Abstract

There is no consensus within the field of psychology on whether there are sex differences in intelligence. To test whether there are, 2,092 effect sizes were gathered that measured differences in mental ability between men and women, representing 15,981,672 individuals. Men scored 2.58 IQ points (95% CI [1.91, 3.25], I^2 = 99.2%, k = 47) above women on general ability tests within adults. Whether this difference is due to general intelligence (*g)* is not clear, though it is likely. Two of the three methods used to test the developmental theory of sex differences suggested that the male advantage in ability increases with age.

Keywords

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

1. Introduction

Men have historically been considered more intelligent than women (Lips, 2020), which is a view that fell out of favour among academics once IQ tests became widely used to measure levels of intelligence. Although there has been no large-scale survey of the opinions of intelligence researchers on the existence of sex differences in intelligence, historically, most experts on intelligence have claimed there are negligible sex differences in intelligence (Thorndike, 1910; Burt & Moore, 1912; Cattell, 1971). This view has largely held up in the modern age (Jensen, 1998; Nisbett et al., 2012; Ritchie, 2015; Murray, 2020), though some intelligence researchers have now contested this consensus with the developmental theory of sex differences in intelligence (Lynn, 1994). This theory posits that the sex difference in intelligence is a function of age, where the difference at the age of 12 is close to null, but that the male advantage of 3-4 IQ points gradually emerges as children develop into adulthood. Most subsequent work was able to replicate the developmental effect (Nyborg, 2005; Colom & Lynn, 2004; Arribas-Aguila et al., 2019; Bakhiet et al., 2015), with a few exceptions that could not (Reynolds et al 2008; Keith et al, 2008).

Besides the disagreements over the developmental theory, it has been debated whether the sex differences in observed IQ is reflective of differences in *g*, that is, intelligence that generalizes to all cognitive tasks. An early method of testing this hypothesis was the method of correlated vectors, which tests whether the g-loadings (loading on the first general factor of mental ability) of the subtests are correlated with the associations the individual subtests have with another variable. This method has been used on various cognitive batteries, and the consistent finding is that the sex difference in subtests are not on *g* (Jensen, 1998). It's not clear if this is the most powerful method to detect a small sex difference in intelligence, as there are large sex differences in specific abilities that cannot be explained by a difference in general intelligence; women score higher in some abilities, and men do better at others.

A more popular method of testing for a sex difference in general intelligence has been using latent modeling; the results of such studies were summarized in Reynolds et al (2022), and they found that 5 out of 7 found a small female advantage in general intelligence. Of these seven, one (Härnqvist, 1997) tested young subjects (ages 11 to 16) where an overall difference is unlikely to be detected. This was also the case for Palejwala & Fine (2015) who tested children between the ages of 2 and 7, Pezzuti & Orsini (2016) who tested children between the ages of 6 and 16 and Rosen (1995) who tested 13 year olds. Keith et al (2008) did find evidence for a sex x age interaction across a wide age range (6 to 54), but it went in the opposite direction in comparison to the convention: the female advantage in *g* increased with age. Reynolds et al (2008), in contrast, found a sex difference in general intelligence at all ages that did not vary by age. Last of all, Keith et al (2011) found no latent difference in intelligence between men and women and no developmental effect either.

Given that the developmental theory of sex differences in intelligence has in some cases failed to replicate, the primary objective of this meta-analysis is to test the robustness of the effect. Although it may be that the studies that use latent methods to calculate the sex difference in intelligence favour women more often than men, there are not enough studies that use the latent methodology for the statistical comparison between the observed and latent differences to be informative; for that reason the issue is ignored.

2. Materials

Studies were gathered from search results using three different search engines: google scholar, yandex, and google. Six different search phrases were used: “sex differences in intelligence”, “sex differences in mental ability”, “sex differences in Raven’s matrices, “gender differences in Raven’s matrices”, “gender differences in intelligence”, and “gender differences in mental ability”. No restrictions were made with regard to year of publication. Although this process was not formally tracked, studies that were done on overlapping samples were excluded, one of which was Nyborg (2005) who tested the developmental theory of sex differences in the NLSY79. To gather effect sizes more efficiently, prior meta-analyses and reviews (Lynn, 2017; Reynolds et al., 2022; Voyer et al., 1995; Lynn & Irwing, 2004) were consulted as well. Lastly, data sources that include cognitive tests such as 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). At the end of this process, 2402 effect sizes had been collected, representing 51,975,891 individuals. Developed countries were expectedly overrepresented when collecting samples, as shown in Figures 1 and 2.

Fig 1 Sample size by country.

