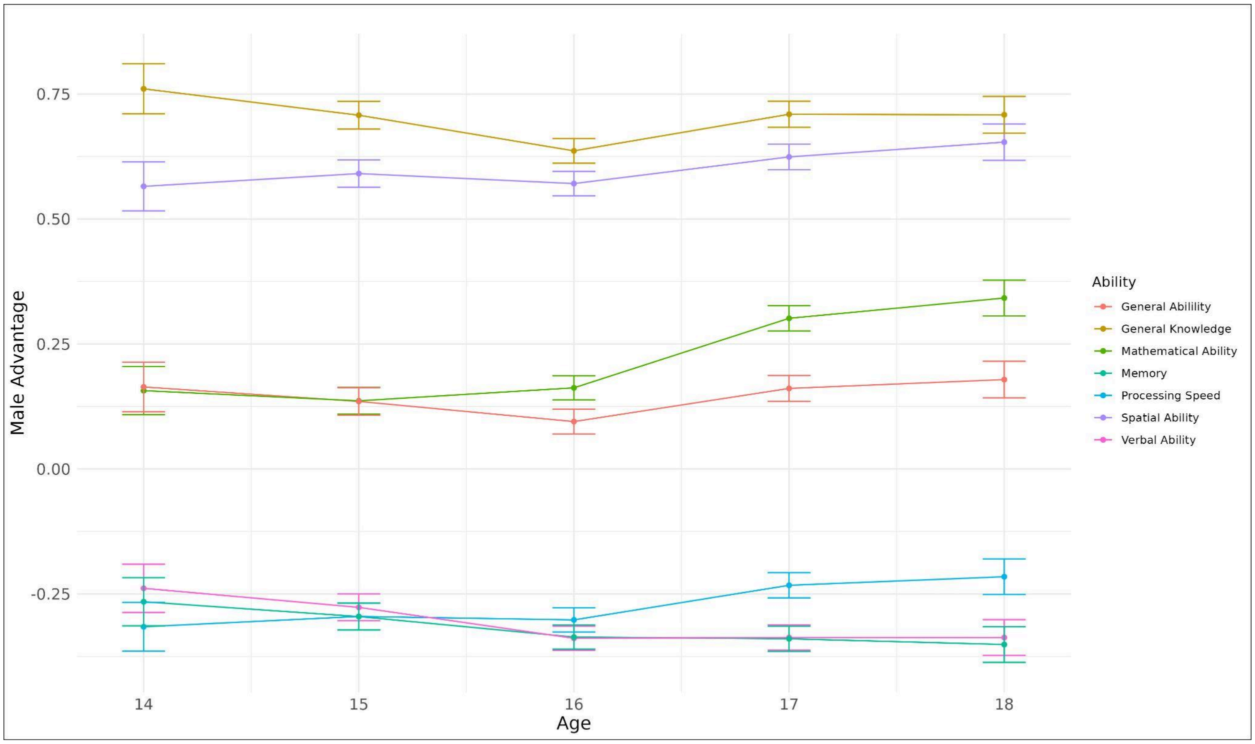


Table A3. Meta-analytic moderator analyses which regress male advantages onto mean ages in several different datasets and with different controls.

Parameter	Only FSIQ tests	All tests	All tests	Only explicit tests	Only explicit tests
Intercept	-0.046 (0.020)	-0.17 (0.11)	-0.14 (0.01)	0.00033 (0.069)	0.051 (0.020)*
Mean age	0.0023 (0.00056)***	0.031 (0.00033)***	0.0051 (0.00054)***	0.0010 (0.00054)	-0.00010 (0.00072)
Controls for ability tested	No	Yes	No	Yes	No