Meta-analysis on Gender Differences in Math Tasks

(authors)

Methods

Data analysis

Standardized Mean Differences (SMDs) between males and females in math scores were meta-analyzed using formulae and analytical strategies suggested by Borenstein et al. (2021). Random effects multilevel models were fitted with the "metafor" package (Viechtbauer, 2010) in R (R Core Team, 2023). Random intercepts were set for studies, samples (as some studies included multiple samples), and effect sizes. Heterogeneity was quantified using the following estimates: τ as the standard deviation (*SD*) of the effects across studies, ω_{sample} as the *SD* across samples, and ω as the *SD* across underlying effect sizes. Effect sizes larger than an absolute value of $d = 2.00$ were preliminarily excluded as they are considered implausible to reflect true effects and are too influential in the analysis.

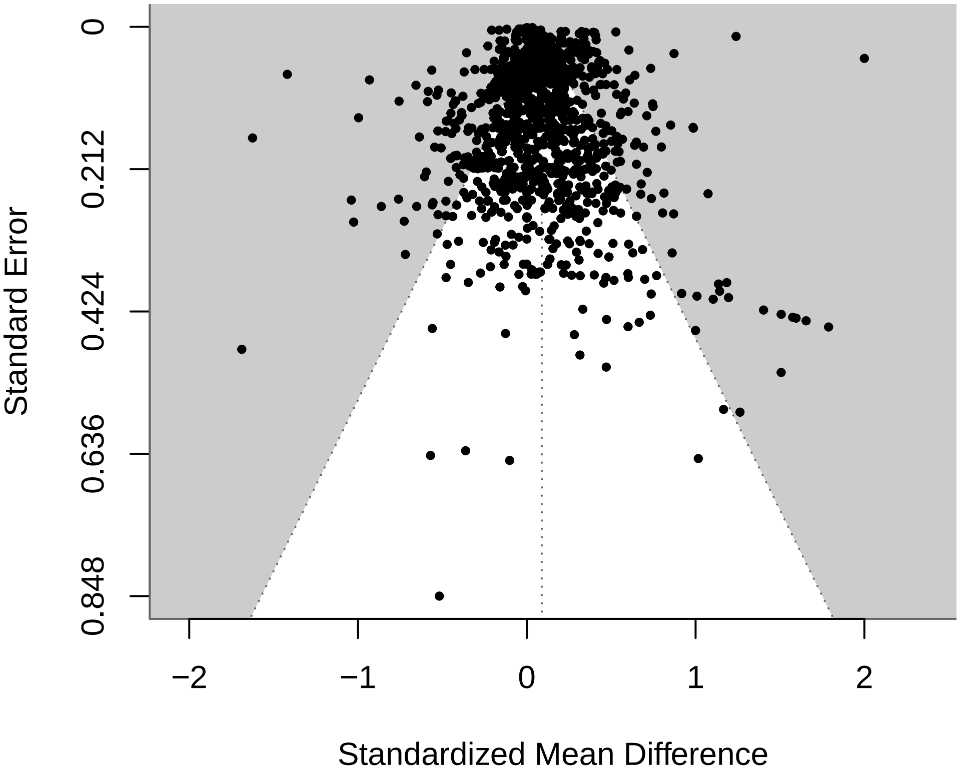
The moderators considered were the year of publication (quantitative: 2010-2022), the type of math task (categorical with 5 levels: advanced maths, basic numeracy, broad mathematics, computation, geometry), the geographical region (categorical with 7 levels: africa, central europe, east asia & oceania, middle east & russia, north america, north europe, south america), and the mean age of the sample (where available). Moderator analysis was conducted via meta-regression. Since the mean age could not be coded for all samples—either due to unavailability or because the age range was too broad—and only 25% of samples provided reliable age information, this moderator was examined in a separate analysis. For the other moderators (year of publication, type of math task, geographical region), a set of alternative models featuring all possible additive combinations and interactions were fitted, and the Akaike Information Criterion (AIC; smaller is better) (Akaike, 2011) was used to determine the best fitting model during model selection.

Results

Main effect

A total number of 440 studies, 785 samples, and 1198 effect sizes were included in the quantitative synthesis. The estimated overall number of individuals involved is 11344712. The median sample is 268. The overall meta-analytic effect estimated via multilevel random effect model (with intercept only) was $d = 0.09$ [95%CI: 0.06, 0.12], suggesting that males present an average score slightly but significantly higher than females, although the effect is virtually negligible in standardized terms. There was substantial heterogeneity across studies: $\tau = 0.22$, no estimated heterogeneity across samples, $\omega_{sample} = 0.00$, and some heterogeneity across individual effect sizes, $\omega = 0.16$. The overall heterogeneity is significant, $Q(1197) = 43437.57$, $p < 0.001$. Due to the number of studies and effects, the forest plot is impossible to represent, but the funnel plot is shown below in Figure 1.