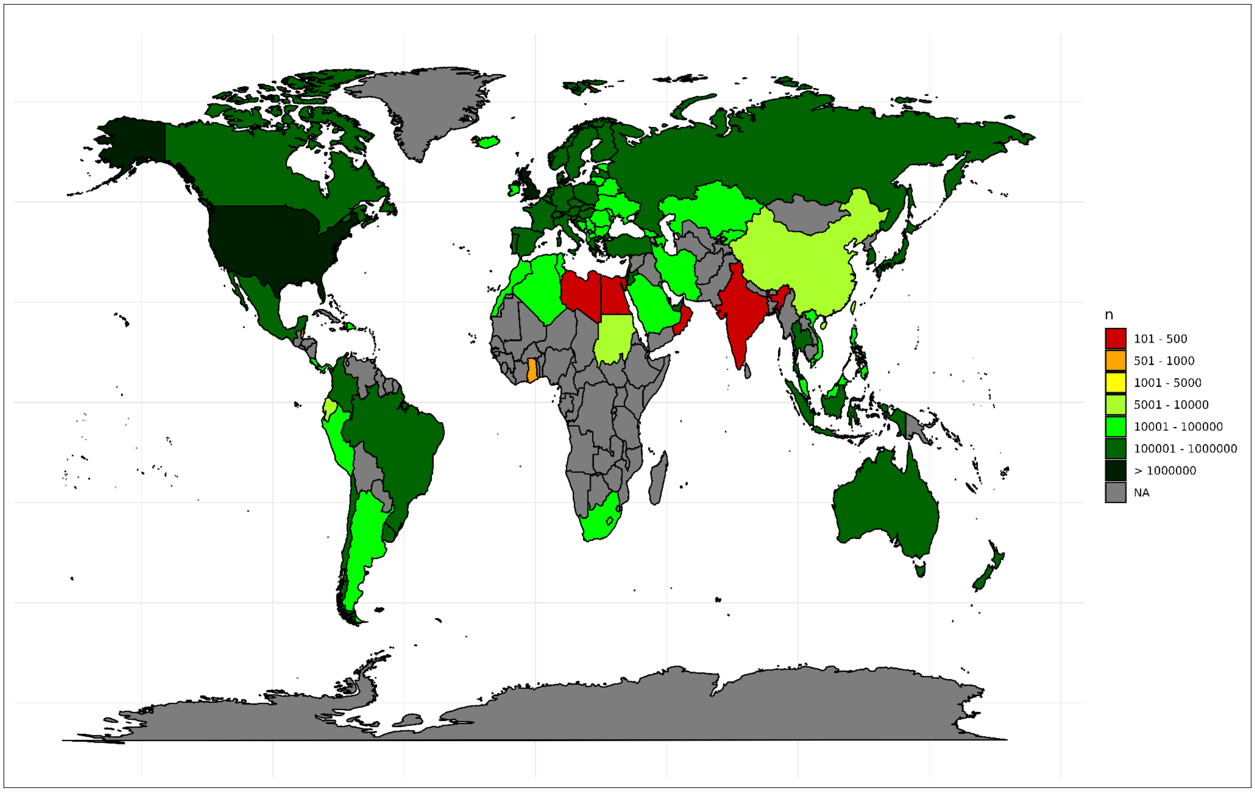
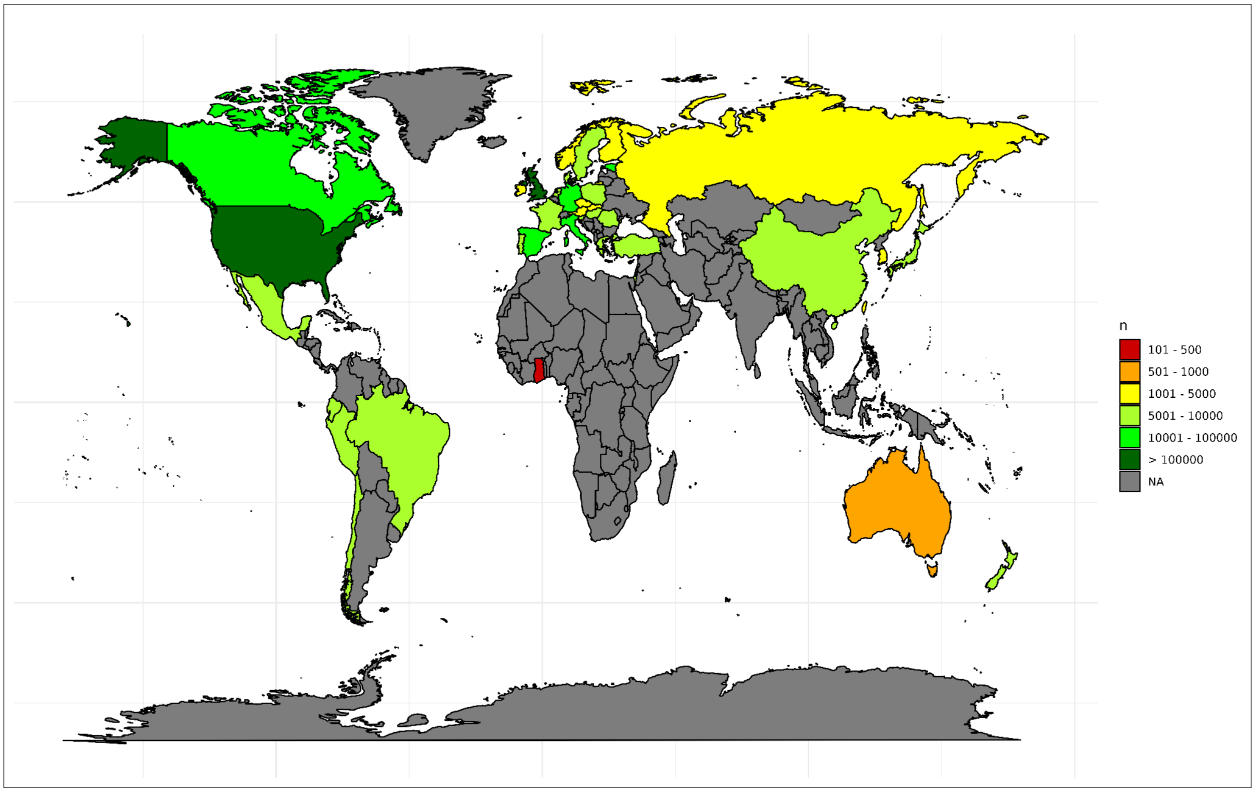


Fig 2 Sample size by country, only for tests that measure general ability.



Within the NLSY79, the ASVAB was administered to 11,914 respondents in 1981. The first unrotated general factor was extracted from the 10 subtests using factor analysis and the standardized difference in ability was calculated using the Cohen’s d statistic. Within the NLSY97, the same methodology was used, but the differences in performance within each ability and age group (12, 13, 14, 15, 16-18) were calculated.

