Table of Contents

[Differences due to pre-natal androgen exposure: 2D:4D, CCFIT and study choice 1](#_Toc48411854)

[Method 3](#_Toc48411855)

[Data 3](#_Toc48411856)

[Participants 4](#_Toc48411857)

[Measures 4](#_Toc48411858)

[Analysis 4](#_Toc48411859)

[Results 5](#_Toc48411860)

[Replication of previous differences 5](#_Toc48411861)

[2D:4D and CFIT 5](#_Toc48411862)

[2D:4D and study choice 7](#_Toc48411863)

[Discussion 9](#_Toc48411864)

[Limitations 10](#_Toc48411865)

[Future research 11](#_Toc48411866)

[Conclusion 11](#_Toc48411867)

[References 12](#_Toc48411868)

[Appendix A: Descriptive statistics. 18](#_Toc48411869)

[Appendix B: Figures representing distributions. 19](#_Toc48411870)

[Appendix C: Pre-registration 23](#_Toc48411871)

Differences due to pre-natal androgen exposure: 2D:4D, CCFIT and study choice

It is commonly reported that adult males on average perform better in tasks assessing visuo-spatial and quantitative ability and are oriented more toward ‘things’ and ‘objects’ than towards people, while adult women on average perform better in tasks assessing verbal fluency, perceptual speed and are oriented more towards people than ‘things’ and ‘objects’ (Neisser et al., 1996; Halpern, 1997; Diekman et al., 2010). Women are also less likely to seek to study or work in STEM fields (Beede et al., 2011). Various measures and markers have been employed to identify sex differences and relationships to developmental features or other traits and behaviour. This paper discusses: fluid intelligence as measured by the Cattell Culture Fair Intelligence Test (CCFIT); the choice to study science or arts; and ‘2D:4D’, the ratio of second to fourth digit length.

Pre-natal androgen exposure appears related to sex differences, as seen in individuals with congenital adrenal hyperplasia (CAH) and androgen insensitivity syndrome (AIS), both of which reliably modify the processing of relevant hormones (Berenbaum, 1999; Cohen-Bendahan, van de Beek & Berenbaum, 2005). 2D:4D, as a proxy for pre-natal androgen exposure (Manning, 1998; Galis et al., 2010), is used to investigate differences in non-clinical populations.

The male-typical low 2D:4D (indicating higher androgen exposure) correlates with ‘masculine’ traits: higher numerical intelligence, lower verbal intelligence and lower agreeableness (Luxen & Buunk, 2005); interest in male-typical occupations (Hell & Päßler, 2010; Nye & Orel, 2015) and self-reported masculinity (Manning, Trivers & Fink, 2017). However, results are mixed. As summarised by Wong and Hines, correlations have been found, not found, found but non-linear (Sanchez et al., 2014), found for left hand ratio but not for right (Nye & Orel, 2015) and found contrary to prediction and reliant on weak evidence (2016; Lippa, 2006; Putz et al., 2004). Methodological differences contribute to the inconsistency (Manning et al., 2005; Ribeiro et al., 2016) but 2D:4D remains common enough as a biomarker for pre-natal androgen exposure to generate meta-analyses of investigations of differences in cognition and behaviour (Hönekopp, & Watson, 2010; Hampson & Sankar, 2012). Some analyses focus particularly on within-sex variation and outcome variables (eg. women with lower 2D:4D *than other women*) (Bull et al, 2010; Nye & Orel, 2015).

Support for sex differences varies between measures of attainment or ‘crystallised intelligence’ and of ‘fluid intelligence’ (Kaufman et al., 2009). Crystallised intelligence is seen as the result of learning, while fluid intelligence is a more heritable endowment of problem-solving ability and, theoretically, unrelated to sex (Kent, 2017). Outcomes in crystallised intelligences, occupational interests and orientation to things vs people are thought to be mediated by non-cognitive interests and preferences (Ackerman & Heggestad, 1997; von Stumm & Ackerman, 2013), which may have a more direct relationship with androgen exposure. It will be of interest, therefore, to consider differences in fluid intelligence inasmuch as it represents a heritable starting point antecedent to the effect of environment or culture.

Lynn and Irwing’s meta-analysis challenged the consensus that fluid intelligence showed no sex difference, finding an advantage for men over women in Raven’s matrices (Lynn & Irwing, 2004). And, while Kaufman et al. found no effect of gender on fluid intelligence using KBIT-2 matrices (2009), a large effect size (*d* = 0.869) was found in Bosnia & Herzegovina among seventeen year olds using Raven’s advanced progressive matrices (Dapo & Kolenovic-Dapo, 2012).

We predict no relationship between 2D:4D and scores on the CCFIT but a significant relationship between 2D:4D and study choice (arts or science). This would be consistent with a world in which individuals are endowed, independently of sex, with varying degrees of broad-ranging, non-specific problem solving ability but their differentiated disposition towards the domains in which they apply that ability has a relationship with pre-natal androgen exposure. Our hypotheses are:

H1: *that there will be no significant correlation in our sample between 2D:4D and scores on the CCFIT.*

H2: *that individuals in our sample studying science will on average have lower 2D:4D than individuals studying arts.*

# Method

## Data

The University of Western Australia (UWA) recruited 186 men and women aged between 18 and 47 years old from the student population for direct and indirect (photocopy) measurement of fingers and to complete eight various measures of cognition. Demographic information was also collected (age, sex, degree type).

