The Role of Gender, Task Success Probability Estimation and Scores as Predictors of the Domain-Masculine Intelligence Type (DMIQ).

Josephine Storek[a] and Adrian Furnham[ab]

[a] Research Department of Clinical, Educational and Health Psychology, University College London

[b] Norwegian Business School, Nydalsveien 27, Oslo, Norway

Abstract

This paper reports two experimental studies aimed at understanding correlates of self-estimated intelligence. In both studies participants twice estimated their mathematical and spacial intelligence (called Domain Masculine Intelligence Type: DMIQ) on a normal distribution. They also completed a number of short numerical and logical ability tests after which they estimated their performance at a similar, more difficult task. Males gave higher estimates than females and did better on the tests. Their estimates of their DMIQ reduced on the second occasion after testing. Gender, task score and estimated performance were all significant predictors of both DMIQ scores. Task confidence was the best predictor of DMIQ1 and DMIQ2, over and above gender and test score, explaining 17% and 23% of variance respectively. Results are discussed in terms of the expanding literature on self-estimated intelligence

Introduction

While an extensive body of self-estimated intelligence (SEI) research is available, only few SEI studies used psychometric measures to compare the accuracy and validity of SEI estimates (e.g. Batey et al., 2009; Chamorro-Premuzic, Moutafi, & Furnham, 2005; Furnham & Mottabu, 2004; Holling & Preckel, 2005). To our knowledge this research is the first experimental design in SEI that focuses on assessing gender differences in self-estimated intelligence using ability tests, repeated measurement as well as investigating the role of task confidence.

Evidence from more than thirty studies shows that stable and consistent universal gender differences in SEI exist in general population (Furnham, 2001; Furnham & Budhani, 2002; Furnham, Crawshaw, & Rawles, 2006; Furnham & Shagabutdinova, 2012; Stieger et al., 2010; von Stumm, Chamorro-Premuzic, & Furnham, 2009), with the strongest gender differences observed on mathematical/ logical and spatial intelligences, followed by overall (‘*g*’) and verbal intelligences, with significantly higher self-estimates provided by males than females (Furnham, 2001; Furnham, Hosoe, & Tang, 2003; Rammstedt & Rammsayer, 2002a,b). The magnitude of gender differences in mathematical/logical, spatial, overall and verbal self-assessed intelligences were further revealed in meta-analytical study (Szymanowicz, Chamorro-Premuzic, & Furnham, 2011), with the biggest weighted mean effect sizes for mathematical/logical, (d = .44), followed by spatial (d = .43), overall (d =.37) and verbal (d =.07) intelligences, with males providing higher estimates in all but verbal intelligence.

This phenomenon is known as the *‘hubris-humility effect’* (HHE) (Beloff, 1992; Storek & Furnham, 2012, 2013a,b). It is unclear whether HHE correctly depicts male and female understanding of their cognitive abilities or whether the inflated and deflated self-perceptions impact one’s behaviour and performance. Equally, it remains to be answered whether female *‘humility’* is a reflection of an accurate female self-estimation or whether it is a direct outcome of negative female self-assessments, performance expectancies, stereotypical self- beliefs or low self-confidence. In fact, female self-estimates were shown to be significantly more accurate than were males’. Male self-estimates were significantly inflated compared to their actual psychometric scores (Rammstedt & Rammsayer, 2002a,b; Reilly & Mulhern, 1995). These findings were further substantiated by Carr et al. (2008) who reported that girls were more accurate in assessing their mathematical skills and knowledge, despite low math ability confidence. Unsurprisingly, boys were overconfident, with poor performance.

To further explore the '*male-normative*' perception of intelligence (Furnham, 2000), the '*domain-masculine intelligence type*' (DMIQ) which is a composite of mathematical/logical and spatial intelligences (Storek & Furnham, 2012, 2013 ab) was introduced. Accordingly, the investigation of the relationship between DMIQ and HHE and DMIQ's role in the prediction of HHE as well as the confirmation of DMIQ as the most sensitive predictor of gender differences in SEI and identification of HHE determinants were central to this research.

This research summarises two experimental studies that were designed to ascertain the determinants of gender differences in the '*domain-masculine intelligence type* (DMIQ) by introducing a number of timed psychometric tasks (TCAP) and confidence assessments (TSP). As in previous research (e.g.: Storek & Furnham, 2012, 2013 ab) gender was expected as the best predictor of DMIQ. The experimental design allowed for in-depth examination of the role gender plays in the repeated measurement of DMIQ as well as in the relationships between DMIQ and TCAP and DMIQ and TSP. Equally, gender differences in TCAP and TSP were examined in an attempt to understand the conflicting claims in current literature and to clarify whether they have any bearing on the gender differences in the intelligence type.

Although the studies were identical in overall design and execution, the content and format of the psychometric task and the number of task-success probes differed per study. This was done to test whether alternating numerical, reasoning, spatial and crystallised knowledge problems and varying the number of TSP probes impacts on the DMIQ estimation process, the hubris-humility effect and the role of gender herein. In addition, TCAP content alternation was expected to be gender-stereotype inducing as it contained items that are perceived as domain masculine, especially by females.

Study 1

This is the first of two experimental studies that sets out to examine the existence of HHE on DMIQ at both pre- and post-task estimation conditions. Repeated measures are included to validate assertions that they influence behaviour and performance and as such change mood and confidence (Bartsch & Nesselroade, 1973). Gender-stereotypes and self-confidence are likely to play a role in HHE or the display of male hubris and female humility in estimation of abilities. Paricipants were asked to undertake a gender-stereotype inducing task, i.e. numerical and reasoning aptitude problems that are likely to increase hubris and humility (Betsworth, 1999; Beyer, 1990, 1998; Dar-Nimrod, 2007; Ehrlinger & Dunning, 2003; Hoffman & Hurst, 1990;

Steele & Aronson, 1995) as well as task-success estimates or confidence probes that will enable the assessment of confidence (Burson et al., 2006; Carr et al., 2008; Dunning et al., 1990; Pallier, 2003). After each block, participants were asked to estimate their task-success confidence.

Thus, it was predicted that HHE will be confirmed on DMIQ at the pre-task (T1) and post-task (T2) estimating conditions (H1) and that there will be a significant decrease in DMIQ estimates from T1 to T2 following the gender-stereotype inducing task (H2).

Existing literature suggests that males have higher self-confidence, despite being inaccurate about their (math) skills or underperforming, whereas females are lacking confidence, while being accurate or outperforming males (Carr et al., 2008; Eccles-Parsons et al., 1984; Pallier, 2003). Consequently, males are expected to provide significantly higher task-success probability estimations (TSP) than females (H3).

Given the ample evidence about sex differences in cognitive abilities (Halpern et al., 2007; Hyde, Fennema, & Lamon, 1990a; Jackson & Rushton, 2006; Lynn & Irwing, 2004; Ogle et al., 2003; Novell & Hedges, 1998; Voyer, Voyer, & Bryden, 1995), sex differences are expected on the numerical and reasoning problems (TCAP), with males providing more correct answers than females (H4).

Gender is expected to be the best predictor of DMIQ1 (H5) and DMIQ2 (H6) over and above TSP and TCAP. Finally, gender is presumed to influence the relationship between TSP and DMIQ1 (H7) and DMIQ2 (H8). Gender is also expected to affect the relationship between TCAP and DMIQ1 (H9) and DMIQ2 (H10).

Method

Participants

A total of four hundred and eighty-eight participants from general public took part in this experimental online study. There were 326 females (67%) and 164 males. Their age ranged from 17 to 70 (M = 22.33, SD = 6.86) years. All participants were fluent in English and no language or other problems were reported.

Measures

Repeated Measure of Domain-Masculine Intelligence Type (DMIQ).

Based on the self-estimated intelligence measure (Furnham & Gasson, 1998) this shortened version had the same properties and layout, but only included mathematical/logical and spatial intelligences that together form the Domain-Masculine Intelligence Type.

Participants were shown a bell curve with IQ scores and asked to estimate their mathematical/logical and spatial intelligences, which were provided with detailed descriptions. Participants were asked to estimate their mathematical/logical and spatial intelligences on two occasions, prior (T1) and post (T2) to completing a psychometric task (TCAP) and assessing their task-success confidence (TSP). Individual scores for DMIQ were computed. Alpha for DMIQ1 was .82 and for DMIQ2 .88.

Psychometric Aptitude Task - Total Correct Aptitude Problems (TCAP)
Numerical and Reasoning Problems (Bryon, 2006)

Fifteen numerical and reasoning problems that were taken from an intelligence test training book were presented in five blocks of three analogous problems (Bryon, 2006). Participants were informed that items in each block varied in difficulty level, ranging from elementary to difficult. A time limit of 90 seconds was given for each block of problems. Participants were advised to leave unanswered problems blank, in order not to exceed the time limit, or face disqualification. The time limit was set to reflect a real-life intelligence testing situation, with the entire task taking 7.5 minutes to complete. Correct answers were available at the end of the survey. Alpha for the fifteen items was .93.

Task Success Probability Estimation Measure (TSP) (Storek, 2007)

After each problem block, participants were asked to indicate how likely they felt they would succeed on a similar task but with increased difficulty, e.g.*" Using the scale, indicate how likely you are to succeed on the same task, but with increased difficulty"* using a rating scale where 1 was *Very Unlikely* and 5 *Very Likely.* The five task success probability statements made up the Task Success Probability measure, with individual scores computed for all participants. The alpha for the five-item measure was .82. As such, the measure was a calibration measure of individual differences.

Procedure

Participants were members of public who were recruited to participate in an online experiment with the help of a snow-balling technique. The data was gathered through an online survey engine and participation was voluntary. Detailed scoring instructions were given at the beginning of each measure, including timing instructions for the psychometric problems. Debrief feedback, correct answers and an opportunity to leave comments about the survey was provided. Ethical permission was applied for and granted.

Results

Domain-masculine intelligence and the Hubris and Humility Effect in T1 and T2

An independent samples t-tests, t(385) = 6.16, p = .001, two-tailed, confirmed significant differences between males (M =120.64, SD = 18.13) and females (M = 108.55, SD = 18.70) in the DMIQ at T1. The magnitude of differences in the means (Means Difference = 12.09, 95% CI: 8.23 to 15.95) was large (η^2 =.09, Hedge's Adjustment *d* = .66).

An independent samples t-tests, t(227) = 4.68, p = .001, two-tailed, confirmed significant differences between males (M =116.02, SD = 21.58) and females (M = 102.57, SD = 21.14) in the DMIQ at T2. The magnitude of differences in the means (Means Difference = 13.56, 95% CI: 7.79 to 19.12) was large (η^2 =.09, Hedge's Adjustment *d* = .63). Hypothesis 1 was confirmed.