Participants in the Add Health study were administered the Peabody Picture Vocabulary Test (PPVT) twice; once in wave 1, and again in wave 3. The scores on all times for both tests into one composite score. Within wave 1, the ages of the participants were segregated into five different bins: 12 years of age, 13 years of age, 14 years of age, 15 years of age, and 16 years and older, and the raw difference within each age group was calculated. For the 3rd wave, as all subjects were above the age of 16, the raw difference within the entire cohort was calculated.

The Programme for the International Assessment of Adult Competencies administered tests of literacy, numeracy, and problem solving ability to several countries in an effort to measure differences in skills between them. A measurement of general ability was calculated by extracting the first unrotated general factor from the three subtests and then the standardized difference was calculated. In a few countries, including Spain, France, and Italy, the problem solving test was not administered, in these cases, the standardized difference in the composite of numeracy and literacy was calculated instead. For the PISA data, the same methodology was used, though the differences in subtest performance were calculated instead of the general difference.

The WORDSUM test was administered to respondents in the General Social Survey, a 10 item multiple choice vocabulary test. The number of correct answers on the test was calculated, and the standardized difference in performance between men and women was calculated for each individual year, from 1974 to 2018. On average, men performed better than women ($d = .039$, $n = 31,950$, $p < .001$). This data was excluded from the final analysis over concerns about the reliability and validity of the WORDSUM test, due to its small number of items.

A dataset of 28,699 employees who took the GATB was privately sent to the authors; the sex differences in both specific and general ability were calculated. The difference in general ability between the sexes was -.22, favouring women, though the sample was excluded over concerns regarding how representative employees are of the general population.

In the Project Talent, the 61 subtests that were administered were grouped into six different factors: general knowledge, memory ability, processing speed, verbal ability, mathematical ability, and visual-spatial ability, and the first general unrotated factor was extracted from each of them. General ability was calculated by taking the first unrotated general factor of all 61 subtests. Then, the sex difference in each ability at the ages of 13, 13, 15, 16, 17, and 18 was calculated. Besides that, the interaction between age and sex in predicting each subtest was calculated, as well as the baseline male advantage within the whole sample. Individuals who did not have data on age were excluded from the analysis.

Cohen's d was used as the preferred measurement of effect size, though Hedge's g was used as an alternative if it was not available. Within this meta-analysis, positive values indicate male advantages, while negative values indicate female advantages. Latent differences in abilities were accepted, though the difference was only calculated using this method in 2.8% of cases. Intersex and transgender individuals were removed from the analyses when possible. If there was no sex variable, self-reported gender identity was used as a proxy for it. No corrections for test reliability were made; given that the reliability of subtests of intelligence is lower than the reliability of the full scale estimates (Weiss et al., 2010), this will cause the estimated differences in group factors of intelligence to be attenuated more than the estimated difference in general ability.

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 (German equivalent of a grammar school) students, college applicants, twins, convenience samples, job applicants, employees, and high school graduates), tested ability poorly, or had unbalanced sex ratios (over 60% female or male). This elimination procedure, summarized in Table 1, shrunk the amount of available effect sizes to 2,092, representing 15,981,672 individuals.

Table 1. Procedure used to eliminate studies of low quality.

Restriction	Number of Remaining Effect Sizes
Initial count	2,402
Only representative samples, including school students	2,145
Only samples that are over 40% or under 60% female	2,112
Exclusion of low quality tests	2,086

These effect sizes were classified based on various moderators. The quantitative ones have been statistically described in Table 2; the nonquantitative ones are the sample type (e.g. school students), country, test type (e.g. WAIS-IV), and ability (e.g. spatial reasoning). Samples were also classified according to the age of their participants: those that tested only children (under the age of 16), those that tested only adults (ages 16 and over), and those that tested both children and adults.

Tests were also evaluated in terms of their quality -- those that tested more than 3 abilities and 4 subtests were assigned the label "high quality". Two tests stood out in terms of poor quality: the WORDSUM, a 10 item multiple choice vocabulary test, and the UK Biobank's fluid intelligence test, a 13 item multiple choice test that tested people's verbal and numerical reasoning. Both tests were excluded due to their brevity and lack of items. Results that included either of these tests were excluded from the meta-analysis.

Table 2. Descriptive statistics of the moderators.

Moderator	Minimum	Maximum	Median	Standard Deviation
Sample size	44	893,000	4,715	48,090
Sample size (weighted)	18.75	893,000	4,668	46,660
% 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. Studies that tested the mental abilities of both adults and children but did not report the effect sizes separately were excluded from this analysis. The differences in specific abilities where there were not enough participants (500) or studies (1) were not reported as the magnitude of the difference would not be detected accurately. To avoid age and country segregation from biasing the results, effect sizes from individual samples that separated results by age and country were combined into one effect size, though this was done after the effect sizes were separated into the two major age groups. The composite differences were calculated with a random-effects meta-analysis using the R package *metafor* (Viechtbauer, 2009), which takes heterogeneity into account when calculating the mean differences.

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 3, which reduced the number of effect sizes to 121 and the effective sample size to 390,749. If a study reported multiple effect sizes, these effect sizes were averaged into one effect size, a process which was only done if said effect sizes were testing the same abilities. This is to avoid spurious 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 the Egger's regression test was used to assess whether there was publication bias in the meta-analysis. These meta-analytic moderator analyses were also done with the R package *metafor* and they have been posted in the Appendix.

Table 3. 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.

Restriction	Number of Remaining Effect Sizes
No restrictions	2,086
Only Full Scale Ability	239
Only Samples With Adults (16 or older)	158
Only High Quality Tests	121
Pooling effect sizes from the same studies	47

Besides this, the developmental theory of sex differences was formally tested using several different methods. First, a meta-analysis was conducted only within studies that tested full scale ability and the average age of the samples was used as a moderator. Then, a second meta-analysis was conducted within all samples that tested the effect of the average age of the sample on the sex difference in intelligence independent of the ability it was testing. Last, a meta-analysis of studies that reported effect sizes for separate age groups was conducted to test for whether the effect existed within the same sample. Sex ratio and year of publication were also considered as moderators. Various methods of testing the theory were undertaken, as each of these methods could be confounded by different biases: age effects within the same study could be confounded by attrition, while age effects between studies could be confounded by moderators such as year of publication or the selectivity of each sample.