Our variables of interest were: 2D:4D, CCFIT, sex and degree type. One data set per hypothesis was created, such that participants could be in one, both or neither depending on whether they provided the data points relevant to the hypothesis. Of three values for degree (‘arts’, ‘science’ and ‘other’) we compared two (arts and science), believing that these map well onto the domains previously studied (ie. ‘things’ vs ‘people’ and ‘numerical’ vs ‘verbal’) (Diekman et al., 2010; Beltz et al., 2011).

## Participants

In the *H1* sample, there were 96 participants (mean age 21 years), of whom 17 were male, 60 female, with 19 not reporting sex. In *H2*, there were 68 participants (mean age 21 years), of whom 15 were male and 53 female, and of whom 36 were science students and 32 arts students.

## Measures

We used the raw sum of correct scores across four sub-tests in CCFIT Scale 3 Form A (ie. no scaling or reference to norms), a well-respected measure of fluid intelligence (Gregory, 2004 cited in Ruiz, 2009; Dapo & Kalenovic-Dapo, 2012). A Shapiro-Wilks test on the hypothesis one sample CCFIT data suggested normality (W = 0.98, p = 0.07).

Selecting a measure for 2D:4D is rather more complex and this paper considers most justified the mean of two raters’ direct measurements of the right hand ratio. Putz et al.’s unfavourable results used measurements from photocopies (2004), an approach which has since been found to reliably distort the measurement (Manning, Fink & Neave, 2005; Ribeiro et al., 2016). Direct measurements are preferred, particularly when made by experts and of the right hand (Hönekopp & Watson, 2010). A Shapiro-Wilks test on the 2D:4D ratios of both samples suggested the same degree of normality (W = 0.99, p = 0.8). The average measure intra-class correlation based on a mean-rating (*k* =2), absolute agreement, two-way mixed effects model was .59 with a 95% confidence interval from .45 to .70 (*F*(173, 173)= 2.4, p < .001), which represents moderate reliability (Koo & Li, 2016).

## Analysis

Our pre-registered secondary analysis involves a Pearson’s correlation test between 2D:4D and CFIT scores and a Welch’s t-test to compare mean 2D:4D across study groups, both using α = .05. Welch’s t-test is preferred to the student’s t-test as it is robust to non-normal data (Delacre, Lakens & Leys, 2017). Pearson’s r and Cohen’s d are used for comparison to effect sizes in the literature.

Prior to the pre-registered analysis we sought to determine whether our sample replicated the sexual differentiation of 2D:4D and the lack of differentiation in fluid intelligence. We added equivalence tests to complement our correlation and t-test results (Harms & Lakens, 2018). Also, after completing the pre-registered analysis, we further divided the participants in order to investigate within-sex variation as has been done in previous studies (Hell & Päßler, 2010; Bull et al., 2010; Nye & Orel, 2015; Manning, Trivers & Fink, 2017). These were corrected using Bonferroni.

# Results

## Replication of previous differences

Our sample did not reflect the expected finger ratio differentiation. Neither the difference in sample one (*t* (27) = -0.56, *p*= .29, *d*= 0.15) nor sample two (*t* (24) = -0.46, *p*= .32, *d*= 0.13) is close to significance. Our sample did show a difference between sexes on the CFIT measure. However, though our effect size (*t* (27) = 1.61, *p*= .06, *d*= 0.43) was larger than in Lynn and Irwing (*d =* .33), it did not reach significance (2004).

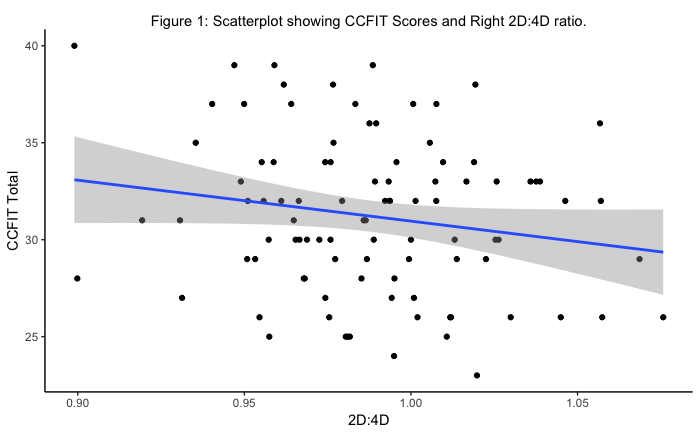
## 2D:4D and CFIT

Firstly, we checked for a relationship between 2D:4D and scores on the CFIT. Previous effect sizes in the 2D:4D literature which reached significance are small: r=-.055 for mean 2D:4D and ‘realistic interests’ in men (Hell & Päßler, 2011); r=-.22 for 2D:4D and arithmetic ability and r=-.23 for 2D:4D and number sense in girls (Bull et al., 2010). A two-sided power analysis suggests the minimum correlation coefficient identifiable in our sample is r=.28.