A paired samples t-test[1] was conducted to test whether DMIQ estimates decreased significantly from T1 to T2. There was a statistically significant decrease in DMIQ from T1 (*M* = 113.49, *SD* = 19.40) to T2 (*M* = 108.21, *SD* = 22.04), t(224) = 5.66, *p* = .00, two-tailed, *r* = .78, *p* =.00. The mean decrease in domain-masculine intelligence self-estimates was 5.28 (14.00) with 95% CI: 3.44 to 7.12. Cohen's *d* statistic (.38) indicated a small effect size. Hypothesis 2 was confirmed.

Gender Differences in Task Success Probability Estimation (TSP) and Psychometric Aptitude Task (TCAP)

Table 1 gives an overview of independent-samples t-tests and effect sizes for the five individual TSP probes and the overall TSP measure. With the exception of TSP4, the independent-samples t-tests were significant, with males providing higher TSP estimates than

[1] Paired t-test is used when the samples are dependent, i.e. when there is only one sample that has been tested twice (repeated measures) or when there are two samples that have been matched or "paired". The appropriate equation is $t = \bar{X}_D - \mu 0 / s_D / \sqrt{n}$. The differences between all pairs must be calculated. The pairs are either one person's pre-test and post-test scores or between pairs of persons matched into meaningful groups. The average (X_D) and standard deviation (s_D) of those differences are used in the equation. The constant μ_0 is non-zero if one needs to test whether the average of the difference is significantly different from μ_0. The degree of freedom used is $n-1$

females. The observed effect sizes were small. Inspection of the correlational results (see Table 2) revealed a negative correlation between gender and TSP (TSP) (r = -.18, p <.01), with males providing higher TSP estimates than females (M_{Males} = 3.18, SD_{Males} = .80; $M_{Females}$ = 2.88, $SD_{Females}$ = .81).Hypothesis 3 was confirmed.

Equally, inspection of the correlational results (see Table 2) revealed a small negative correlation between gender and Total Correct Aptitude Problems (TCAP), (r =-.18, p =.00), with males correctly solving more problems then females. An independent-samples t-test for TCAP revealed significant gender differences t(307) = 3.96, p =.00, two-tailed between males (M_{Males} = 5.47, SD_{Males} = 4.60) and females ($M_{Females}$ = 3.77, $SD_{Females}$ = 4.27). The magnitude of the differences in the means (mean difference = .43, 95% CI: .86 to 2.55) was small (η^2 = .05; Hedge's Adjustment =.01).

Insert Table 1 here

2x2 χ^2 tests[2] and effect sizes for the 5x3 numerical and reasoning problem blocks were computed. Out of fifteen problems, significant gender differences were observed on twelve problems. Despite the unequal gender distribution (67% of participants were females), more males solved correctly the psychometric problems. Phi coefficient effect sizes, using Cohen's effect size criteria (1988), were small. Hypothesis 4 was confirmed.

Gender, Task-Success Probability (TSP) and Total Correct Aptitude Problems (TCAP) as Predictors of DMIQ1 and DMIQ2

Firstly, the relationships between the DMIQ1 and DMIQ2, gender, TSP and TCAP were explored. Table 2 shows the results of the correlational and partial correlational analyses. DMIQ1 and DMIQ2 were strongly intercorrelated (r = .78, p =.00). Gender correlated negatively (r = -.30, p =.00), with DMIQ1 as well as DMIQ2 (r = -.30, p =.00), with females providing lower scores than males. A positive relationship was observed between DMIQ1 and TSP (r = .47, p =.00) and DMIQ2 and TSP (r = .62, p =.00). DMIQ1 also correlated positively with TCAP (r= .16, p <.01) as did DMIQ2 (r= .40, p =.00). The correlations between TSP, TCAP and DMIQ2 were stronger than with DMIQ1. A medium positive correlation was observed between TSP and TCAP (r =.43, p =.00).

[2] $\chi^2_{(1)} = Z^2 = r^2{+}N$. Phi ($\varphi$) is the best measure of association for χ^2 test (2x2 contingency table); it estimates the extent of the relationship between the variables. For a 2x2 matrix the following formula is used: $\varphi = \sqrt{}\ \chi^2 / N$, where N is the number of subjects

Insert Table 2 here

As in previous studies (Storek & Furnham, 2012, 2013a,b), the role of age in the DMIQ estimation process was examined. Despite the wide age range (53 years), no significant relationships were observed between age and DMIQ1 and DMIQ2. A negative relationship was observed between age and gender (r= -.14, p <.01) indicating that females in this sample were younger than males. A positive relationship between age and TCAP (r= .12, p =.01) indicated that older participants solved more TCAP problems. This finding is contrary to assertions that fluid cognitive ability declines with age (e.g. Beier & Ackerman, 2001, 2003; Deary et al., 2003).

The correlations were re-run, with age partialled out. Preliminary analyses were performed to ensure no violation of the assumptions of normality, linearity and homoscedasticity. An inspection of the zero order correlation matrix suggested that controlling for age had little impact on the strength of the observed relationships, with values slightly higher.

Subsequently, the data was split per gender and the correlational analysis recomputed. The results are presented in Table 3. TSP displayed a strong positive relationship with DMIQ1 and DMIQ2 for both genders, with stronger correlations between TSP and DMIQ2 than between TSP and DMIQ1. Medium positive correlations were observed between TCAP and DMIQ2 for both genders, but no significant relationships were observed between TCAP and DMIQ1. These findings indicate that the relationships between TSP and TCAP and DMIQ became stronger following the task.

Insert Table 3 here

Gender as the best predictor of DMIQ1 and DMIQ2

To determine the best predictor of DMIQ1 and DMIQ2 simultaneous multiple regressions were performed. Results are reported in Table 4.

The first model predicting DMIQ1 was significant F(3,212) = 26.48, p =.00, Adjusted R^2 =.26, f^2=.37), with the overall model explaining 27% of total variance. Gender (β = -.23, p =.00, r_{part} = -.22) and TSP (β = .46, p =.00, r_{part} = .41) were significant predictors of DMIQ1,with gender accounting for 5% and TSP for 17% of variance. TCAP did not significantly contribute to the prediction of DMIQ1. Contrary to prediction, TSP and not gender was the best predictor of the DMIQ1. Hypothesis 5 was not supported.

The second model, predicting DMIQ2 was also significant F(3,205) = 53.43, p =.00, Adjusted R^2 = .43, f^2=.79), with the overall model explaining 44% of total variance. Gender (β = -.18, p <.01, r_{part} = -.17), TSP (β = .54, p =.00, r_{part} = .48) and TCAP (β = .14, p <.05, r_{part} = .12) were significant predictors, explaining 3%, 23% and 1% of variance respectively. As in DMIQ1, TSP, and not gender, was the best predictor of DMIQ2. Hypothesis 6 was also not supported.

Insert Table 4 here

Impact of Gender on the Relationship between TSP and DMIQ1 and DMIQ2

Two 2-way between-groups analysis of variance were conducted to explore whether gender influences the relationship between TSP and DMIQ1 and DMIQ2. Results are presented in Table 5. For DMIQ1, Levene's Test of Equality of Error Variance was significant (*p* <.05), indicating that the variance across the groups was not equal. As a result, a more stringent significance level, *p* =.01, was set for evaluating the results of the analysis. The interaction effect between gender and TSP estimation conditions was not significant, F(2,210) = .30, *p* = .74, η_p^2 = .00. There was a statistically significant main effect for TSP, F(2,210) = 19.56, *p* =.00, η_p^2 = .16 with large effect size. The main effect for gender was also significant, F(1,210) = 13.26, *p* =.00, η_p^2 = .06, with medium effect size.

Insert Table 5 here

Planned contrasts revealed significant differences between Group 1 and Group 2, (Contrast Estimate -13.68, *p* =.00) and between Group 2 and Group 3 (Contrast Estimate -10.93, *p* =.00). Post-hoc comparisons using the Games-Howell test indicated that the mean score for Group 1 (<= 3) was significantly different from Group 2 (3-4) as well as Group 3 (4+). The mean score for Group 2 was also significantly different from Group 3. Results were confirmed by the Ryan-Einot-Gabriel-Welch Range test of homogenous subsets. Hypothesis 7 was partially confirmed.

For DMIQ2, the interaction effect between gender and TSP was not significant, F(2,203) = .16, *p* = .86, η_p^2 = .00. There was a statistically significant main effect for TSP, F(2,203) = 34.82, *p* =.00, η_p^2 = .26, with large effect size, and for gender, F(1,203) = 11.10, *p* <.01, η_p^2 = .05, with medium effect size. Planned contrasts revealed significant differences between Group 1 and Group 2, (Contrast Estimate -21.46, *p* =.00) and between Group 2 and Group 3 (Contrast Estimate -12.47, *p* =.00). Post-hoc comparisons using the Tukey HSD and Bonferroni tests indicated that the mean score for Group 1 (<=3) was significantly different from Group 2 (3-4) as well as from Group 3 (4+). Group 2 mean scores were also significantly different from Group 3. Results were confirmed by the Ryan-Einot-Gabriel-Welch Range test of homogenous subsets. Hypothesis 8 was partially confirmed.

Insert Figure 1 here

Impact of Gender on the Relationship between TCAP and DMIQ1 and DMIQ2

Two 2-way between-groups analyses of variance were conducted to explore whether gender influences the relationship between TCAP and DMIQ1 and DMIQ2. Results are presented in Table 6. For DMIQ1, the interaction effect between gender and TCAP was significant, $F(2,381) = 3.26$, $p < .05$, $\eta^2 = .02$, with small effect size. The main effect for TCAP, $F(2,381) = 19.56$, $p = .00$, $\eta^2 = .09$, was also significant, with medium effect size. The main effect for gender $F(1,381) = 26.49$, $p = .00$, $\eta^2 = .07$ was also significant, with medium effect size. Planned contrasts revealed significant differences between Group 2 and Group 3, (Contrast Estimate -14.73, *p* =.00).

Post-hoc comparisons using the Tukey HSD and Bonferroni tests indicated that the mean score for Group 1 (<=0) was significantly different from Group 2 (1-8). Group 1 also significantly differed from Group 3 (9+). Group 2 mean scores were also significantly different from Group 3. This was confirmed by the Ryan-Einot-Gabriel-Welch Range test of homogenous subsets.

Insert Table 6 here

As the main interaction effect was significant, further investigation of the relationship was warranted. Simple effects analysis was conducted. The data was split per gender and two one-way between-groups analyses of variance were conducted. For males, the one-way between-groups analysis of variance for DMIQ1 was significant, $F(2,135) = 16.01$, $p = .00$, $\eta^2 = .19$, with large effect size. The robust tests of equality of means, Welch (2, 72) = 12.83, p =.00; Brown-Forsythe (2, 97) = 14.67, p =.00 were also significant. Post-hoc comparisons using the Tukey HSD and Bonferroni tests revealed significant differences in mean scores between Group 1

(<=0) (M = 122.50, SD = 16.05) and Group 2 (1-8) (M = 107.41, SD = 19.70) as well as between Group 2 (1-8) and Group 3 (9+) (M = 126.73, SD = 14.60).