Figure 1: Scatter plot of random data.



Moderators

Year of publication, type of math task, geographical region were simultaneously tested in meta-regression. The best fitting model (lowest AIC) featured all three moderators simultaneously in an additive model (i.e., without interaction), AIC = 130.44. Second best fitting model ($\Delta AIC = +1.55$) also featured an interaction between year of publication and type of math task. All other models presented $\Delta AICs > +4.70$ compared to the best fitting model. Also, all three effects were statistically significant when examined using likelihood ratio test: for year of publication, $Q(1) = 6.97$, $p = 0.008$; for type of math task, $Q(4) = 32.74$, $p < 0.001$; for geographical region, $Q(6) = 17.94$, $p = 0.006$. Effect of year of publication is easily interpretable due to positive meta-regression coefficient of year, $B = 0.01$, $p = 0.008$ (this suggests an increase of about $\Delta d = 0.10$ over 10 years). As other moderators are categorical with many levels, interpretation of meta-regression coefficients is difficult as it involves considering several contrasts whose p-values should be corrected for multiple testing. Therefore, since the purpose of assessing moderators was widely explorative in nature, we resorted to visual inspection of predicted effects. For simplicity, this was conducted on models fitted separately for each moderator. Confidence interval may be used for a visual inspection of reliability of effects: where an interval excludes zero, the effect may be considered non-null, while where confidence intervals overlap by less than 50% we may consider them as probably different from each other.

Figure 2 shows meta-regression effect of year of publication on male-female difference in math scores. Meta-regression curve is titled upwards, confirming that the effect is estimated as being increasing over the time span considered. LOESS smoother (dashed curve) closely follows the meta-regression curve, confirming that the increase is approximately linear.

Figure 3 shows meta-regression effect of type of math task on male-female difference in math scores. Visual inspection suggests that gender differences in both advanced maths and basic numeracy are probably null (i.e., zero on average) while in all other three areas they may be positive (i.e., presenting higher scores in males than females on average). The largest effect is estimated in geometry, with a SMD of about 0.20.

Figure 4 shows meta-regression effect of type of geographical region on male-female difference in math scores. Visual inspection suggests that gender differences are larger in Central Europe countries (estimated SMD about 0.20) than in any other geographical area.

Lastly, we examined the meta-regression effect of mean age of sample. Since mean age could be reliably coded for only 25% of samples, this analysis was conducted separately. The meta-regression effect was statistically significant, $Q(1) = 13.94$, $p < 0.001$, $B = 0.01$. The estimated increase in SMD was about 0.14 for every +10 years of age. Figure 5 shows the estimated meta-regression effect: the curve suggests that the male-female SMD is about null at the onset of formal education (6 years), and it may exceed 0.20 after 20 years of age.

Figure 2: Meta-regression effect of year of publication on male-female difference in math scores. Dots represent all (limited to the [-1, +1] range for ease of graphical representation). The solid line represents the meta-regression curve. Dashed line represents a LOESS smoother fitted for checking that the predicted effect is actually sufficiently linear. Shaded areas represent 95% confidence bands.