The raw CFIT scores range from 23 to 40. The overall mean raw score in the sample was 31.21 (*n*=96, *SD*= 4.02) and the male mean (*Mm* = 32.65, *n* = 17, *SD* = 3.87) was greater than the female (*Mf* = 30.92, *n* = 60, *SD* = 4.03). As mentioned above, this difference is not significant at α = .05. The data was overall normally distributed, though not within the male subset. See Table 1.1 in Appendix A for descriptive statistics and Appendix B for distributions.

2D:4D ratio measurements in both samples ranged from 0.9 to 1.08. The overall mean 2D:4D in this sample was 0.99 (*n* = 96, *SD* = 0.03) and the male (*Mm* = 0.98, *n* = 17, *SD* = 0.03) was indeed lower than the female (*Mf* = 0.99, *n* = 60, *SD* = 0.04), though as mentioned, not a significant difference. See Table 1.2 in Appendix A for descriptive statistics.

Between CCFIT score and 2D:4D ratio we found a small, negative but not significant relationship (*r* (94) = -0.18, *CI* = [-0.37, 0.02], *p*= .08), see Figure 1.



Among males there was a medium, negative correlation with lower 2D:4D being associated with higher raw CFIT scores and for females a neglible relationship, neither reaching significance. The power analysis indicates that our study is underpowered for all observed effects (minimum observable whole sample *r* = .28, male *r*=.63, female *r*= .35). See Table 2 for results.

Table 2: Pearson’s correlation results between 2D:4D and CCFIT scores.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *r* | *n* | | 95% CI | *p* |
| Total | -.18 | 96 | -0.37, 0.02 | | .08 |
| Males | -.47 | 17 | -0.78, 0.01 | | .17 |
| Females | -.01 | 60 | -0.27, 0.24 | | 1 |

Note: Bonferroni corrections applied for multiple comparisons.

## 2D:4D and study choice

In Hell and Päßler, effect sizes for sex differences in occupational interests ranged from *d* = 0.12 to 0.80 (2011). Our study has the power to identify a minimum effect size of *d* = 0.63 in the comparison of mean 2D:4D by study choice.

2D:4D measurements in both samples ranged from 0.9 to 1.08. The overall mean in this sample was 0.99 (*n* = 68, *SD* = 0.04) and among science students (*MSc* = 0.98, *n* = 36, *SD* = 0.04) was indeed lower than among arts students (*Ma* = 0.99, *n* = 32, *SD* = 0.03). See Table 3 in Appendix A for descriptive statistics.

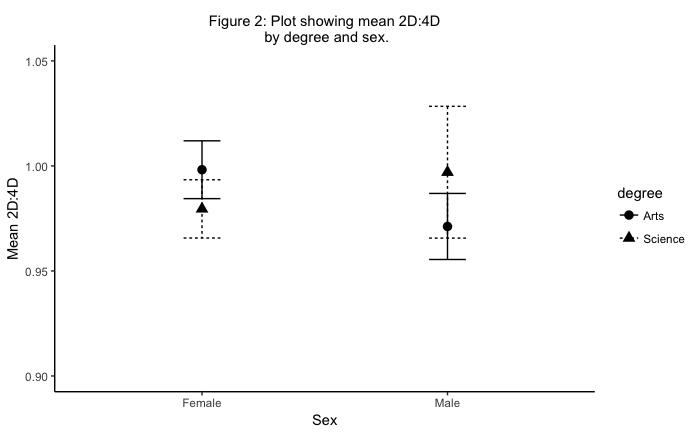
One-tailed Welch’s t-test results are consistent with science students having lower 2D:4D than arts students but it is not significant (*t* (66) = 0.96, *p*= 0.17, *d*= 0.23, 95% *CI* = [-0.25, 0.71]). The observed effect size is lower than our minimum observable (*d* = 0.63), suggesting again our study was underpowered. Results within sex were similarly underpowered, with low degrees of freedom and lacking in significance: for males, running contrary to the alternative hypothesis and for females in line with the alternative hypothesis (minimum observable *d* = 0.73 and *d* = 1.41 respectively). See Table 4 for results.

Table 4: Two-tailed Welch’s t-test results 2D:4D by study choice and sex.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | *t* | *df* | *p* | *d* | 95%CI |
| Total | 0.96 | 66 | .17 | 0.23 | -0.25, 0.71 |
| Male | 1.44 | 9 | 1 | 0.78 | -0.34, 1.84 |
| Female | 1.87 | 51 | .1 | 0.51 | -0.05, 1.07 |

*Note: using the Bonferroni corrected p-values*

This latter relationship suggesting lower digit ratio among females was associated with study of science was perhaps the most interesting of our results, see Figure 2.



A non-significant result does not indicate the truth of the null hypothesis. One aid to interpreting non-significant results is the equivalence test, which asks whether we can conclude the absence of an effect within a given range of effect sizes (Harms & Laken, 2018). Using the smallest effect our study was powered to observe as the smallest effect size of interest (no theoretically justified SESOI being available), (Lakens, Scheel & Isager, 2018), the equivalence test suggests we cannot dismiss the possibility of such an effect in comparing 2D:4D by subject choice (*t* (66) = 1.63, *p* = .054). For the 2D:4D by degree in only the female subset, again we cannot dismiss the possibility of an effect (*t* (51) = 0.783, *p* = .219).