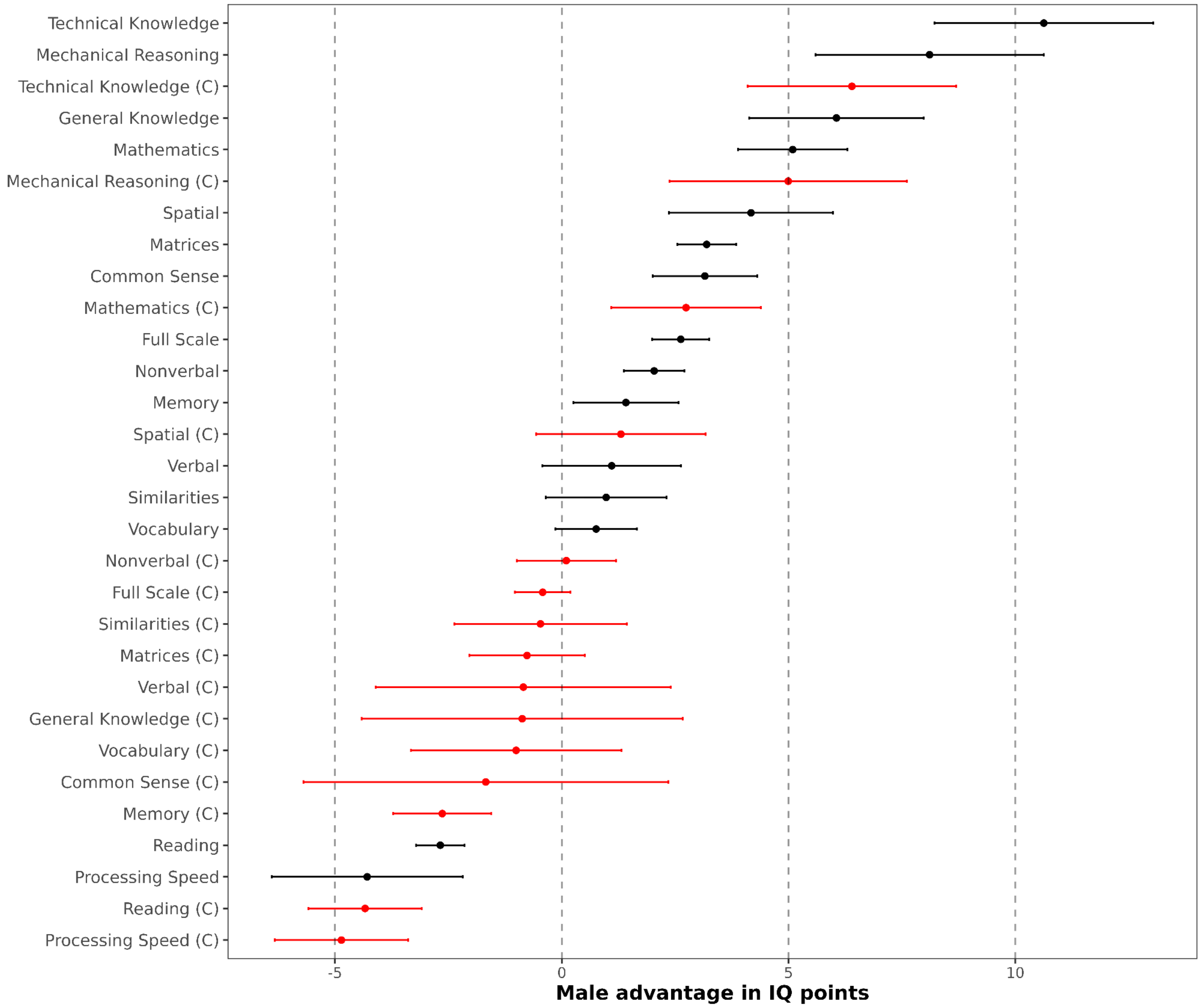
Of interest was whether some nations have larger sex differences in cognitive ability. Prior research has indicated that nations differ in gender differences in scholastic ability (OECD, 2019). To test this hypothesis, sex differences found in international assessments of student learning (e.g. PISA, 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 standardized tests as well.

Other analyses that did not fit into the main body, or were not highly relevant, were placed in the Appendix. One is an analysis of the Project Talent data which tests for whether tests with high baseline male advantages show stronger developmental effects. Another is a violin plot which plots the mean difference in general ability by the test administered (e.g. Weschler, Woodcock-Johnson), to show how spread out the effect sizes are between and within the tests. The Appendix also contains the meta-analytic moderator analyses, where the developmental theory of sex differences is tested using various different methods.