The Levene's Test of Equality of Error Variance was significant (p <.05) in the female sub-sample. As a result, a more stringent significance level, i.e. p =.01, was set for evaluating the results of the analysis. For females, the one-way analysis of variance was also significant, F (2,246) = 5.87, p<.01, η^2 =.05, with medium effect size. The robust tests of equality of means, Welch (2, 160) = 7.55, p <.01; Brown-Forsythe (2,227) = 6.14, p <.01 were significant. The post-hoc comparisons using the Games-Howell test revealed significant differences between Group 1 (<=0) (M = 107.65, SD = 18.70) and Group 3 (9+) (M = 114.69, SD = 13.38) and between Group 3 and Group 2 (1-8) (M = 114.69, SD = 13.54). Hypothesis 9 was confirmed.

For DMIQ2, the interaction effect between gender and TCAP was not significant, F(1,225) = .01, p = .94, η^2 = .00. The main effect for TCAP, F(1,225) = 28.35, p =.00, η^2 = .11 was significant, with medium effect size. The main effect for gender, F(1,225) = 12.99, p =.00, η^2 = .06 was significant with medium effect size. Planned contrasts revealed significant differences between Group 1 and Group 2, (Contrast Estimate -15.18, p =.00). Post-hoc comparisons were not computed as for TCAP only two categories were available, i.e. Group 2 and Group 3 were available. Hypothesis 10 was partially confirmed.

Thus, hypotheses 1, 2, 3, 4 and 9 were confirmed and hypotheses 5 and 6 were not confirmed. Hypotheses 7, 8 and 10 were partially supported.

Discussion

The aim of this study was to confirm the occurrence of HHE on DMIQ1 and DMIQ2. The results confirmed the existence of gender differences on the numerical-spatial factor of SEI (η^2 =.09, d =.66 for DMIQ1 and η^2 =.09, d =.63 for DMIQ2). Equally, a significant decrease in DMIQ estimates was observed from the pre-task to post-task estimation condition (d =.38). The results also revealed significant gender differences in the task-success probes, with males providing higher task-success estimates than females. Yet, males also solved correctly more psychometric problems than did females. The observed effect sizes for both TSP and TCAP were small.

The findings also revealed a stronger relationship between TSP, TCAP and DMIQ2, compared to DMIQ1. This pattern was also observed when the data was split per gender, with TSP and DMIQ2 having stronger relationship than TSP and DMIQ1. Interestingly, for both genders, TCAP only correlated with DMIQ2 and not with DMIQ1. These results indicate that although TSP and TCAP were not assessed during DMIQ1, TSP or task confidence already played a role in the estimation process, indicating the individuals rely on their confidence before they are prompted to do so.

As in previous studies, gender was expected to be the best predictor of DMIQ. The results failed to validate this claim, with TSP confirmed as the best predictor of DMIQ1 and DMIQ2, over and above gender and TCAP, explaining 17% and 23% of variance respectively. Thus, it appears that TSP or task confidence plays an important role in the prediction of the intelligence type.

Subsequently, the role gender plays in the relationship between TSP and DMIQ1 and DMIQ2 was investigated. For DMIQ1, results revealed significant task-success effect, with significant differences between the lowest, average and high task-success groups, with the lowest DMIQ1 estimates provided in the lowest TSP group, average estimates in the average TSP group and the highest DMIQ1 estimates in the highest TSP estimates group. Equally, a significant gender effect revealed that males were more confident than females across the three groups. These results provided further support for the role of confidence in the self-estimation process as well as for male hubris. The results were identical for DMIQ2.

Finally, gender's role in the relationship between TCAP and DMIQ1 and DMIQ2 was examined. For DMIQ1, the results revealed a significant interaction effect as well as significant TCAP and gender effects. Significant differences between the three TCAP groups were observed; with lowest DMIQ1 estimates provided by the group that solved an average number of psychometric problems, average DMIQ1 estimates by the group that did not solve any problems and the highest estimates by the group that solved most psychometric problems. Identical estimation patterns were observed for males and females respectively. These results provided additional support for the role of Better-Than-Average Effect and Worse-Than-Average Effect biases in the self-estimation process (e.g. Alicke et al., 1995;Kruger & Dunning, 1999).

Still, males provided higher DMIQ1 estimates than females in all three groups. Further analyses showed that males'DMIQ1 estimates were significantly different in the lowest and medium TCAP groups as well as between the medium and the highest TCAP groups. Significant differences were also observed for females, with DMIQ1 estimates significantly different in the lowest and highest as well as between medium and highest TCAP groups.

For DMIQ2, the results revealed a significant TCAP effect, with findings identical to the DMIQ1 estimation pattern. Equally, a significant gender effect revealed that males provided higher DMIQ2 estimates than females across the three groups, providing further support for the hubris-humility effect in self-estimated intelligence.

Thus, while gender differences exist in self-estimated intelligence, and in particular in the domain-masculine intelligence type, one's confidence in ability to succeed on a gender stereotype-inducing task, was a better determinant of performance than gender itself. Equally, contrary to some assertions (Ehrlinger & Dunning, 2003; Johnson & Bouchard, 2007; Kruger & Dunning, 1999), the results demonstrated that individuals were capable of making accurate self-estimates that match their confidence levels. Likewise, the existence of the hubris-humility effect, and in particular of the male hubris, was established in the pre- and post- task conditions. As the psychometric task was likely to activate gender-stereotypical biases, it was unsurprising that the provided self-estimates did not match the number of correctly resolved problems, with only the most capable problem solvers providing accurately matching self-estimates, while inflated self-estimates were provided by the average and the least capable problem solvers. Hence, self-confidence seems to positively influence the accuracy of self-estimates, but the psychometric task that evokes cognitive stereotypical biases, seems to impact the accuracy of self-estimates.

Study 2

Introduction

This study set to validate the findings of the previous study. It was identical in set-up and execution, except two numerical problems that yielded no correct answers were dropped. The other measures remained unchanged. In order to further substantiate the previous results, this study ensured that the gender groups were homogeneous in size.

Thus, it is predicted that HHE would prevail on DMIQ1 and DMIQ2 (H1) and that a significant reduction will occur in DMIQ2 (H2). Males were expected to give significantly higher TSP estimations than females (H3). Sex differences are expected to be observed in the psychometric problems, with males providing more correct answers (H4). Further, gender was expected to be the best predictor of the DMIQ1 (H5) and DMIQ2 (H6), over and above TSP and TCAP. Based on previous findings, gender is expected to influence the relationship between TSP and DMIQ1 (H7) and DMIQ2 (H8). Gender was also expected to affect the relationship between TCAP and DMIQ1 (H9) and DMIQ2 (H10). Male and female DMIQ2 estimates were expected to differ in response to TSP probes, while DMIQ1 estimates are controlled for (H11). Equally, males and female DMIQ2 estimates are expected to differ in response to the psychometric problems, while DMIQ1 estimates are controlled for (H12).

Method

Participants

A total of one hundred and eighty-two participants took part in the second experimental online study. There were 92 females (50.5%) and 90 males (49.5%). Their age ranged from 17 to 50 (M = 22.84, SD = 6.51) years. All participants were fluent in English and no language or other problems were reported. 55% had completed A-levels, 21% achieved BA/BSc level, and 10% MA/MSc/MBA or equivalent level of education.

Measures

Repeated Measure of Domain-Masculine Intelligence Type (DMIQ).

See above Alpha for DMIQ1 was .85 and DMIQ2 .88.

Psychometric Aptitude Task Total Correct Aptitude Problems (TCAP)
Numerical and Reasoning Problems (Bryon, 2006)

Thirteen numerical and reasoning problems that were based on actual intelligence test items were presented in three blocks of three and two blocks of two analogous problems (Bryon, 2006). For an overview of the problems see Appendix. Participants were informed that items in each block varied in difficulty level, ranging from elementary to difficult. A time limit of 60 or 90 seconds was given for each block. Participants were advised to leave unanswered problems blank in order to not exceed the time limit, or be disqualified. The time limit was set to reflect a real-life testing situation, with the entire task taking 6.5 minutes. Correct answers were available at the end of the survey. Alpha for the thirteen items was .53.

Task Success Probability Estimation Measure (TSP) (Storek, 2007)

See above The alpha for the five-item measure was .81.

Procedure

Participants were from general public. They were recruited in a same fashion as were participants in the previous study. Data was gathered through an online survey engine and participation was voluntary.

Detailed scoring instructions were given at the beginning of each measure, including timing instructions for the numerical and reasoning problems. Participants were aware that the study was approved by UCL Ethics Committee, meeting confidentiality and Data Protection requirements. Debrief feedback, correct answers and opportunity to leave survey feedback were provided at the end.

Results

HHE and DMIQ1 and DMIQ2

Two independent samples t-tests were computed to assess whether significant gender differences or HHE occurred on DMIQ1 and DMIQ2. Results are presented in Table 7. Significant gender differences, with males providing higher DMIQ estimates in T1 and T2 estimation conditions, were observed. Hypothesis 1 was confirmed.

Insert Table 7 here

To test hypothesis 2 whether significant change occurred from DMIQ1 to DMIQ2 following the intervention task, a paired-samples t-test was conducted. There was a statistically significant decrease in DMIQ1 (M = 112.68, SD = 19.93) to DMIQ2 (M = 106.59, SD = 21.48), t(181) = 7.77, p = .00, two-tailed, r = .87, p =.00. The mean decrease in DMIQ was 6.09 (SD = 10.57) with 95% confidence interval ranging from 4.54 to 7.64. Cohen's d (.58) indicated a medium effect size. Hypothesis 2 was confirmed.

Gender Differences in Task Success Probability Estimation (TSP) and Psychometric Aptitude Task (TCAP)

Table 8 gives an overview of independent-samples t-tests and effect sizes for the five individual task-success probability (TSP) estimation probes and the Total TSP measure. The independent samples t-tests for the five TSP probes and the Total TSP measure were significant, with males providing higher TSP estimates than females. The observed effect sizes were small to medium. Inspection of the correlational results (see Table 9) revealed a medium negative correlation between gender and TSP (r = -.32, p =.00), with males providing higher TSP estimates than females (M_{Males} = 3.24, SD_{Males} = .79; $M_{Females}$ = 2.71, $SD_{Females}$ = .77). Hypothesis 3 was confirmed.