# Discussion

This paper has addressed the questions, firstly, of a relationship between 2D:4D and scores on the CCFIT and secondly, of a meaningful difference in the 2D:4D of those who chose to study science and those who chose arts. Both of our main tests produced effects within the ranges seen in the literature, though results were not significant and the sample size in both cases did not afford sufficient power for the effect sizes identified.

We found, contrary to our hypothesis, a small correlation between CFIT and 2D:4D but this was not comparable to the effects in Lynn and Irwing (relation between sex and Raven’s matrices) or Dapo and Kolenovic-Dapo (relation between 2D:4D and Raven’s matrices) (2004; 2012). This correlation was stronger within the male subset, CFIT increasing with lower 2D:4D, suggesting that pre-natal androgen exposure could be related to increased fluid intelligence among males.

The difference in 2D:4D between arts and science students showed a medium effect comparable to the effect size in differential preference for the Investigative dimension of occupational interests (Beltz et al., 2011).

The results within sex groups also offered weak evidence, though a potentially meaningful result is the medium-sized effect between 2D:4D and study choice among the female subset, suggesting female arts students have higher ratios than female science students with wide confidence intervals. That is to say, female science students in our sample may have more stereotypically ‘masculine’ interests. This places our findings in line with Beltz and colleagues, who found that degree of interest in Things (vs People) varied with pre-natal androgen exposure (2011); distinct from Hell and Päßler who found no relationships with women; but also in line with Nye and Orel’s finding that women with lower 2D:4D than their female average were in careers associated with more ‘masculine’ traits (2015).

## Limitations

Sample size and power have been discussed. These are related to the dataset to which we had access. Underlying our results is a sample, from the UWA dataset, heavily weighted toward the female. Although the primary hypotheses did not refer to sex, the low number of males may render the sample unrepresentative, particularly as the key measure 2D:4D was not normally distributed in that small group, (see Appendix A and a Shapiro-Wilks test (*W* = 0.88, *p* = .04)). This could explain the fact that our sample did not demonstrate reliable differentiation of 2D:4D. Such a lack of clear differentiation has occurred in other research (Bull et al., 2010). Potential confounding factors include ethnicity and sexuality, about which we have no information but both of which relate to 2D:4D (Manning & Fink, 2018; Manning, Trivers & Fink, 2017; Cohen-Bendahan et al., 2005).

Further, the validity of 2D:4D as a proxy for pre-natal androgen exposure remains a matter of contention. There is no theory to distinguish correlated and uncorrelated traits (beyond the timing of differentiation) and the evidence linking ratios to the ultimate variable of interest, pre-natal androgen exposure, is limited (Putz et al., 2004; Cohen-Bandahan et al., 2004). A relation of 2D:4D to genetic polymorphisms coding for androgen receptors was not replicated in a large sample (Hampson & Sankar, 2012). The selection of one measure (right, direct 2D:4D) is justified not by theory but by the strength of previous results (Cohen-Bandahan et al., 2004).

A final limitation of our study is the over-reliance on dated research in the area. Perhaps a more thorough and systematic review of recent literature would have provided more relevant effect sizes for comparison and reinforced the legitimacy of 2D:4D as a proxy.

## Future research

Non-significant results do not entail the absence of an effect and the equivalence tests conducted on selected results suggest the presence of some effects. Larger samples are required in order to more confidently dismiss or support the observed effects in 2D:4D and CCFIT (240) or in 2D:4D and study choice (300) . However, only ten more participants per group would sufficiently increase power for the effect on study choice within the female group. It is plausible that the effects we found exist and our study was simply unable to offer stronger evidence. At the moment, we neither reject nor affirm our null hypotheses.

# Conclusion

This paper’s contribution is in testing 2D:4D as a predictor of two new outcomes and identifying potential relationships and size of sample required for further study. The present results are inconclusive re pre-natal androgen exposure’s contribution to potential differentiation in fluid intelligence and study choice. More robust study design could reward attention to: the correlation between CCFIT and 2D:4D within sex and the apparent increased likelihood of studying science among females with lower 2D:4D. Nonetheless, we also add to the voices urging caution regarding the us of biomarkers as predictors in real world issues, particularly as effect sizes are so often small and non-significant (Turanovic et al., 2017).

References

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits.*Psychological Bulletin, 121*(2), 219-245. doi:10.1037//0033-2909.121.2.219

Beede, D., Julian, T., Langdon, D.,McKittrick, G., Khan, B., & Doms,M. (2011). Women in STEM: A gender gap to innovation. Washington, DC: U.S. Department of Commerce, Economics and Statistics Administration.