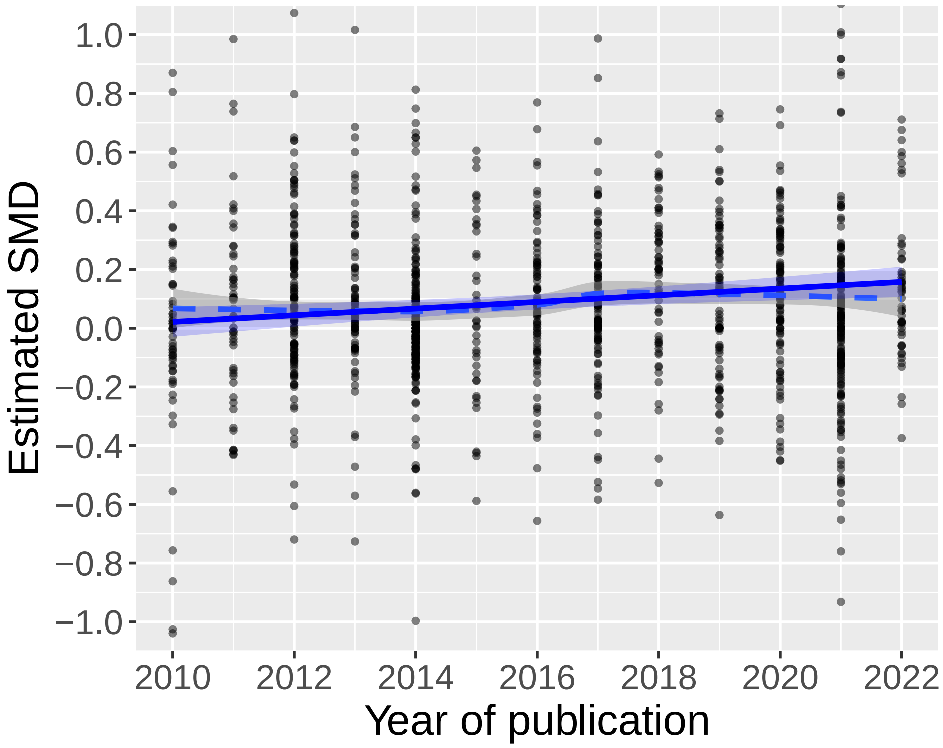


Figure 3: Meta-regression effect of math content on male-female difference in math scores. Error bars represent 95% confidence intervals.

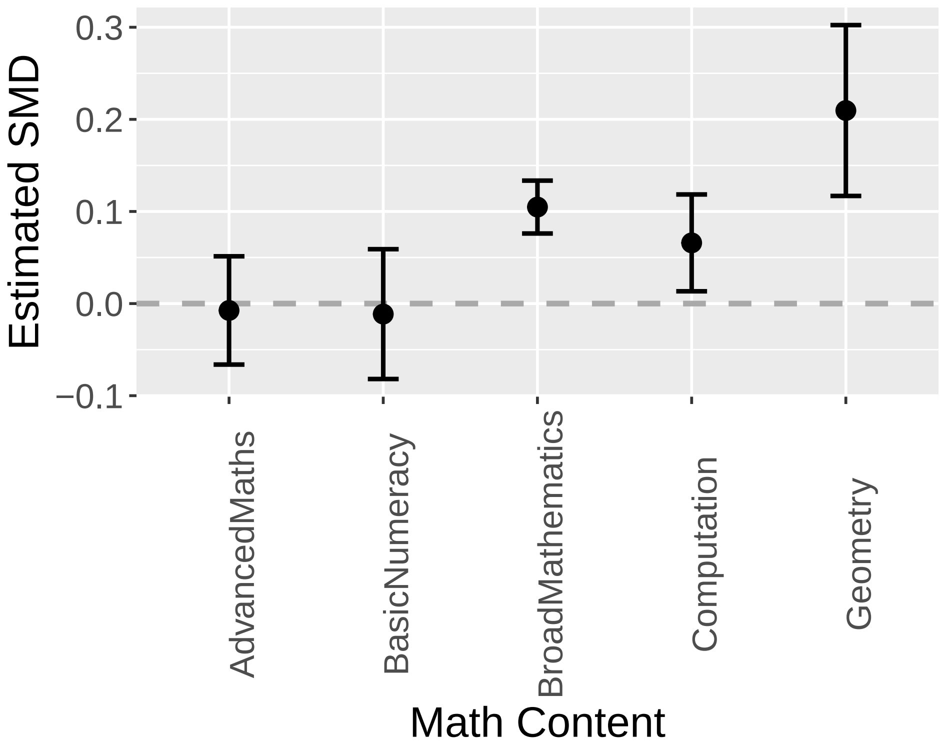


Figure 4: Meta-regression effect of geographical region on male-female difference in math scores. Error bars represent 95% confidence intervals.