Insert Table 8 here

Inspection of the correlational results (see Table 9) revealed a small negative correlation between gender and TCAP (r =-.26, *p* =.00), with males correctly solving more problems then females (M_{Males} = 9.04, SD_{Males} = 1.87; $M_{Females}$ = 7.95, $SD_{Females}$ = 2.24). 2x2 χ^2 tests[3] and effect sizes for the thirteen psychometric problems were computed. Significant gender differences were observed only on four problems, i.e. Q12A, Q16, Q17 and Q20, with males providing significantly more correct answers than females. This finding differs from the previous study where thirteen problems (87%) revealed significant gender differences. Phi coefficient values, using Cohen's effect size criteria (1988), were small. An independent samples t-test revealed significant gender differences on TCAP, t(180) = 3.60, *p* =. 00 two-tailed, with males (M = 9.04, SD = 1.87) correctly solving more psychometric problems than females (*M* = 7.95, *SD* = 2.24). The magnitude of the differences in the means (mean difference = 1.10, 95% CI: .50 to 1.70) was medium (η^2 = .07; Cohen's *d* =.53). Hypothesis 4 was confirmed.

Gender, TSP and TCAP as Predictors of DMIQ1 and DMIQ2

Firstly, the relationship between the DMIQ1 and DMIQ2, gender, TSP, TCAP and age was explored. Table 9 shows the correlational results. DMIQ1 and DMIQ2 were strongly intercorrelated, which is not surprising (r = .87, *p* =.00). Gender correlated negatively with DMIQ1 (r = -.41, *p* =.00) and DMIQ2 (r = -.50, *p* =.00), with females providing lower scores than males (DMIQ1M_{Males} = 120.94, SD_{Males} = 17.96; DMIQ1$M_{Females}$ = 104.59, $SD_{Females}$ = 18.46; DMIQ2M_{Males} = 117.46, SD_{Males} = 18.10; DMIQ1$M_{Females}$ = 95.96, $SD_{Females}$ = 19.13).

Strong positive correlations were observed between TSP and DMIQ1 (*r* =.50, *p* =.00) and between TSP and DMIQ2 (*r* =.60, *p* =.00). Strong positive correlations were also observed between TCAP and DMIQ1 (*r* =.45, *p* =.00) and between TCAP and DMIQ2 (*r* =.51, *p* =.00). A strong positive relationship was observed between TSP and TCAP (r = .53, *p* =.00). These results are similar to previous study, yet, the correlations between TSP, TCAP and DMIQ1 are even stronger.

---

[3] $\chi^2_{(1)} = Z^2 = r^2$+N. Phi (φ) is the best measure of association for χ^2 test (2x2 contingency table); it estimates the extend of the relationship between the variables. For a 2x2 matrix the following formula is used: $\varphi = \sqrt{\ } \chi^2$ / N, where N is the number of subjects

As in previous studies and given the age range of the participants, i.e. 33 years, age was included in the analysis to explore whether it had an impact on DMIQ. No significant relationships were observed.

Insert Table 9 here

Gender as the best predictor of DMIQ1 and DMIQ2

To determine the best predictor of DMIQ1 and DMIQ2 two simultaneous multiple regressions were performed. The dependent variables were DMIQ1 and DMIQ2 and the independent variables were gender, TSP and TCAP. Results are reported in Table 10. Preliminary analyses were conducted to ensure no violation of the assumptions of normality, linearity, multicollinearity and homoscedasticity.

The first model predicting DMIQ1 was significant $F(3,163) = 30.44$, $p = .00$, Adjusted $R^2 = .35$, $f^2 = .56$), with the overall model explaining 36% of total variance. Gender (β = -.26, p =.00, r_{part} = -.24), TSP (β = .30, p =.00, r_{part} = .25) and TCAP (β = .23, p <.01, r_{part} = .19) were significant predictors of DMIQ1, accounting for 6%, 6% and 4% of variance respectively. As in previous study, TSP was the best predictor of the DMIQ1. Hypothesis 5 was not supported.

The second model, predicting DMIQ2 was also significant $F(3,163) = 55.74$, $p = .00$, Adjusted $R^2 = .50$, $f^2 = 1.04$), with the overall model explaining 51% of total variance. Gender (β = -.32, p =.00, r_{part} = -.30), TSP (β = .38, p =.00, r_{part} = .31) and TCAP (β = .23, p <.01, r_{part} = .19) were significant predictors, explaining 9%, 10% and 4% of variance respectively. As in DMIQ1 and identical to the previous study, TSP was the best predictor of DMIQ2. Hypothesis 6 was also not supported.

Insert Table 10 here

Impact of Gender on the Relationship between TSP on DMIQ1 and DMIQ2

Two 2-way between-groups analyses of variance were conducted to explore whether gender influences the relationship between TSP and DMIQ1 and DMIQ2. Results are presented in Table 11. For DMIQ1, the interaction effect between gender and TSP estimation conditions was not significant, $F(2,161) = 2.39$, $p = .10$, $\eta_p^2 = .03$. There was a statistically significant main effect for TSP, $F(2,161) = 16.12$, $p = .00$, $\eta_p^2 = .17$ with large effect size. The main effect for gender was also significant, $F(1,161) = 13.23$, $p = .00$, $\eta_p^2 = .08$, with medium effect size.

Planned contrasts revealed significant differences between Group 1 and Group 2, (Contrast Estimate -16.21, p =.00) and between Group 2 and Group 3 (Contrast Estimate -9.39, p <.01). Post-hoc comparisons using the Tukey HSD and Bonferroni tests indicated that the mean score for Group 1 (<= 3) was significantly different from Group 2 (3-4) as well as from Group 3 (4+). The mean score for Group 2 was also significantly different from Group 3. Results were confirmed by the Ryan-Einot-Gabriel-Welch Range test of homogenous subsets. Hypothesis 7 was partially confirmed.

Insert Table 11 here

For DMIQ2, the interaction effect between gender and TSP estimation conditions was not significant, $F(2,161) = .40$, $p = .67$, $\eta_p^2 = .01$. There was a statistically significant main effect for TSP, $F(2,161) = 24.53$, $p = .00$, $\eta_p^2 = .23$, and for gender, $F(1,161) = 28.04$, $p = .00$, $\eta_p^2 = .15$, both with large effect sizes. Planned contrasts revealed significant differences between Group 1 and Group 2, (Contrast Estimate -19.93, p =.00) and between Group 2 and Group 3 (Contrast Estimate -11.87, p =.00). Post-hoc comparisons using the Tukey HSD and Bonferroni tests indicated that the mean score for Group 1 (<=3) was significantly different from Group 2 (3-4) as well as from Group 3 (4+). Group 2 mean scores were also significantly different from Group 3. Results were confirmed by the Ryan-Einot-Gabriel-Welch Range test of homogenous subsets. Hypothesis 8 was partially confirmed.

Insert Figure 3 here

Impact of Gender on the Relationship between TCAP and DMIQ1 and DMIQ2

Two 2-way between-groups analyses of variance were conducted to explore whether gender influences the relationship between TCAP and DMIQ1 and DMIQ2. Results are presented in Table 12. For DMIQ1, the interaction effect between gender and TCAP was not significant, $F(2,176) = .29$, $p = .75$, $\eta_p^2 = .00$. The main effect for TCAP, $F(2,176) = 18.77$, $p = .00$, $\eta_p^2 = .17$, was significant, with large effect size. The main effect for gender $F(1,176) = 20.64$, $p = .00$, $\eta_p^2 = .11$ was also significant, with medium effect size. Planned contrasts revealed significant differences between Group 1 and Group 2, (Contrast Estimate -14.75, p =.00). Post-hoc comparisons using the Tukey HSD and Bonferroni tests indicated that the mean score for Group 1 (<=8) was significantly different from Group 2 (8-9) as well as from Group 3 (10+).

This was confirmed by the Ryan-Einot-Gabriel-Welch Range test of homogenous subsets. Hypothesis 9 was partially confirmed.

For DMIQ2, the Levene's Test of Equality of Error Variance was significant (p <.05), indicating the DMIQ2 variance across the groups was not equal. Firstly, the largest and the smallest standard deviations were squared. The largest squared SD was divided by the smallest squared SD, with resulting value of 1.41, which is smaller than the recommended value of 2, suggesting that the group variances were not unacceptably unequal. Equally, a more stringent significance level, p =.01, was set for evaluating the results of the analysis.

Insert Table 12 here

The interaction effect between gender and TCAP was not significant, F (2,176) = .48, p = .62, η_p^2 = .01. The main effect for TCAP, F (2,176) = 20.12, p =.00, η_p^2 = .19 was significant, with large effect size. The main effect for gender, F (1,176) = 39.19, p =.00, η_p^2 = .18 was also significant, with large effect size. Planned contrasts revealed significant differences between Group 1 and Group 2, (Contrast Estimate -15.61, p =.00). Post-hoc comparisons using the Games-Howell test indicated that the mean score for Group 1 (<= 8) was significantly different from Group 2 (8-9) as well as from Group 3 (10+). Results were confirmed by the Ryan-Einot-Gabriel-Welch Range test of homogenous subsets. Hypothesis 10 was partially confirmed.

Insert Figure 4

Gender Differences in DMIQ2 Estimates in Response to TSP

A 2-by-2 between-groups analysis of covariance[4] was conducted to assess the influence of the TSP probes on the DMIQ2 estimates for males and females. The independent variables were TSP and gender. The dependent variable was DMIQ2. DMIQ1 was used as a covariate to control for individual differences. Preliminary checks were conducted to ensure that there was no violation of the assumptions of normality, linearity, homogeneity of variances, homogeneity of regression slopes, and reliable measurement of the covariate.

[4] Analysis of covariance (ANCOVA) is recommended in situations with two-group pre-test/post-test design. The pre-test scores are treated as a covariate to control for pre-existing differences between the groups. Thus, ANCOVA is particularly useful in situations with small sample size and only small or medium effect sizes. (Pallant, 2007, p. 291).

Levene's Test of Equality of Error Variance was significant (p <.05), indicating the groups variances were not equal. An alternative check for comparing variances was used. Firstly, the largest and the smallest standard deviations were squared. The largest squared SD was divided by the smallest squared SD, with resulting value of 1.36, which is smaller than the recommended value of 2, suggesting that the group variances were not unacceptably unequal. Subsequently, a more stringent significance level, p =.01, was set for evaluating the results of the analysis. Homogeneity of regression slopes assumption was not violated, F(2,159) = 1.23, p =.29 for the TSP by DMIQ1 interaction, nor for the gender by DMIQ1 interaction, F(1,159) = .52, p =.47.