Beltz, A. M., Swanson, J. L., & Berenbaum, S. A. (2011). Gendered occupational interests: Prenatal androgen effects on psychological orientation to things versus people. Hormones and Behavior, 60(4), 313-317. doi:10.1016/j.yhbeh.2011.06.002

Berenbaum, S. A. (1999). Effects of early androgens on sex-typed activities and interests in adolescents with congenital adrenal hyperplasia.*Hormones and Behavior, 35*(1), 102-110. doi:10.1006/hbeh.1998.1503

Boyle, G. J., Stankov, L., Martin, N. G., Petrides, K. V., Eysenck, M. W., & Ortet, G. (2016). Hans J. Eysenck and Raymond B. Cattell on intelligence and personality. Personality and Individual Differences, 103, 40-47. doi:10.1016/j.paid.2016.04.029

Bull, R., Davidson, W. A., & Nordmann, E. (2010). Prenatal testosterone, visual-spatial memory, and numerical skills in young children. Learning and Individual Differences, 20(3), 246-250. doi:10.1016/j.lindif.2009.12.002

Cai, H., Luo, Y. L. L., Shi, Y., Liu, Y., & Yang, Z. (2016). Male = science, female = humanities: Both implicit and explicit gender-science stereotypes are heritable. Social Psychological and Personality Science, 7(5), 412-419. doi:10.1177/1948550615627367

Cohen-Bendahan, C. C. C., van de Beek, C., & Berenbaum, S. A. (2005). Prenatal sex hormone effects on child and adult sex-typed behavior: Methods and findings. Neuroscience and Biobehavioral Reviews, 29(2), 353-384. doi:10.1016/j.neubiorev.2004.11.004

Colom, R. and García-Lopez, O. (2002), Sex differences in fluid intelligence amongst high school graduates. Personality and Individual Differences, 32, 445-451.

Đapo, N., & Kolenović-Đapo, J. (2012). Sex differences in fluid intelligence: Some findings from bosnia and herzegovina.*Personality and Individual Differences, 53*(7), 811-815. doi:10.1016/j.paid.2012.05.036

Delacre, M., Lakens, D., & Leys, C. (2017). Why Psychologists Should by Default Use Welch’s *t*-test Instead of Student’s *t*-test. *International Review of Social Psychology*, *30*(1), 92–101. DOI: <http://doi.org/10.5334/irsp.82>

Diekman, A. B., Brown, E. R., Johnston, A. M., & Clark, E. K. (2010). Seeking congruity between goals and roles: A new look at why women opt out of science, technology, engineering, and mathematics careers.*Psychological Science, 21*(8), 1051-1057. doi:10.1177/0956797610377342

Galis, G., Broek, C. M. A. T., Dongen, S. V., & Wijnaendts, L. C. D. (2010). Sexual dimorphism in the prenatal digit ratio (2D:4D).*Archives of Sexual Behavior, 39*(1), 57-62. doi:10.1007/s10508-009-9485-7

Halpern, D. F. (1997). Sex differences in intelligence: Implications for education.*American Psychologist, 52*(10), 1091-1102. doi:10.1037//0003-066X.52.10.1091

Hampson, E., & Sankar, J. S. (2012). Re-examining the Manning hypothesis: Androgen receptor polymorphism and the 2D:4D digit ratio.*Evolution and Human Behavior, 33*(5), 557-561. doi:10.1016/j.evolhumbehav.2012.02.003

Harms, C., & Lakens, D. (2018) Making 'null effects' informative: Statistical techniques and inferential frameworks. *Journal of Clinical and Translational Research, 3*, 382-393, http://dx.doi.org/10.18053/jctres.03.2017S2.007

Hell, B. and Päßler, K. (2011). Are occupational interests hormonally influenced? the 2D:4D-interest nexus. Personality and Individual Differences, 51(4), 376-380. doi:10.1016/j.paid.2010.05.033

Hönekopp, J., & Watson, S. (2010). Meta-analysis of digit ratio 2D:4D shows greater sex difference in the right hand.*American Journal of Human Biology : The Official Journal of the Human Biology Council, 22*(5), 619-630. doi:10.1002/ajhb.21054

Kaufman, A. S., Kaufman, J. C., Liu, X., & Johnson, C. K. (2009). How do educational attainment and gender relate to fluid intelligence, crystallized intelligence, and academic skills at ages 22-90 years?*Archives of Clinical Neuropsychology, 24*(2), 153-163. doi:10.1093/arclin/acp015

Kent, P. (2017). Fluid intelligence: A brief history.*Applied Neuropsychology: Child, 6*(3), 193-203. doi:10.1080/21622965.2017.1317480

KIDD, A. H. (1962). The culture-fair aspects of cattell's test of g: Culture-free. The Journal of Genetic Psychology, 101(2), 343-362. doi:10.1080/00221325.1962.10533635

Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine*, *15*(2), 155–163. doi:10.1016/j.jcm.2016.02.012

Lakens, D., Scheel, A. M., & Isager, P. M. (2018). Equivalence testing for psychological research: A tutorial.*Advances in Methods and Practices in Psychological Science, 1*(2), 259-269. doi:10.1177/2515245918770963

Lippa, R. A. (2006a). The gender reality hypothesis.*The American Psychologist, 61*(6), 639-640. doi:10.1037/0003-066X.61.6.639

Lippa, R. A. (2006b). Finger lengths, 2D:4D ratios, and their relation to gender-related personality traits and the big five.*Biological Psychology, 71*(1), 116-121. doi:10.1016/j.biopsycho.2005.02.004

Luxen, M. F., & Buunk, B. P. (2005). Second-to-fourth digit ratio related to verbal and numerical intelligence and the big five. Personality and Individual Differences, 39(5), 959-966. doi:10.1016/j.paid.2005.03.016

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

Manning, J. T., & Fink, B. (2018). Sexual dimorphism in the ontogeny of second (2D) and fourth (4D) digit lengths, and digit ratio (2D:4D). American Journal of Human Biology, 30(4), e23138-n/a. doi:10.1002/ajhb.23138