4. Results

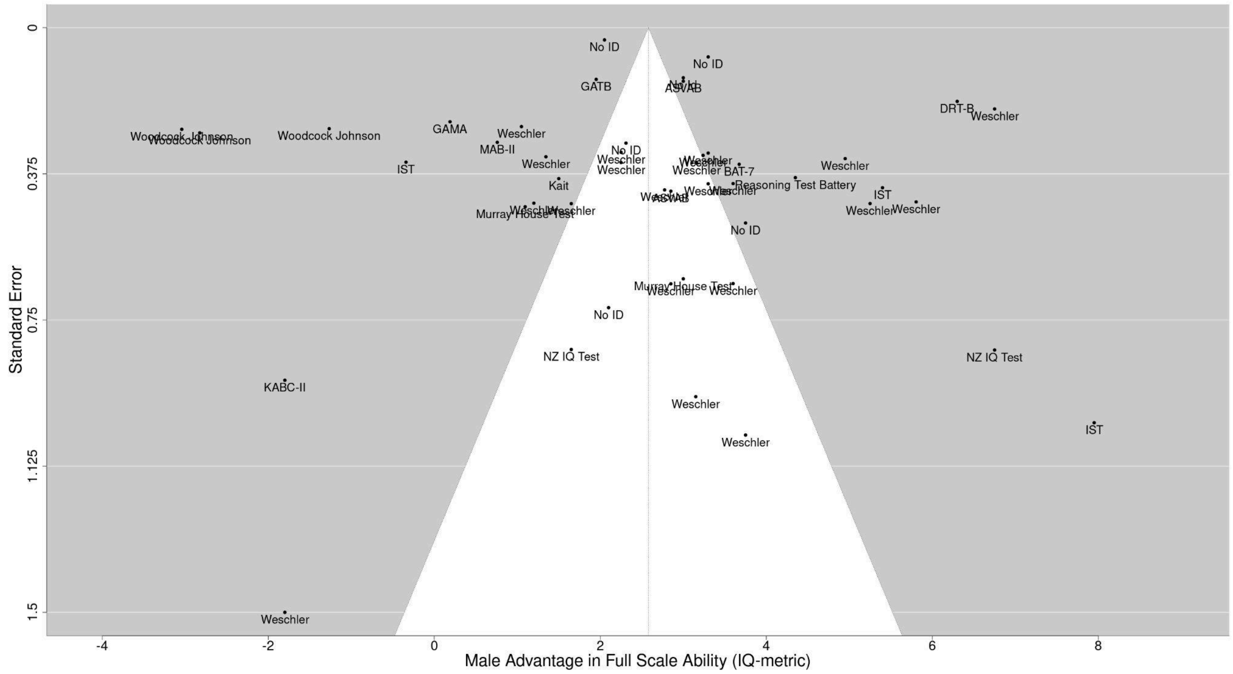
The results suggest that among adults, men score better than women on measurements of technical knowledge, general knowledge, mechanical reasoning, common sense, spatial ability, mathematical ability, memory, matrix reasoning, nonverbal tests, and full scale ability. Men and women scored about equally on measurements of vocabulary and similarities (a type of vocabulary test). Women substantially outscored men on measurements of reading comprehension and processing speed. Within children, there were no sex differences in intelligence on most tests, with the exception being that boys outscored girls on tests of technical knowledge, mechanical reasoning, and mathematics, but that girls outscored boys in tests of reading comprehension and processing speed. In every single case, the male advantage on a given sub-factor of ability was larger within adults than children. A plot of the standardized sex difference by ability and age group is displayed in Figure 3.

Fig. 3 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.



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 a sample that included only the highest quality samples (d = .17, 95% CI [.13, .22], I^2 = 99.2%, p < .00001). Publication bias in favor of either sex was not statistically significant (p = .78) according to the Egger's regression test, and the funnel plot in Figure 4 shows no visual signs of publication bias.

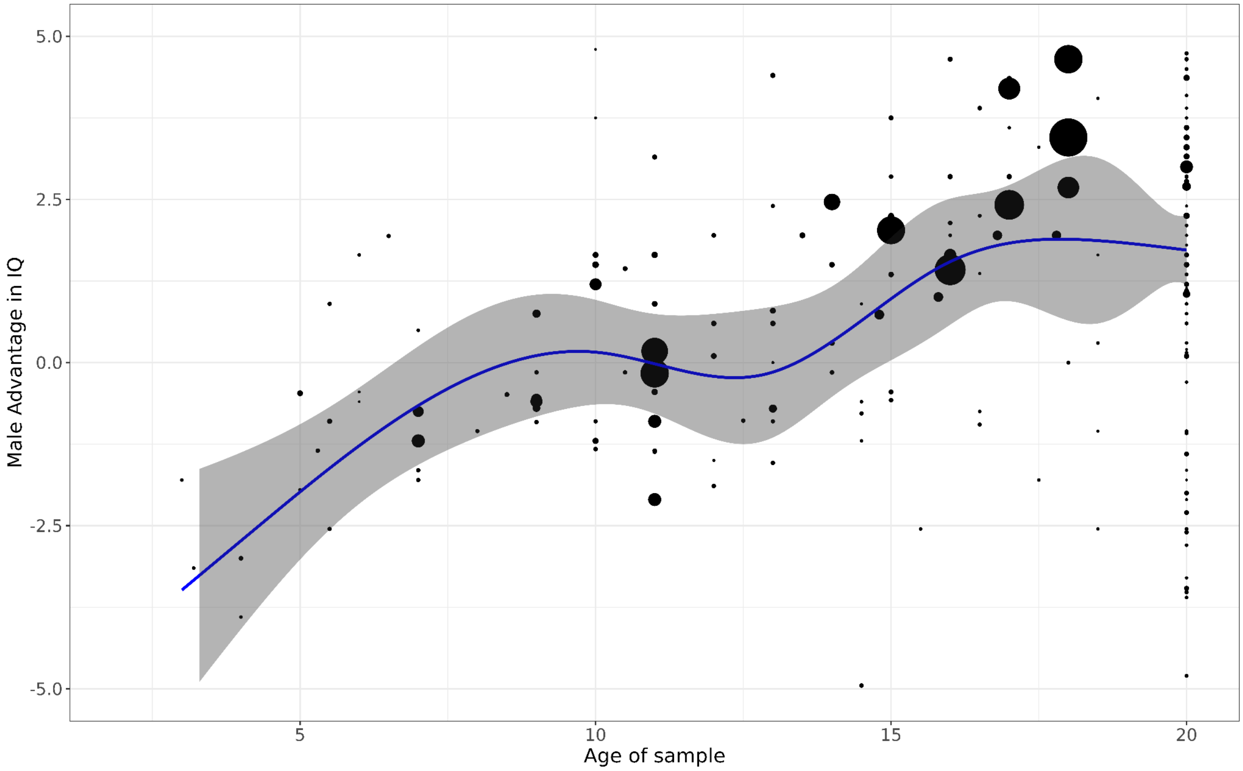
Fig. 4 Funnel plot of the difference in full scale ability between adult men and women



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 = .0021, p < .001), and when all tests were analyzed, an age effect was found even when the ability tested was controlled for (b = .0026, p < .001). 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 = .00077, p = .16).