After adjusting for DMIQ1 estimates, there was a non-significant interaction effect between TSP and gender, F (2,160) = 1.80, p = .17, η_p^2 = .02. The main effect for TSP was significant, F(2,160) = 6.97, p < .01, η^2 = .08, with medium effect size. The main effect of gender was significant, F(1,160) = 14.94, p =.00, η_p^2 = .09, with medium effect size. The main effect for the covariate variable DMIQ1 was also significant, F(1,160) = 324.31, p =.00, η_p^2 = .67, with the covariate significantly and positively related to DMIQ2 and a large effect size.

Planned comparisons analysis revealed significant differences between Group 2 and Group 1, (Contrast Estimate 4.60, p <.05), between Group 3 and Group 1 (Contrast Estimate 8.75, p =.00) and between the genders (Contrast Estimate 6.56, p =.00). Males provided higher self-estimates of ability (Group 1: M_{Male} = 99.75, SD_{Male} =16.93; M_{Female} = 88.09, SD_{Female} =20.26; Group 2: M_{Male} = 115.46, SD_{Male} =16.60; M_{Female} = 100.48, SD_{Female} =15.63; Group 3: M_{Male} =128.98, SD_{Male} =13.05; M_{Female} = 110.69, SD_{Female} =13.87). The results confirmed that gender, and in particular male hubris plays, as well as task-success probability, a role in DMIQ2. Equally, DMIQ1 contributed to DMIQ2 estimations. Hypothesis 11 was partially confirmed.

7.8. Gender Differences in DMIQ2 in Response to TCAP

A 2-by-2 between-groups analysis of covariance was conducted to assess the influence of TCAP on DMIQ2 estimates for males and females. The independent variables were TCAP and gender. The dependent variable was DMIQ2. DMIQ1 was used as a covariate to control for individual differences. Preliminary checks were conducted to ensure that there was no violation of the assumptions of normality, linearity, homogeneity of variances, homogeneity of regression slopes, and reliable measurement of the covariate.

Homogeneity of regression slopes assumption was not violated for the TCAP by DMIQ1 assumption, F (2,174) = .58, p =.56 nor for the gender by DMIQ1 interaction, F (1,174) = .36, p

=.55. After adjusting for DMIQ1 estimates, there was a non-significant interaction effect between TCAP and gender, F (2,175) = .23, p = .80, η_p^2 = .00. The main effect for TCAP was not significant, F (2,175) = 2.30, p = .10, η_p^2 = .03. The main effect for gender was significant, F(1,175) = 17.20, p =.00, η_p^2 = .09, with medium effect size. The main effect for the covariate variable DMIQ1 was significant, F(1,175) = 330.60, p =.00, η_p^2 = .65, with the covariate significantly and positively related to DMIQ2 and of very large effect size.

Planned comparisons analysis revealed significant differences between Group 3 and Group 1, (Contrast Estimate 4.01, p <.05) and between the genders (Contrast Estimate 6.94, p =.00). Males provided higher self-estimates of ability (Group 1: M_{Male} = 105.30, SD_{Male} =18.66; M_{Female} = 89.66, SD_{Female} =20.17; Group 2: M_{Male} = 117.24, SD_{Male} =15.58; M_{Female} = 102.97, SD_{Female} =16.61; Group 3: M_{Male} =126.31, SD_{Male} =13.79; M_{Female} = 105.83, SD_{Female} =11.02). The results confirmed that gender, and in particular male hubris play a role in DMIQ2 but TCAP did not. Equally, DMIQ1 contributed to DMIQ2 estimations. Hypothesis 12 was partially confirmed.

Thus, hypotheses 1, 2, 3, and 4 were confirmed and hypotheses 5 and 6 were not supported. Hypotheses 7, 8, 9, 10, 11 and 12 were partially supported.

Discussion

This study set out to validate the findings of the previous study. The results confirmed the existence of HHE on DMIQ1 (η^2 =.17, d =1.19 for DMIQ1 and on DMIQ2 (η^2 =.25, d =1.15). Equally, a significant decrease in DMIQ estimates was observed from the pre-task to post-task estimation condition (d =.58). The results also revealed significant gender differences in the task-success probes, with males providing higher task-success estimates than females. Males also correctly solved more psychometric problems than did females. The observed effect sizes for both TSP and TCAP were small to medium. Stronger relationships were also observed between TSP, TCAP and DMIQ2 than between TSP, TCAP and DMIQ1.

As in previous studies, gender was expected to be the best predictor of DMIQ. Results failed to validate this claim, with TSP confirmed as the best predictor of DMIQ1 and DMIQ2, over and above gender and TCAP, explaining 6% and 10% of variance respectively. As in previous study, task confidence plays an important role in the prediction of the intelligence type.

The role that gender plays in the relationship between TSP and DMIQ1 and DMIQ2 was investigated next. For DMIQ1, results revealed a significant task-success effect, with significant differences between the lowest, average and high task-success groups, with the lowest DMIQ1 estimates provided in the lowest TSP group, average estimates in the average TSP group and highest DMIQ1 estimates in the highest TSP estimates group. Equally, a significant gender effect revealed that males were more confident than females across the three groups. These results provide added support for the role of task-confidence in the SEI estimation process and for the display of male hubris in the estimation process. Identical results pattern was observed for DMIQ2.

Subsequently, the role gender plays in the relationship between TCAP and DMIQ1 and DMIQ2 was investigated. For DMIQ1, results revealed a significant TCAP effect, with significant differences between the lowest, average and high TCAP groups, with the lowest DMIQ1 estimates provided by the group that solved fewest TCAP problems, average estimates by the average TCAP group and highest DMIQ1 estimates by the group that solved the most TCAP problems. Equally, significant gender effects revealed that males provided higher DMIQ1 estimates than females across the three groups. These results provide additional support for the assertion that individuals are aware of their abilities and thus capable of accurate self-assessment (e.g. Ackerman et al., 2002; Chamorro-Premuzic et al., 2010; Swim, 1994) of ability as well as for male hubris. An identical result pattern was observed for DMIQ2.

Lastly, two 2x2 between-groups analyses of covariance were conducted to assess whether males and females provided different DMIQ2 estimates in their response to TSP probes as well as the psychometric problems. Both analyses confirmed gender differences in DMIQ2 but not as a result of TSP probes or psychometric problems.

Thus, the results of this study replicated the findings of the previous study in that the existence of the hubris-humility effect was confirmed on the domain-masculine intelligence type in both estimation conditions. Confidence in one's ability to succeed on a psychometric stereotype-inducing task was again the best predictor of the intelligence type. Equally, the results confirmed that the provided self-estimates accurately matched individuals' confidence levels. Contrary to the previous study, the supplied self-estimates were also accurately provided by subjects in all three ability groups, providing further support for the assertion that individuals are capable of accurate self-assessments of ability.

Conclusion

While the causes and working mechanisms of HHE remain to be identified, the following causes have been suggested to play a role: diverse child rearing and socialisation practices (Beloff, 1992), social and gender-role normative stereotyping and self-stereotyping (Guimond et al., 2006), self-enhancement and self-derogatory evaluation biases (Beyer, 1990, 1998, 1999; Furnham, 2001; Kwan et al., 2008), lack of confidence and/or overconfidence (Sleeper & Nigro, 1982), gender differences in self-concept and inaccurate self-estimates (Pallier, 2003; Roberts, 1991), personality traits and male superiority in certain areas of cognition (Chamorro-Premuzic & Furnham, 2005; Hyde et al., 1990a,b; Lynn et al., 2002; Maccoby & Jacklin, 1974; Voyer et al., 1995).

To date, no experimental studies have been conducted in the SEI research programme and only a few SEI studies used 'objective' or psychometric measures to compare the accuracy and validity of SEI estimates (e.g. Batey et al., 2009; Chamorro-Premuzic, Moutafi, & Furnham, 2005; Furnham & Fong, 2000; Furnham & Mottabu, 2004; Holling & Preckel, 2005; Reilly & Mulhern, 1995).

Likewise, the majority of SEI studies were conducted with university students. Participants in the studies reported here were from the general public, making the results more generalisable and robust.

The experimental design allowed for in-depth examination of the role gender plays in the repeated measurement of Domain-Masculine Intelligence Type (DMIQ) as well as in the relationships between the intelligence type and psychometric tasks (TCAP) and the type and task-success confidence probes (TSP). Equally, gender differences in TCAP and TSP were examined in an attempt to understand the conflicting claims in current literature and to clarify whether they have any bearing on the gender differences in the intelligence type.

The repeated measurement of DMIQ aimed to ascertain that HHE can be manipulated or reduced following the psychometric and task-success task, based on the assertions that repeated measures affect mood, confidence and behaviour (Bartsch & Nesselroade, 1973; Ryckman et al., 1971). The results of the two studies confirmed the existence of HHE in the pre- and post-task DMIQ estimates as well as significant reduction in the intelligence type estimates from pre- to post-task estimation condition. The effect sizes for HHE's occurrence on DMIQ1 and DMIQ2 ranged from medium to very large and the effect sizes (NUMBERS) for the DMIQ estimate reduction ranged from small to medium. These results validated the findings of the previous studies (Storek & Furnham, 2012, 2013 ab) as well as provided further support for the role gender plays in HHE and DMIQ.

The gender-stereotype literature has provided abundant evidence for female underperformance on domain-masculine tasks (e.g. Dar-Nimrod, 2007; Ehrlinger & Dunning, 2003; Hyde et al., 1990a,b). The results here established that the psychometric and task-confidence task caused both genders to lower their post-task estimates, although female estimates were lower than males'. These findings are surprising as the existing literature shows that men have higher self-confidence and report higher self-perceived ability on domain-masculine tasks, e.g. mathematics (Meelissen & Luyten, 2008). Thus, the task seems to have affected both genders similarly, impacting on male and female self-perceptions and ability beliefs and causing both genders to reduce their post-task estimates. In other words, the task brought about skill and ability realisation that in turn affected self-perceptions.

Contrary to prediction, gender was not the best predictor of DMIQ1 and DMIQ2 in both studies. Task-success confidence (TSP) was the best predictor with βs ranging from .30 to .54. The role of TSP as the best predictor was unforeseen, and revealed that the task-confidence probes or participants' perceived task-success, had the biggest impact on the post-task estimates. These results provide additional support for the impact of the psychometric and task confidence task, and in particular TSP probes, on the DMIQ estimation pattern by both genders.

Gender differences in math achievement, attitudes and affect have been extensively researched and documented (cf. Halpern et al., 2007), with females displaying more negative or self-handicapping math attitudes, having lower math self-confidence, stereotyping math as domain-masculine, underperforming on standardised math tests, and opting out of STEM careers (Crombie et al., 2005; Beyer, 1990, 1998; Hyde et al., 1990a,b; Linn & Hyde, 1989; Meelissen & Luyten, 2008; Sax & Harper, 2007; The College Board, 1998). On the other hand, males perceive math as a domain-masculine and are more self-confident about their math abilities (Meece et al., 2006; Meelissen & Luyten, 2008; van der Sluis et al., 2010). Thus, males were did better on the psychometric task and were more confident about their success.