Manning, J. T., Fink, B., Neave, N., & Caswell, N. (2005). Photocopies yield lower digit ratios (2D:4D) than direct finger measurements.*Archives of Sexual Behavior, 34*(3), 329-333. doi:10.1007/s10508-005-3121-y

Manning, J. T., Scutt, D., Wilson, J., & Lewis-Jones, D. I. (1998). The ratio of 2nd to 4th digit length: A predictor of sperm numbers and concentrations of testosterone, luteinizing hormone and oestrogen.*Human Reproduction (Oxford, England), 13*(11), 3000-3004. doi:10.1093/humrep/13.11.3000

Manning, J. T., Trivers, R., & Fink, B. (2017). Is digit ratio (2D:4D) related to masculinity and femininity? evidence from the BBC internet study.*Evolutionary Psychological Science, 3*(4), 316-324. doi:10.1007/s40806-017-0098-4

Neisser, U., Boodoo, G., Bouchard, T. J., Jr, Boykin, A. W., Brody, N., Ceci, S. J., . . . Urbina, S. (1996). Intelligence: Knowns and unknowns. American Psychologist, 51(2), 77-101. doi:10.1037//0003-066X.51.2.77

Nenty, H. J., & Dinero, T. E. (1981). A cross-cultural analysis of the fairness of the cattell culture fair intelligence test using the Rasch model. Applied Psychological Measurement, 5(3), 355-368. doi:10.1177/014662168100500309

Nye, J. V. C., & Orel, E. (2015). The influence of prenatal hormones on occupational choice: 2D:4D evidence from moscow. Personality and Individual Differences, 78, 39-42. doi:10.1016/j.paid.2015.01.016

Putz, D. A., Gaulin, S. J. C., Sporter, R. J., & McBurney, D. H. (2004). Sex hormones and finger length: What does 2D:4D indicate? Evolution and Human Behavior, 25(3), 182-199. doi:10.1016/j.evolhumbehav.2004.03.005

Ribeiro, E., Neave, N., Morais, R. N., & Manning, J. T. (2016). Direct versus indirect measurement of digit ratio (2D:4D): A critical review of the literature and new data.Evolutionary Psychology, 14(1), 147470491663253. doi:10.1177/1474704916632536

Ruiz, P. E. (2009). Measuring fluid intelligence on a ratio scale: Evidence from nonverbal classification problems and information entropy.*Behavior Research Methods, 41*(2), 439-445. doi:10.3758/BRM.41.2.439

Sánchez, Á., Sánchez-Campillo, J., Moreno-Herrero, D., & Rosales, V. (2014). 2D:4D values are associated with mathematics performance in business and economics students.Learning and Individual Differences, 36, 110-116. doi:10.1016/j.lindif.2014.10.001

Sullivan, C. J., & McGloin, J. M. (2014). Looking back to move forward: Some thoughts on measuring crime and delinquency over the past 50 years. Journal of Research in Crime and Delinquency, 51(4), 445-466. doi:10.1177/0022427813520446

von Stumm, S., & Ackerman, P. L. (2013). Investment and intellect: A review and meta-analysis.*Psychological Bulletin, 139*(4), 841-869. doi:10.1037/a0030746

Turanovic, J. J., Pratt, T. C., & Piquero, A. R. (2017). Exposure to fetal testosterone, aggression, and violent behavior: A meta-analysis of the 2D:4D digit ratio.*Aggression and Violent Behavior, 33*, 51-61. doi:10.1016/j.avb.2017.01.008

Wong, W. I., & Hines, M. (2016). Interpreting digit ratio (2D:4D)–behavior correlations: 2D:4D sex difference, stability, and behavioral correlates and their replicability in young children. Hormones and Behavior, 78, 86-94. doi:10.1016/j.yhbeh.2015.10.022

# Appendix A: Descriptive statistics.

*Table 1.1*: Descriptive statistics for CFIT scores in sample 1.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | n | Mean Age | CCFIT Mean | CCFIT Median | CCFIT SD | CCFIT Skew | CCFIT Kurtosis | CCFIT SE | CCFIT UpperCI | CCFIT LowerCI |
| Total | 96 | 21.06 | 31.21 | 31 | 4.02 | 0.15 | -0.76 | 0.41 | 32.01 | 30.4 |
| Male | 17 | 22.53 | 32.65 | 33 | 3.87 | -0.06 | -1.05 | 0.94 | 34.49 | 30.81 |
| Female | 60 | 20.65 | 30.92 | 31 | 4.03 | 0.04 | -0.87 | 0.52 | 31.94 | 29.90 |
| NA | 19 | NaN | 30.84 | 30 | 4.06 | 0.74 | -0.45 | 0.93 | 32.67 | 29.02 |

*Table 1.2:* Descriptive statistics for 2D:4D in sample 1.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | n | Mean Age | Digits Mean | Digits Median | Digits SD | Digits Skew | Digits Kurtosis | Digits SE | Digits UpperCI | Digits LowerCI |
| Total | 96 | 21.06 | 0.99 | 0.99 | 0.03 | 0.1 | 0.1 | 0 | 1 | 0.98 |
| Male | 17 | 22.53 | 0.98 | 0.98 | 0.03 | 1.17 | 1.11 | 0.01 | 1.00 | 0.97 |
| Female | 60 | 20.65 | 0.99 | 0.99 | 0.04 | 0.05 | -0.21 | 0.00 | 1.00 | 0.98 |
| NA | 19 | NaN | 0.99 | 1.00 | 0.04 | -0.58 | 0.28 | 0.01 | 1.01 | 0.97 |