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; these were chosen because they have similar sex differences in terms of magnitude. Then, restricted cubic splines were used to calculate the non-linear relationship between the two variables. This was judged to be the best method, as the difference in fit between it and a simple linear model passed significance testing according to an ANOVA (F = 6.7, p < .001). Based on this analysis, male advantages in intelligence increased from d = -0.13 (95% CI [-0.20, -0.064]) at 5 years old, to d = -0.014 at 12 years old (95% CI [-0.073, 0.044]), and finally to d = .12 (95% CI [0.062, 0.17]) at 17 years old, as shown in Figure 5.

Fig. 5 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 consistent sex differences in cognitive ability were replicated in this study. Sex differences in cognitive ability between nations correlated with the differences in international scholastic test scores, regardless of the indicator, as shown in Table 4.

Table 4. 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**	.44**

*** →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 and verbal ability. Adult women surpassed adult men on tests of processing speed and reading comprehension. Most of these findings are uncontroversial and in line with prior literature on the topic, specifically for spatial ability (Voyer et al., 1995), reading ability (Lietz, 2006), general knowledge (Tran et al., 2014), and matrix reasoning (Waschl & Burns, 2020).

Some of the findings here did not corroborate previous research. This paper found a nonsignificant advantage in verbal ability within men, contrary to the results of a previous one which found a small advantage in favour of women (Hyde & Linn, 1988). This is to be expected, as the label ‘verbal ability’ is a generic label that applies to tests of vocabulary, reading comprehension, spelling, tests of analogies, and sentence completion; heterogeneity in results is not unexpected as there is no reason to think that the sex difference within these subtypes of verbal ability tests should be the same. This meta-analysis found that men score higher in tests of mathematical ability by about .3 SD, which does not corroborate results from a previous meta-analysis (Hyde et al., 1990). This meta-analysis argued that the gender difference in mathematical abilities were a result of selective samples having differences that favoured males, and that they did not exist in samples of the general population. This is not supported by the statistical analyses in their own paper; when they assessed the independent effect of sample selectivity and age on the sex difference in mental ability, independent of selectivity, age still had an association with the sex difference in mathematical ability.

Most of the analyses that were conducted supported the developmental theory of sex differences, which is that intelligence tests increasingly favour boys as they mature into their adult years. Figure A2 in the appendix shows that tests with a high baseline male advantage are also the ones that come to favour them more as they mature. In a similar vein, subtests where men score higher also tend to have greater male variance in performance (Giofrè et al., 2024; Bird, 2022).

The results were supportive of the existence of a mCarliale advantage in full scale ability within adults. 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), a difference which exists after controlling for height and weight (Williams et al., 2021). 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. 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. It is not necessarily the case that an advantage in brain volume must result in an advantage in general intelligence, though it does adjust priors towards the existence of one.

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.

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 and technical ability. This lack of representation in the batteries would tend to slightly decrease the male advantage. Similarly, since test constructors are concerned with political opposition to testing, they may intentionally select items that result in smaller sex differences.

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 tend to suggest that there is no sex difference in intelligence or one that favors women (Härnqvist, 1997; Keith et al., 2008; Keith et al., 2011; Palejwala & Fine, 2015; Pezzuti & Orsini, 2016). In some of these cases, such as Pezzuti & Orsini (2016), the observed difference in intelligence is of roughly the same magnitude as the latent difference, so it would be misleading to say that the use of latent methods is responsible for the discrepancy in results. That is not to say that the apparent discrepancy should not be studied, rather it should be considered the next avenue of inquiry.

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 observed difference in general ability. Sex differences in specific abilities (notably mathematical ability, spatial ability, processing speed) exist and a few are large in magnitude.

7. 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.

8. Limitations

Studies that were excluded due to not posting sample sizes or the statistics necessary to calculate the standardized difference were not tracked. This is a rather egregious mistake, though, as noted in the discussion section, the results of this meta-analysis are very consistent with findings from previous studies.

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

Bakhiet, S. F. A., Haseeb, B.-W. M., Seddieg, I. F., Cheng, H., & Lynn, R. (2015). Sex differences on Raven’s Standard Progressive Matrices among 6 to 18year olds in Sudan. *Intelligence*, *50*, 10–13. https://doi.org/10.1016/j.intell.2015.01.013

Bird, I. (2022). Understanding greater male variability. *Patterns in Humanity*. https://inquisitivebird.xyz/p/understanding-greater-male-variability

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

Burt, C., & Moore, R. (1912). The mental differences between the sexes. *Journal of Experimental Ped*, *1*, 355–388.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*.