To better understand the role gender played in TSP and TCAP in both estimation conditions, a series of analyses of variance were conducted. No interaction affects between TSP and gender were observed in both estimation conditions. Significant gender and TSP effects were observed in all analyses, with males providing higher DMIQ estimates across all three TSP groups.

The accuracy of DMIQ1 and DMIQ2 estimates by the three task-success confidence groups was notable. Overall, males and females provided accurate or matching DMIQ1 and DMIQ2 estimates, i.e. low DMIQ estimates by low task-success probability group, average estimates by average group and high DMIQ estimates by high task-success probability group.

Significant interaction effects between TCAP and gender were observed for DMIQ1 but not for DMIQ2. Significant gender and TCAP effects were observed in all analyses, with males across all three TCAP groups providing higher DMIQ estimates than females.

The accuracy of DMIQ1 and DMIQ2 estimates by the three TCAP groups differed from TSP results. Overall, the estimates were less accurate.

As the TCAP and TSP tasks were devised to also validate the claims that individuals overestimate their ability on easy tasks and underestimate their abilities on difficult tasks (e.g. Alicke et al., 1995; Burson et al., 2006; Guenther & Alicke, 2010; Moore & Small, 2007), leading them to make inaccurate performance judgements (Ehrlinger & Dunning, 2003; Kahneman & Tversky, 2000), the observed results are particularly interesting.

Based on the observed data, individuals were capable of more accurate intelligence estimates in the task-success probability conditions then in the psychometric conditions. In particular, the TSP results support the assertions that individuals are capable of accurate self-assessments of ability (e.g. Ackerman et al., 2002; Chamorro-Premuzic et al., 2010; Hall & Carter, 1999; Swim, 1994), but not in the psychometric task condition. Equally, the observed male hubris in DMIQ estimates, provided support for the literature in the field. Thus, gender influenced the relationship between confidence and the Domain-Masculine Intelligence Type as well as between the psychometric tasks and the intelligence type.

Limitations and Future Research

The main limitation of this research was the fact that the intelligence type was assessed through a single estimate that could have been influenced by numerous factors, such as mood fluctuation, fatigue, fear, lack of concentration, socially desirable responding, and stress, at the time of estimation. As such it is possible that the acquired estimates were not only subjective but also unreliable. Still, DMIQ is an individualised score based on a combination of two scores, the mathematical/logical and spatial estimates. Similarly, numerous studies about the accuracy of 'subjective' assessments have shown that individuals are capable of accurate self-assessments of ability and that the current SEI measures are valid proxies of intellectual competence (Ackerman et al, 2002; Chamorro-Premuzic et al., 2010; Swim, 1994). Equally, the introduction of multiple measurements of DMIQ estimates was intended to reduce the possible affects of 'subjective' measurement. The experimental findings replicated the earlier correlation results (e.g.: Storek & Furnham, 2012, 2013ab), providing further support for the observed results.

Based on the findings that largely affirmed the main objectives, the main recommendation for future research is the employment of more sophisticated statistical analyses, such as SEM that allow for in-depth and simultaneous examination of multiple causal relationships and assumptions. Recent studies have demonstrated that the usage of sophisticated techniques and models, such as SEM yield more reliable data as well as exposed faulty assumptions that were made using traditional statistical techniques (e.g.: Chamorro-Premuzic et al., 2010; van der Sluis, 2010; von Stumm et al., 2009).

Likewise, studies with diverse and large study samples, preferably international, are recommended in order to produce more robust and generalisable results. Equally, asking male and female participants whether their perceive the individual self-estimated intelligences as masculine or feminine could help the understanding of the self-perception and gender-stereotypical biases that were shown to play a role in the observed gender differences in DMIQ.

References

Ackerman, P. L., Beier, M. E., & Bowen, K. R. (2002). What we really know about our abiltieis and our knowledge. *Personality and Individual Differences, 33,* 587-605.

Alicke, M. D., Klotz, M. L., Breitenbecher, D. L., Yurak, T. J., & Vredenburg, D. S. (1995). Personal contact, individuation and the Better-Than-Average effect. *Journal of Personality and Social Psychology, 68*, 804-825.

Bartsch, T. W., & Nesselroade, J. R. (1973). Test of the trait-state anxiety distinction using a manipulative, factor-analytic design. *Journal of Personality and Social Psychology, 27,* 58-64.

Batey, M., Chamorro-Premuzic, T., & Furnham, A., (2009). Intelligence and personality as predictors of divergent thinking: The role of general, fluid and crystallised intelligence. *Thinking Skills and Creativity, 4,* 60-69.

Beier, M. E., & Ackerman, P. L. (2001). Current events knowledge in adults: An investigation of age, intelligence, and non-ability determinants. *Psychology and Aging, 16,* 615-628.

Beier, M. E., & Ackerman, P. L. (2003). Determinants of health knowledge: An investigation of age, gender, abilities, personality, and interests. *Journal of Personality and Social Psychology, 84* , 439-448.

Beloff, H. (1992). Mother, father and me: our intelligence. *The Psychologist, 5,* 309-311.

Betsworth, D. (1999). Accuracy of self-estimated abilities and the relationship between self-estimated abilities and realism for women. *Journal of Career Assessment, 7,* 35-43.

Beyer, S. (1990). Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology, 59*, 960-970.

Beyer, S. (1998). Gender differences in self-perception and negative recall bias. *Sex Roles, 38*, 103-133.

Beyer, S. (1999). Gender differences in the accuracy of grade expectations and evaluations. *Sex Roles, 41*, 279-296.

Bryon, M. (2006). *The Ultimate Psychometric Test Book*. London: Kogan Page Ltd

Burson, K. A., Larrick, R. P., Klayman, J. (2006). Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive miscalibration in relative comparisons. *Journal of Personality and Social Psychology, 90*, 60-77.

Carr, M., Hettinger-Steiner, H., Kyser, B., & Biddlecomb, B. (2008). A comparison of predictors of early emerging gender differences in mathematics competency. *Learning and Individual Differences, 18*, 61-75.

Chamorro-Premuzic, T., Harlaar, N., Greven, C. U., & Plomin, R. (2010).
More than just IQ: A longitudinal examination of self-perceived abilities as predictors of academic performance in a large sample of UK twins. *Intelligence, 38,* 385-392.

Chamorro-Premuzic, T., & Furnham, A. (2005). *Personality and intellectual competence.* New Jersey: Laurence Erlbaum Associates.

Chamorro-Premuzic, T., Moutafi, J., & Furnham, A. (2005). The relationship between personality traits, subjectively-assessed and fluid intelligence. *Personality and Individual Differences, 38,* 1517-1528.

Cohen, J. L. (1988). *Statistical power analysis for the behavioural sciences (2nd Ed.)*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Crombie, G., Sinclair, N., Silverthorn, N., Byrne, B. M., DuBois, D. L., & Trinneer, A. (2005). Predictors of young adolescents' math grades and course enrolment intentions: Gender similarities and differences. *Sex Roles, 52,* 351-367.

Dar-Nimrod, I. (2007). Math ability in women- nature versus nurture. *Nano Today, 2,* 56.

Deary, I. J., Thorpe, G., Wilson, V., Starr, J. M., & Whalley, L. J. (2003). Population sex differences in IQ at age 11: The Scottish Mental Survey, 1932. *Intelligence, 31*,533-542.

Dunning, D., Griffin, D. W., Milojkovic, J. D., & Ross, L. (1990). The overconfidence effect in social prediction. *Journal of Personality and Social Psychology, 58,* 568-581.

Eccles-Parsons, J. E., Adler, T., & Meece J. L. (1984). Sex differences in achievement: A test of alternate theories. *Journal of Personality and Social Psychology, 46*, 26-43.

Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology, 84*, 5-17.

Furnham, A. (2000). Parents' estimates of their own and their children's multiple intelligences. *British Journal of Developmental Psychology, 18*, 583-594.

Furnham, A. (2001). Self-estimates of intelligence: culture and gender difference in self and other estimates of both general (*g*) and multiple intelligences. *Personality and Individual Differences, 31*, 1381-1405.

Furnham, A., & Budhani, S. (2002). Sex differences in the estimated intelligence of school children. *European Journal of Personality, 16,* 201-219.

Furnham, A., Crawshaw, J., & Rawles, R. (2006). Sex differences in self-estimates on two validated IQ test subscale scores. *Journal of Applied Social Psychology, 26,* 417-440.

Furnham, A., & Fong, G. (2000). Self-estimated and psychometrically measured intelligence: A cross-cultural and sex differences study of British and Singaporean students. *North American Journal of Psychology, 2,* 191-200.

Furnham, A. & Gasson, L. (1998). Sex differences in parental estimates of their children's intelligence. *Sex Roles, 38*, 151-162.

Furnham, A., Hosoe, T. & Tang, T. (2003). Male hubris and female humility? A cross-cultural study of ratings of self, parental and sibling multiple intelligence in America, Britain, and Japan. *Intelligence, 30,* 101-115.

Furnham, A., & Mottabu, R. (2004). Sex and culture differences in the estimates of general and multiple intelligence: A study comparing British and Egyptian students. *Individual Differences Research, 2,* 82-94.

Furnham, A., & Shagabutdinova, K. (2012). Sex differences in estimating multiple intelligences in self and others: a replication in Russia. *International Journal of Psychology, 47,* 448-459.

Guenther, C. L., & Alicke, M. D. (2010). Deconstructing the Better-Than-Average effect. *Journal of Personality and Social Psychology, 99*, 755-770.

Guimond, S., Martinot, D., Chatard, A., Crisp, R. J., & Redersdorff, S (2006). Social comparison, self-stereotyping, and gender differences in self-construals. *Journal of Personality and Social Psychology*, *90*, 221-242.

Hall, J. A., & Carter, J. D. (1999). Gender-stereotype accuracy as an individual difference. *Journal of Personality and Social Psychology*, *77,* 350-359.

Halpern, D.F., Benbow, C. P., Geary, D.C., Gur, R. C., Hyde, J.S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, *8*, 1-51.

Hoffman, C., & Hurst, N. (1990). Gender stereotypes: Perception or rationalization? *Journal of Personality and Social Psychology, 58*, 197-208.

Holling, H., & Preckel, F. (2005). Self-estimates of intelligence – methodological approaches and gender differences. *Personality and Individual Differences, 38*, 503-517.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990a). Gender differences in mathematics performance : A meta-analysis. *Psychological Bulletin, 107,* 139-155.

Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., & Hopp, C. (1990b). Gender comparisons of mathematics attitudes and affect. A meta-analysis. *Psychology of Women Quarterly, 14,* 299-324.