*Table 3:* Descriptive statistics for 2D:4D and study choice.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 5 | n | Mean Age | Digits Mean | Digits Median | Digits SD | Digits Skew | Digits Kurtosis | Digits SE | Digits UpperCI | Digits LowerCI |
| Total | 68 | 21.12 | 0.99 | 0.98 | 0.04 | 0.33 | -0.24 | 0.00 | 1 | 0.98 |
| Science | 36 | 21.33 | 0.98 | 0.98 | 0.04 | 0.34 | -0.10 | 0.01 | 1 | 0.97 |
| *Male* | 7 | 23.14 | 1.00 | 0.99 | 0.04 | 0.70 | -1.03 | 0.02 | 1.03 | 0.97 |
| *Female* | 29 | 20.90 | 0.98 | 0.98 | 0.04 | 0.18 | -0.33 | 0.01 | 0.99 | 0.97 |
| Arts | 32 | 20.88 | 0.99 | 0.99 | 0.03 | 0.43 | -0.88 | 0.01 | 1 | 0.98 |
| *Male* | 8 | 22.25 | 0.97 | 0.97 | 0.02 | 0.36 | -1.56 | 0.01 | 0.99 | 0.96 |
| *Female* | 24 | 20.42 | 1.00 | 0.99 | 0.03 | 0.22 | -1.11 | 0.01 | 1.01 | 0.98 |

# Appendix B: Figures representing distributions.

Figure 3. Distribution of CCFIT scores.

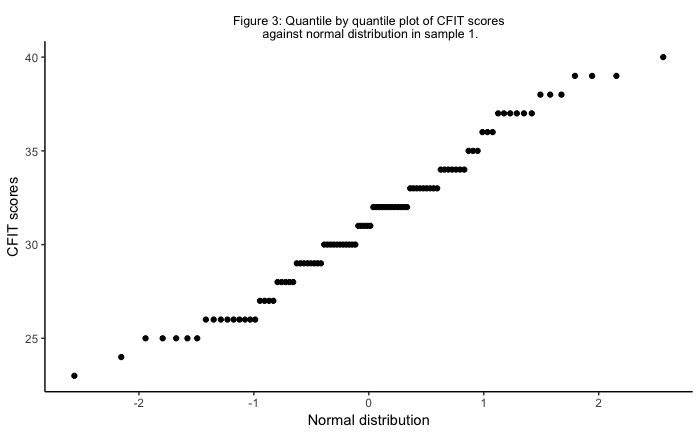


Figure 4. Distribution of 2D:4D scores in sample 1.

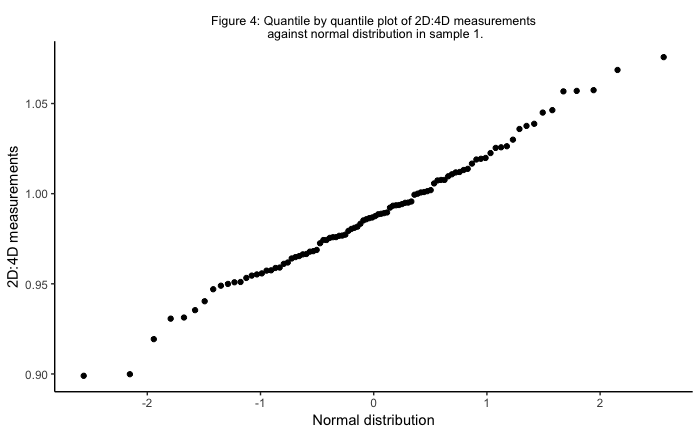


Figure 5. Distribution of 2D:4D scores in men in sample 1.

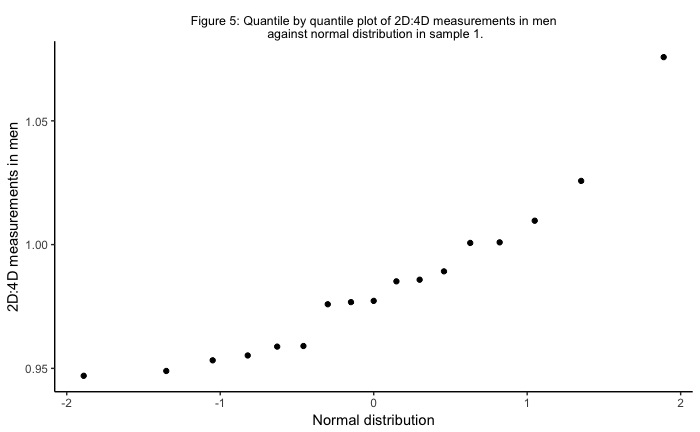
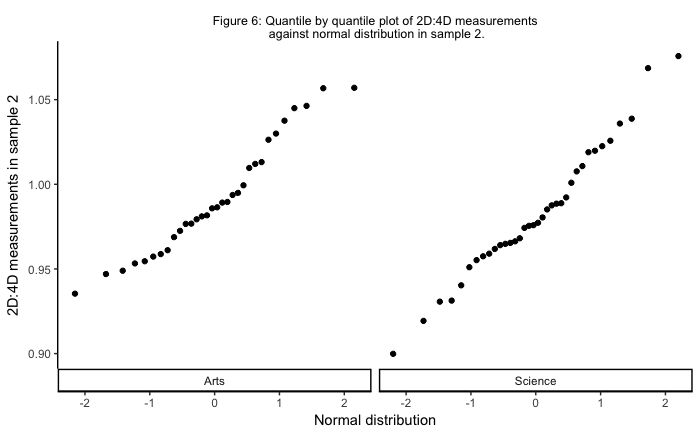