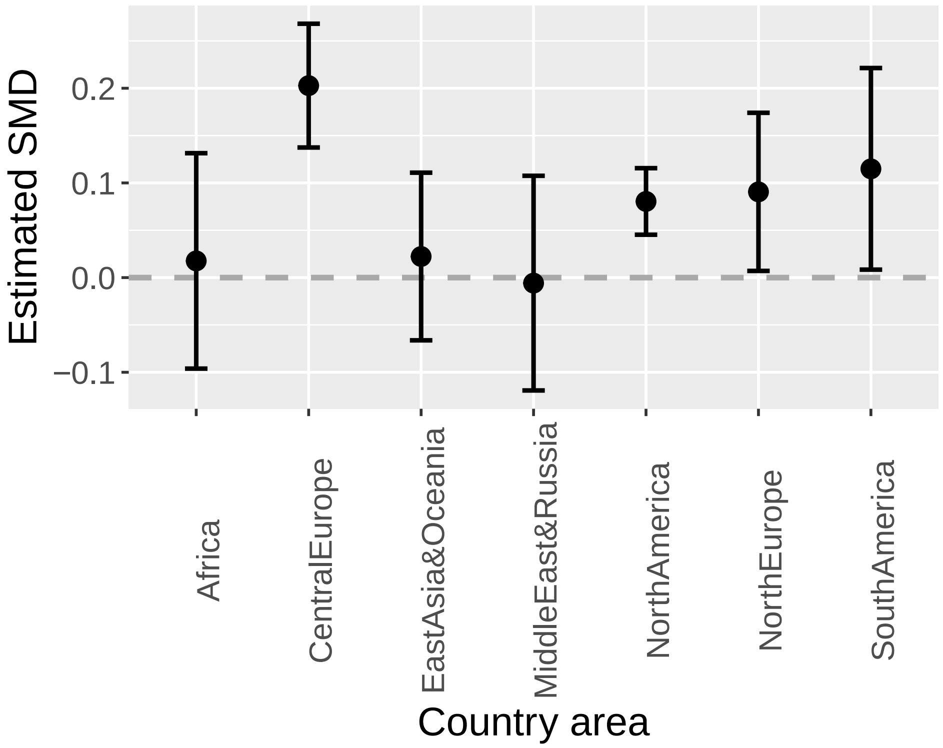
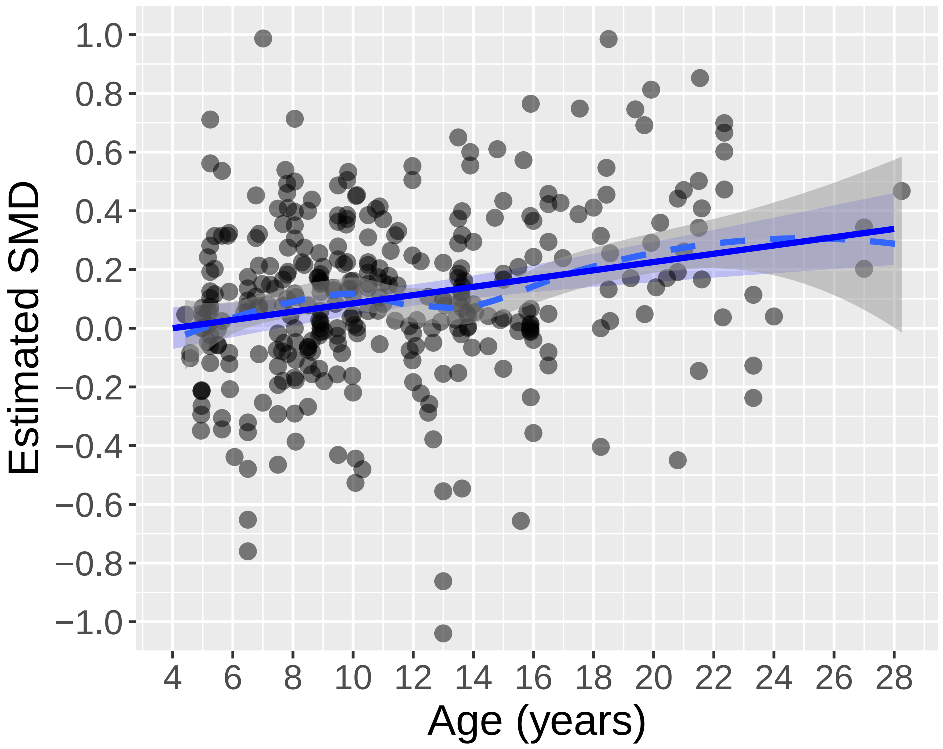


Figure 5: Meta-regression effect of years of age on male-female difference in math scores. Dots represent all (limited to the [-1, +1] range for ease of graphical representation). The solid line represents the meta-regression curve. Dashed line represents a LOESS smoother fitted for checking that the predicted effect is actually sufficiently linear. Shaded areas represent 95% confidence bands.



References

Akaike, H. (2011). Akaike's information criterion. *International Encyclopedia of Statistical Science*, 25–25.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2021). *Introduction to meta-analysis*. John Wiley & Sons.

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48. https://doi.org/10.18637/jss.v036.i03