Jackson, D. N., & Rushton, J. P. (2006). Males have greater *g*: sex differences in general mental ability from 100,000 17- to 18-year olds on the Scholastic Assessment Test. *Intelligence*, *34*, 479-486.

Johnson, W., & Bouchard, T. J., Jr. (2007). Sex differences in mental abilities: *g* masks the dimensions on which they lie. *Intelligence, 35*, 23-39.

Kahneman, D., & Tversky, A. (2000*). Choices, values, and frames*. Cambridge: Cambridge University Press.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognising one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*, 1124-1134.

Kwan, V.S., John, O., Robin, R., & Kuang, L. (2008). Conceptualizing and assessing self-enhancement bias: A componential approach. *Journal of Personality and Social Psychology, 94*, 1062-1077.

Linn, M. C., & Hyde, J. (1989). Gender, mathematics, and science. *Educational Researcher, 18,* 17-27.

Lynn, R., & Irwing, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence, 32*, 481-498.

Lynn, R., Irwing, P., & Cammock, T. (2002). Sex differences in general knowledge. *Intelligence, 30,* 27-39.

Maccoby, E.E., & Jacklin, C. N. (1974). *The psychology of sex differences.* Stanford: CA: Stanford University Press.

Meece, J. L., Glienke, B. B., & Burg, S. (2006). Gender and motivation. *Journal of School Psychology, 44*, 351-373.

Meelissen, M., & Luyten, H. (2008). The Dutch gender gap in mathematics: Small for achievement, substantial for beliefs and attitudes. *Studies in Educational Evaluation, 34,* 82-93.

Moore, D. A., & Small, D. A. (2007). Error and bias in comparative judgment: On being both better and worse than we think we are. *Journal of Personality and Social Psychology, 92,* 972-989.

Novell, A., & Hedges, L.V.(1998). Trends in gender differences in academic achievement from 1960-1994: An analysis of differences in mean, variance and extreme scores. *Sex Roles, 39,* 21-43.

Ogle, L., Sen, A., Pahlke, E., Jocelyn, L., Kostberg, D., Roye, S., & Williams, T. (2003). *International comparisons in fourth-grade reading literacy: Findings from the Progress in international Literacy Study (PIRLS) of 2001* (NCES 2003-073). Washington, DC: U.S. Government Printing Office.

Pallant, J. (2007). *SPSS Survival Manual: A step by step guide to data analysis using SPSS for Windows* (3rd edition). New York: Open University Press.

Pallier, G. (2003). Gender differences in the self-assessment of accuracy on cognitive tasks. *Sex Roles, 48,* 265-276.

Rammstedt, B., & Rammsayer, T. H. (2002a). Gender differences in self-estimated intelligence and their relation to gender-role orientation. *European Journal of Personality, 16,* 382-396.

Rammstedt, B., & Rammsayer, T. H. (2002b). Self-estimated intelligence. Gender differences, relationship to psychometric intelligence and moderating effects of level of education. *European Psychologist, 7,* 275-284.

Reilly, J., & Mulhern, G. (1995). Gender differences in self-estimated IQ: The need for care in interpreting group data. *Journal of Personality, 66,* 523-555.

Roberts, T.-A. (1991). Gender and the influence of evaluations on self-assessments in achievement settings. *Psychological Bulletin, 109,* 297-308.

Ryckman, R. M., Gold, J. A., & Rodda, W. C. (1971). Confidence rating shifts and performance as a function of locus of control, self-esteem, and initial task performance. *Journal of Personality and Social Psychology, 18,* 305-310.

Sax, L. J., & Harper, C. E. (2007). Origins of the gender gap: Pre-college and college influences on differences between men and women. *Research in Higher Education, 48,* 669-694.

Sleeper, L. A., & Nigro, G. N. (1982). It's not who you are but who you're with: Self-confidence in achievement settings. *Sex Roles, 16,* 57-69.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69,* 797-811.

Stieger, S., Kastner, C., Voracek, M., von Stumm, S., Chamorro-Premuzic, T., & Furnham, A. (2010). Independent effects of personality and sex on self-estimated intelligence: Evidence from Austria. *Psychological Reports, 107,* 553-563.

Storek, J. (2007). *Task Success Probability Estimation Measure.* Unpublished measure, University College London, UK.

Storek, J., & Furnham, A. (2012). Gender and gender role differences in Domain-Masculine Intelligence and beliefs about intelligence: A study with Mensa UK members. *Personality and Individual Differences, 53,* 890–895.

Storek, J., & Furnham, A. (2013a).Gender, 'g', and fixed versus growth intelligence mindsets as predictors of self-estimated Domain Masculine Intelligence (DMIQ). *Learning and Individual Differences, 25,* 93–99

Storek, J., & Furnham, A. (2013b). Gender, *g*, gender identity concepts, and self-constructs as predictors of the self-estimated IQ. *The Journal of Genetic Psychology, 10,* 1-13

Storek, J., & Furnham, A. (2014). Gender and Task Confidence as Predictors of the Domain-Masculine Intelligence Type (DMIQ). *Personality and Individual Differences, 69,* 43-49.

Swim, J. K. (1994). Perceived versus meta-analytic effect sizes: An assessment of the accuracy of gender stereotypes. *Journal of Personality and Social Psychology, 66,* 21-36.

Szymanowicz, A., & Furnham, A. (2011). Gender differences in self-estimates of general, mathematical, spatial, and verbal intelligence: Four meta-analyses. *Learning and Individual Differences, 21*, 493–504.

The College Board, Office of Research and Development (1998). *Research Summary: SAT and Gender Differences*. NY: New York.

Van der Sluis, S., Vinkhuyzen, A. A. E., Boomsma, D. I., & Posthuma, D. (2010). Sex differences in adults' motivation to achieve. *Intelligence, 38,* 433-446.

von Stumm, S., Chamorro-Premuzic, T., & Furnham, A. (2009). Decomposing self-estimates of intelligence: Structure and sex differences across 12 nations. *British Journal of Psychology, 100*, 429-442.

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,* 250-270.

Table 1

Independent t-tests and Effect Sizes for Task-Success Probability Estimation and 5 Individual TSP Probes

	Males M (SD) n	Females M (SD) n	*t(df)*	Mean Difference	95% CI		Effect Size	
					L	U	η^2	*d*
Total TSP	3.18 (.80) 90	2.88 (.81) 132	2.75(220)**	.30	.09	.52	.03	.37
TSP1	3.61 (1.09) 99	3.32 (1.04) 154	2.11(251)*	.29	.02	.56	.02	.27
TSP2	2.81 (1.04) 110	2.54 (1.04) 150	2.01(248)*	.27	.01	.54	.02	.48
TSP 3	3.43 (1.02) 98	2.97 (1.10) 143	3.27(237)**	.46	.18	.73	.04	.43
TSP 4	3.40 (.91) 99	3.20 (1.09) 143	1.51(240)	.20	-.06	.46	.01	.20
TSP 5	2.67 (1.15) 96	2.31 (1.13) 140	2.38(234)*	.36	.06	.66	.02	.31

* $p < .05$ ** $p < .01$ *** $p < .001$ (2-tailed). Note: *d* = Hedge's Adjustment or Cohen's d adjusted for sample size.

Table 2

Correlations and Partial Correlations, Means and Standard Deviations between DMIQ1 and DMIQ2, Gender, TSP, TCAP and Age

	DMIQ1	DMIQ2	G	TSP	TCAP	A
	112.86	108.43	1.66	3.00	4.34	22.33
	(19.37)	(21.20)	(.47)	(.82)	(4.45)	(6.86)
DMIQ1						
DMIQ2	.78***					
Gender	-.30***	-.30***				
TSP	.47***	.62***	-.18**			
TCAP	.16**	.40***	-.18***	.43***		
Age	.08	.01	-.14**	-.06	.12*	
-Controlled For Age-						
DMIQ1						
DMIQ2	.78***					
Gender	-.29***	-.30***				
TSP	.48***	.63***	-.19**			
TCAP	.15**	.40***	-.17**	.44***		

* $p < .05$ ** $p < .01$ *** $p < .001$ (2-tailed). N between 198 and 487.

Table 3

Correlations, Means and Standard Deviations between DMIQ1 and DMIQ2, TSP, TCAP and Age – Per Gender

	Males		Females	
	DMIQ1	DMIQ2	DMIQ1	DMIQ2
	120.64	116.02	108.55	102.57
	(18.13)	(21.58)	(18.70)	(21.14)
DMIQ1				
DMIQ2	.64***		.83***	
TSP	.49***	.65***	.41***	.57***
TCAP	.14	.44***	.10	.31***
Age	.01	.08	.07	-.07

* $p < .05$ ** $p < .01$ *** $p < .001$ (2-tailed). N between 47 and 321.

Table 4

Beta coefficients for Simultaneous Multiple Regressions of Gender, TSP and TCAP onto DMIQ1 and DMIQ2

Dependent Variable	***DMIQ1***		***DMIQT2***	
	β	*t*	*β*	*t*
Gender	**-.23**	**-3.83*****	**-.18**	**-3.26****
TSP	**.46**	**7.07*****	**.54**	**9.17*****
TCAP	-.08	-1.20	**.14**	**2.34***
Regression Model	**F(3, 212) = 26.48*****		**F(3, 205) = 53.43*****	
R^2	.27		.44	
R^2 Change	.27		.44	
Adj. R^2	.26		.43	
f^2	.37		.79	

p <.05, ** *p* <.01, *** *p* <.001. Note: Significant values are in bold.

Table 5

2-way ANOVA (TSP and gender) on DMIQ1 and DMIQ2

Variable	TSP Groups	Mean Score (SD)			*F*-score		
		Total	Males	Females	TSP	Gender	TSP x Gender
DMIQ1	G1 (L)	104.43 (20.17)	111.21 (23.80)	100.98 (17.28)	19.56***	13.26***	.30
	G2 (M)	113.76 (16.17)	117.47 (16.23)	111.15 (15.78)			
	G3 (H)	125.33 (15.69)	130.34 (12.75)	120.13 (16.95)			
DMIQ2	G1 (L)	94.56 (23.04)	101.38 (27.69)	91.33 (19.97)	34.82***	11.10**	.16
	G2 (M)	111.01 (15.90)	115.02 (15.55)	108.14 (15.71)			
	G3 (H)	124.04 (16.24)	128.98 (13.05)	119.11 (17.78)			

* *p* < .05 ** *p* < .01 *** *p* < .001 (2-tailed). Note: DMIQ1 = Domain-Masculine Intelligence Type at pre-task estimation condition; DMIQ2 = Domain-Masculine Intelligence Type at post-task estimation condition. TSP = Task-success probability estimation condition.