Figure 6. Distribution of 2D:4D scores by study choice in sample 2.



# Appendix C: Pre-registration

**What are your research questions and hypotheses?**

RQ1: Does the 2D:4D ratio, considered a proxy for exposure to prenatal androgens, correlate with ‘fluid intelligence’ as measured by the Culture Fair Intelligence Test?

H1: We predict that there will be no significant correlation between 2D:4D ratio and scores on the CFIT.

RQ2: Are individuals with a lower 2D:4D ratio more likely to pursue science degrees than arts degrees at university?

H2: We predict that individuals with lower 2D:4D ratios will be significantly more likely to pursue science degrees than arts degrees.

**How many observations will be collected or what will determine the sample size?**

We will be conducting a secondary analysis on an existing data set. Sample size is, therefore, out of our control. It is our intention to use a power analysis to determine the smallest possible effect sizes identifiable, given the sample we have. We will then be able to assess whether our study was sufficiently powered to confidently identify any observed effects.

Research into 2D:4D correlations has looked at other variables: Hell and Päßler (2011) found small effect sizes in a large sample, few of which reached significance (r=-.055 for 2D:4Dmean and ‘realistic interests’ in men was significant). Larger, significant correlations were found by Bull et al. (2010) between 2D:4D and arithmetic (r=-.38 for males) and 2D:4D and number sense (r=.44 for females).

**What are the variables that will be included in the analyses and how they are calculated?**

H1:

Variable 1: individual scores (total number of items correct) on the CFIT.

Variable 2: right hand 2D:4D. This is measured on approved photocopies and actual hands, using digital callipers and by two independent raters (from the mid-point of the basal crease to the mid-point of the finger-tip).

H2:

IV: right hand 2D:4D ratio. As it’s a ratio (based on millimetres), there are no units to report.

DV: study choice (the data set contains three levels, of which we will use two, ‘arts’ and ‘science’).

**What assumption checks will be performed and what you will do if your data fail these assumption checks?**

H1:

Assumption: When conducting a Pearson’s correlation (between CFIT scores and 2D:4D scores), we need paired values.

Check: ensure data set is complete for the relevant variables.

Assumption: When conducting a Pearson’s correlation, we assume a linear relationship between the two quantitative variables.

Check: a scatterplot of the two variables. If it appears the relationship is not linear (as is true in some research involving 2D:4D), we can apply Spearman’s rho instead (or check for a relationship with 2D:4d2 as Sanchez et al. did)?

Assumption: using Pearson’s r for significance checks requires that the data are normally distributed.

Check: boxplot and violinplot visualisations plus descriptive statistics. If the data are not normally distributed, we can use Spearman’s rho instead.

Assumption: the raters reliably measured the 2D:4D ratio. We plan to use the right hand, direct measure (mean between 1st and 2nd rater).

Check: find a reliability score for the ratings. If they’re not reliable… acknowledge in the discussion/limitations?

H2:

Assumption: There will be no outliers in the data.

Check: 1) look at data in violin plot to assess distribution and identify potential outliers; 2) investigate the potential outliers to see examine their cause (eg badly entered data) 3) see how the outliers effect the assumptions and results 4) decide whether to run the analysis with or without the outliers.

**What do you know about missing data in the dataset already? How will you deal with missing or incomplete data?**

We will delete participants who do not have the data required for our analyses. However, as the analyses are independent, these deletions will only be made in the analysis for which the data is missing (ie. one participant may be present in the analysis for hypothesis 1 but not hypothesis 2). This will help to preserve the largest sample size to help with power.

**What descriptive statistics will you calculate for each variable?**

Right hand 2D:4D: N, mean, median, standard deviation, skew, kurtosis.

CFIT score: N, mean, median, standard deviation, skew, kurtosis.

Study choice: number in arts and number in science.

**What kind of plots will you create?**

Plot 1: boxplot and violinplot showing distribution of CFIT scores.

Plot 2: boxplot and violinplot showing distribution of right hand 2D:4D ratios.

Plot 3: scatterplot of CFIT scores and right hand 2D:4D ratios.

Plot 3: violinplot of 2D:4D by study choice.

**What inferential analyses will you conduct for each hypothesis?**

Hypothesis 1 will involve a Pearson’s r correlation.

Hypothesis 2 will involve a Welch’s t-test (Delacre, Lakens, Leys, in press).

**What inference criteria will you use?**

We will use an alpha of .05 as the threshold for significance. As we are not going to perform more than one test, we will not be applying multiple test corrections. We will calculate effect sizes using Cohen’s d.