Colom, R., & Lynn, R. (2004). Testing the developmental theory of sex differences in intelligence on 12–18 year olds. *Personality and Individual Differences*, *36*(1), 75–82. https://doi.org/10.1016/s0191-8869(03)00053-9

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

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

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

Giofrè, D., Toffalini, E., Perugini, A., Esposito, L., Amoretti, G., & Geary, D. C. (2024). Sex differences in cognition: A meta-analysis of variance ratios in the Wechsler Intelligence Scales for Children. *Personality and Individual Differences*, *229*, 112776. https://doi.org/10.1016/j.paid.2024.112776

Härnqvist, K. (1997). Gender and grade differences in latent ability variables. *Scandinavian Journal of Psychology*, *38*(1), 55–62. https://doi.org/10.1111/1467-9450.00009

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

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

Keith, T. Z., Reynolds, M. R., Roberts, L. G., Winter, A. L., & Austin, C. A. (2011). Sex differences in latent cognitive abilities ages 5 to 17: Evidence from the Differential Ability Scales—Second Edition. *Intelligence*, *39*(5), 389–404. https://doi.org/10.1016/j.intell.2011.06.008

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

Lietz, P. (2006). A META-ANALYSIS OF GENDER DIFFERENCES IN READING ACHIEVEMENT AT THE SECONDARY SCHOOL LEVEL. *Studies in Educational Evaluation*, *32*(4), 317–344. https://doi.org/10.1016/j.stueduc.2006.10.002

Lips, H. M. (2020). *Sex and gender: An introduction, seventh edition*. Waveland Press.

Lynn, R. (1994). Sex differences in intelligence and brain size: A paradox resolved. *Personality and Individual Differences*, *17*(2), 257–271. https://doi.org/10.1016/0191-8869(94)90030-2

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., & 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

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.

Palejwala, M. H., & Fine, J. G. (2015). Gender differences in latent cognitive abilities in children aged 2 to 7. *Intelligence*, *48*, 96–108. https://doi.org/10.1016/j.intell.2014.11.004

Pezzuti, L., & Orsini, A. (2016). Are there sex differences in the Wechsler Intelligence Scale for Children — Forth Edition? *Learning and Individual Differences*, *45*, 307–312. https://doi.org/10.1016/j.lindif.2015.12.024

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

Reynolds, M. R., Hajovsky, D. B., & Caemmerer, J. M. (2022). The sexes do not differ in general intelligence, but they do in some specifics. *Intelligence*, *92*, 101651. https://doi.org/10.1016/j.intell.2022.101651

Reynolds, M. R., Keith, T. Z., Ridley, K. P., & Patel, P. G. (2008). Sex differences in latent general and broad cognitive abilities for children and youth: Evidence from higher-order MG-MACS and MIMIC models. *Intelligence*, *36*(3), 236–260. https://doi.org/10.1016/j.intell.2007.06.003

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

Rosén, M. (1995). Gender differences in structure, means and variances of hierarchically ordered ability dimensions. *Learning and Instruction*, *5*(1), 37–62. https://doi.org/10.1016/0959-4752(95)00002-k

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

Thorndike, E. (1910). *Educational Psychology*. Houghton Mifflin.

Tran, U. S., Hofer, A. A., & Voracek, M. (2014). Sex differences in general knowledge: Meta-Analysis and new data on the contribution of school-related moderators among high-school students. *PLoS ONE*, *9*(10), e110391. https://doi.org/10.1371/journal.pone.0110391

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

Viechtbauer, W. (2009). metafor: Meta-Analysis Package for R. *CRAN: Contributed Packages*. https://doi.org/10.32614/cran.package.metafor

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

Waschl, N., & Burns, N. R. (2020). Sex differences in inductive reasoning: A research synthesis using meta-analytic techniques. *Personality and Individual Differences*, *164*, 109959. https://doi.org/10.1016/j.paid.2020.109959

Weiss, L. G., Saklofske, D. H., Coalson, D., & Raiford, S. E. (2010). *WAIS-IV clinical use and interpretation*. Academic Press.

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 Violin plot of sex difference in full scale ability within adults by test. Positive values indicate a male advantage. 'No ID' implies that the construction of the test was improvised or that it was made using a combination of various professional tests.

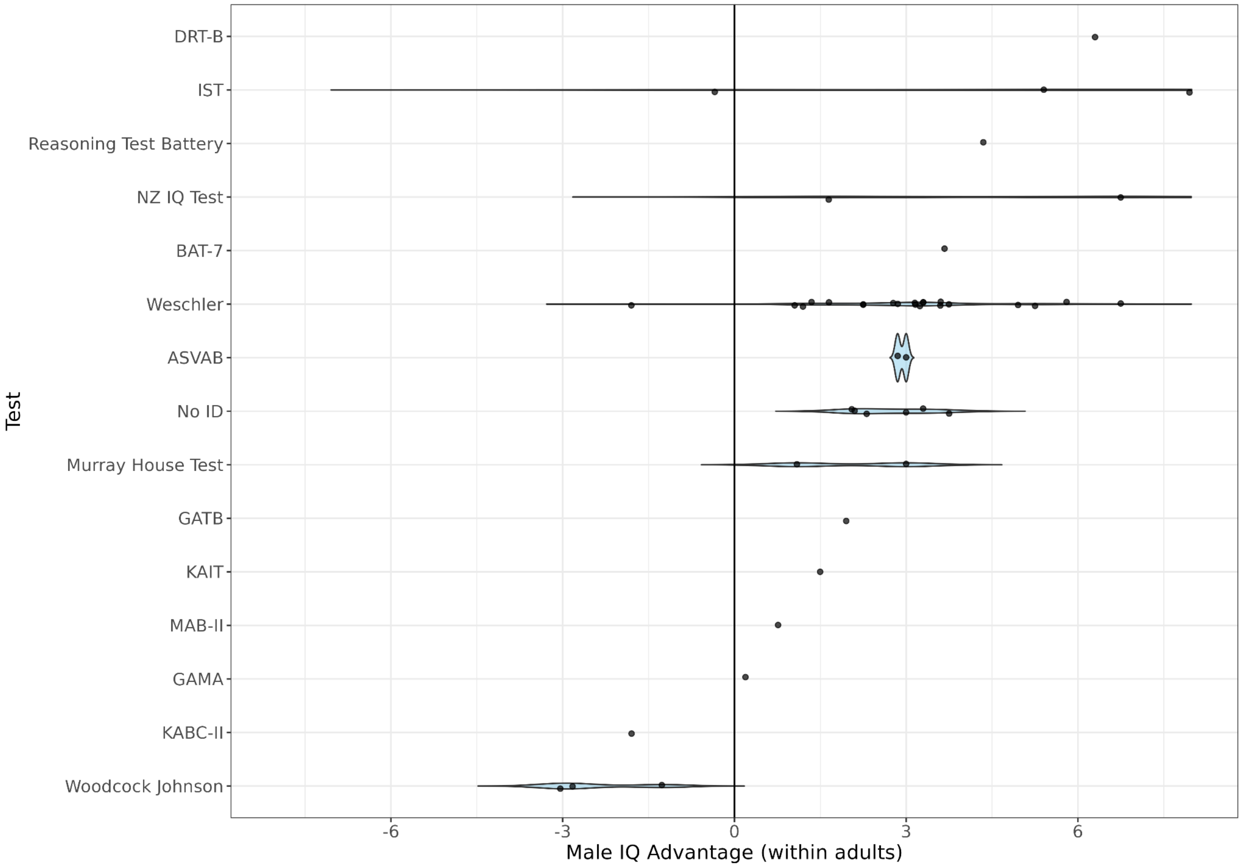


Fig. A2 Relationship between male advantages and the growth in the male advantage that occurs with age within the Project Talent. The y variable is the baseline sex difference between each subtest at all ages and the the x variable is the interaction between sex and age when predicting subtest performance, where positive values indicate increasing male advantages with age.

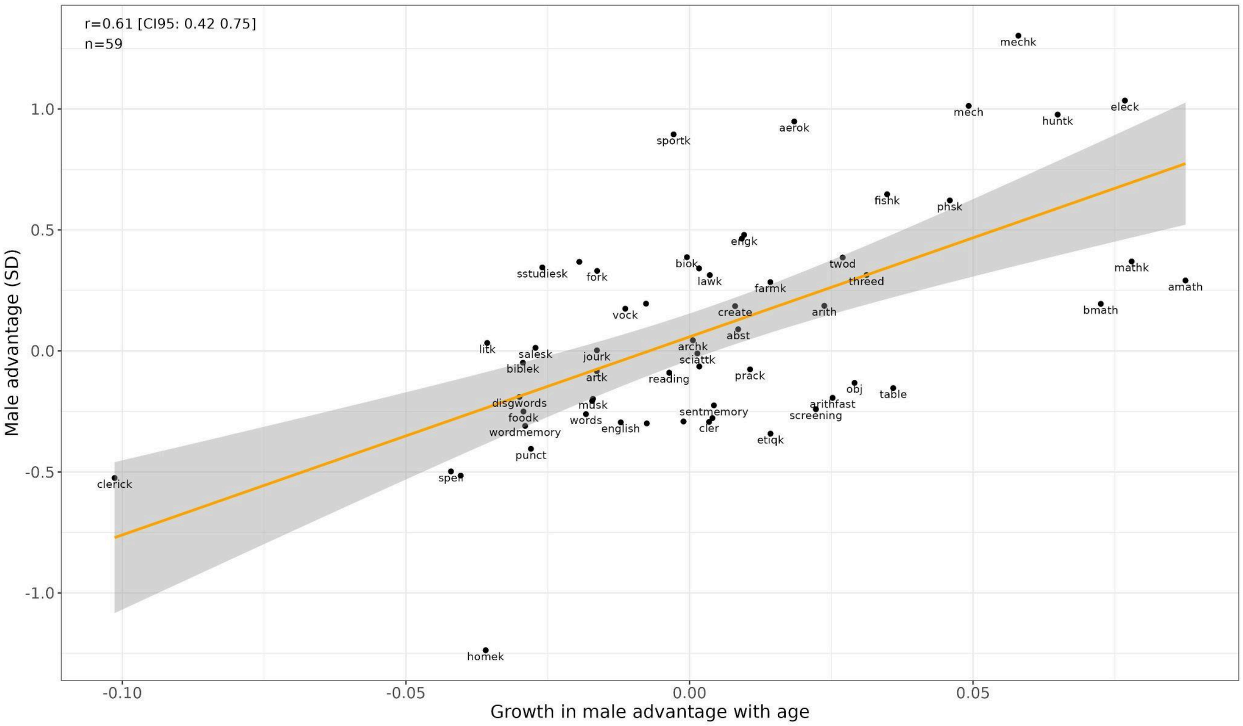


Table A1. Regression model based on the Project Talent data predicting male advantages, with the independent variables being the subtest g-loadings and the rate of growth of the male advantage. Standard error in parenthesis.

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

Table A2. Meta-analytic moderator analyses which regress male advantages onto mean ages in several different datasets and with different controls. ‘Only explicit tests’ denotes the effect sizes from studies that test the developmental hypothesis by segregating statistics by age. Regression coefficients are unstandardized, and the dependent variable is the sex difference in ability, with positive values indicating male advantages.

Parameter	Only FSIQ tests	All tests	All tests	Only explicit tests	Only explicit tests
Intercept	2.00 (1.25)	-0.17 (0.11)	5.26 (0.82)***	2.55 (0.94)**	1.95 (1.13)

Mean age	0.0021 (0.00057)***	0.031 (0.00033)***	0.0047 (0.00046)***	0.0007 (0.0005)	-0.00011 (0.00073)
Year of publication	-0.0012 (0.00062)	-0.0015 (0.0028)***	-0.0028 (0.00041)***	-0.0015 (0.0005)***	-0.0012 (0.00056)*
% Female	0.69 (0.36)	0.42 (0.14)**	0.40 (0.22)	1.12 (0.29)***	0.81 (0.41)*
Controls for ability tested	No	Yes	No	Yes	No