Figure 1 Two 2-way ANOVAs (TSP and Gender) on DMIQ1 and DIMQ2

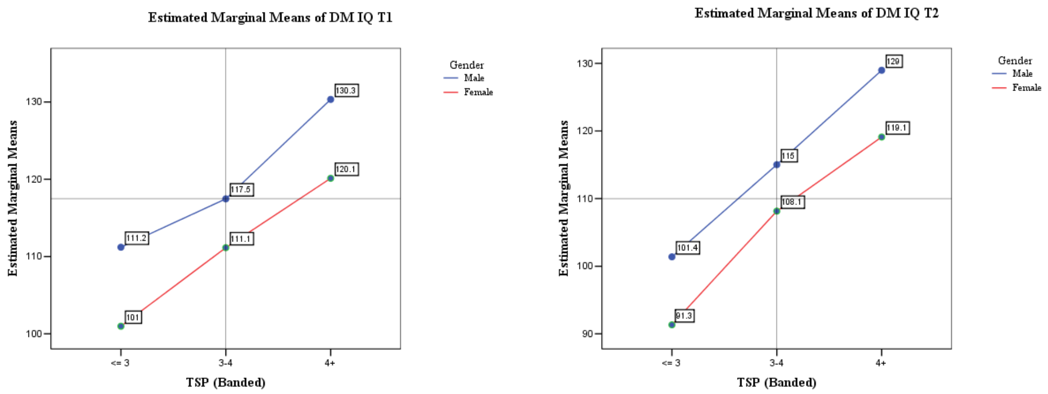

Table 6

Two 2-way ANOVAs (TCAP and gender) on DMIQ1 and DMIQ2

Variable	TCAP Groups	Mean Score (SD)			*F*-score		
		Total	Males	Females	TCAP	Gender	TCAP x Gender
DMIQ1	G1 (L)	111.84	122.50	107.65	19.56***	26.49***	3.26*
		(19.15)	(16.04)	(18.70)			
	G2 (M)	105.41	107.41	104.55			
		(20.68)	(19.70)	(21.15)			
	G3 (H)	120.53	126.73	114.69			
		(15.26)	(14.60)	(13.54)			
DMIQ2	G1 (L)	98.12	105.30	95.25	28.35***	12.99***	.01
		(22.44)	(18.66)	(23.29)			
	G2 (M)	115.34	120.71	110.21			
		(19.15)	(21.21)	(15.42)			
	G3 (H)	107.97	116.02	102.57			
		(22.27)	(21.58)	(21.14)			

* $p < .05$ ** $p < .01$ *** $p < .001$ (2-tailed). Note: DMIQ1 = Domain-Masculine Intelligence Type at pre-task estimation condition; DMIQ2 = Domain-Masculine Intelligence Type at post-task estimation condition. TCAP = Total correct aptitude problems.

Figure 2 Two 2-way ANOVAs (TCAP and Gender) on DMIQ1 and DMIQ2

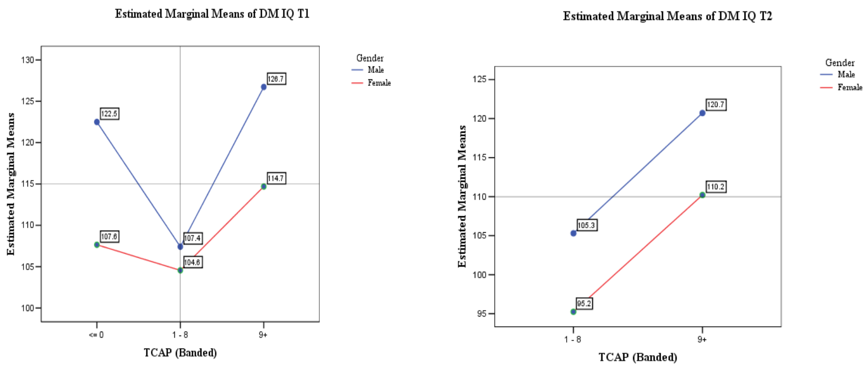

Table 7
Overview of Independent t-Tests and Effect Sizes for DMIQ1 and DMIQ2

	Males M (SD) n	Females M (SD) n	*t(df)*	Mean Diff.	95% CI L	95% CI U	Effect Size η^2	Effect Size *d*
DMIQ1	120.94 (6.06) 90	104.59 (18.46) 92	6.06(180)***	16.35	11.02	21.68	**.17**	**1.19**
DMIQ2	117.46 (18.10) 90	95.96 (19.13) 92	7.78(180)***	21.50	16.05	26.95	**.25**	**1.15**

* *p* < .05 ** *p* < .01 *** *p* < .001 (2-tailed). Note: *d* = Cohen's *d*. Large effect sizes are in bold

Table 8

Independent t-tests and Effect Sizes for Task-Success Probability Estimation and 5 Individual TSP Probes

	Males M (SD) n	Females M (SD) n	*t(df)*	Mean Difference	95% CI L	95% CI U	Effect Size η^2	Effect Size *d*
Total TSP	3.24 (.79) 82	2.71 (.77) 85	4.39(164)***	.53	.29	.77	.10	.68
TSP1	3.69 (1.03) 88	3.20 (1.05) 91	3.19(177)**	.50	.19	80	.05	.47
TSP2	2.82 (1.86) 88	2.36 (1.01) 89	2.95(175)**	.50	.15	.77	.05	.31
TSP 3	3.48 (1.02) 88	2.79 (1.12) 89	4.29(175)***	.69	.37	1.01	.10	.64
TSP 4	3.44 (.91) 90	3.13 (1.06) 89	2.10(177)*	.31	.02	.60	.02	.31
TSP 5	2.72 (1.16) 88	2.09 (1.08) 88	3.69(174)***	.63	.29	.96	.07	.56

* $p < .05$ ** $p < .01$ *** $p < .001$ (2-tailed). Note: *d* = Hedge's Adjustment or Cohen's d adjusted for sample size.

Table 9

Correlations, Means and Standard Deviations between DMIQ1, DMIQ2, Gender, TSP, TCAP and Age

	DMIQ1	DMIQ2	G	TSP	TCAP	A
	112.68	106.59	1.51	2.97	8.49	22.84
	(19.93)	(21.48)	(.50)	(.82)	(2.13)	(6.51)
DMIQ1						
DMIQ2	.87***					
Gender	-.41***	-.50***				
TSP	.50***	.60***	-.32***			
TCAP	.45***	.51***	-.26***	.53***		
Age	.05	-.02	-.11	-.10	.10	

* *p* < .05 ** *p* < .01 *** *p* < .001 (2-tailed). N = between 167 and 182.

Table 10

Beta coefficients for Simultaneous Multiple Regressions of Gender, TSP and TCAP onto DMIQ1 and DMIQ2

Dependent Variable	***DMIQ1***		***DMIQT2***	
	β	*t*	*β*	*t*
Gender	**-.26**	**-3.83*** **	**-.32**	**-5.47*** **
TSP	**.30**	**3.98*** **	**.38**	**5.68*** **
TCAP	**.23**	**3.05** **	**.23**	**3.53** **
Regression Model	**F(3, 163) = 30.44*** **		**F(3, 163) = 55.74*** **	
R^2	.36		.51	
R^2 Change	.36		.51	
Adj. R^2	.35		.50	
f^2	.56		1.04	

p <.05 ** *p* <.01 *** *p* <.001. *Note:* Significant values are in bold.

Table 11

Two 2-way ANOVAs (TSP and gender) on DMIQ1 and DMIQ2

Variable	TSP Groups	Mean Score (SD)			*F*-score		
		Total	Males	Females	TSP	Gender	TSP x Gender
DMIQ1	G1 (L)	101.62 (19.65)	102.75 (20.49)	101.16 (19.55)	16.12***	13.23***	2.39
	G2 (M)	114.08 (17.09)	120.58 (15.84)	106.36 (15.40)			
	G3 (H)	125.60 (14.86)	130.34 (12.75)	115.38 (14.30)			
DMIQ2	G1 (L)	91.42 (19.94)	99.75 (16.93)	88.09 (20.26)	24.53***	28.04***	.40
	G2 (M)	108.61 (17.72)	115.46 (16.60)	100.48 (15.63)			
	G3 (H)	123.18 (15.71)	128.98 (13.05)	110.69 (13.87)			

* $p < .05$ ** $p < .01$ *** $p < .001$ (2-tailed). Note: DMIQ1 = Domain-Masculine Intelligence Type at pre-task estimation condition; DMIQ2 = Domain-Masculine Intelligence Type at post-task estimation condition. TSP = Task-success probability estimation condition.

Figure 3 Two 2-way ANOVAs (TSP and Gender) on DMIQ1 and DMIQ2

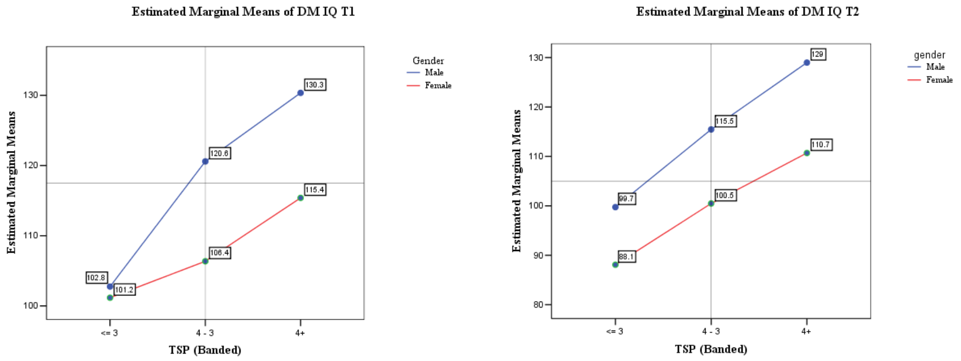

Table 12

Two 2-way ANOVAs (TCAP and gender) on DMIQ1 and DMIQ2

Variable	TCAP Groups	Mean Score (SD)			*F*-score		
		Total	Males	Females	TCAP	Gender	TCAP x Gender
DMIQ1	G1 (L)	102.39 (20.26)	109.34 (19.70)	98.72 (19.75)	18.77***	20.64***	.29
	G2 (M)	116.40 (15.92)	121.07 (14.39)	110.44 (16.17)			
	G3 (H)	124.03 (14.29)	129.21 (13.80)	114.40 (9.55)			
DMIQ2	G1 (L)	95.07 (20.93)	105.30 (18.66)	89.66 (20.17)	20.12***	39.19***	.48
	G2 (M)	110.98 (17.39)	117.24 (15.58)	102.97 (16.61)			
	G3 (H)	119.14 (16.14)	126.31 (13.79)	105.83 (11.02)			

* $p < .05$ ** $p < .01$ *** $p < .001$ (2-tailed).

Note: DMIQ1 = Domain-Masculine Intelligence Type at pre-task estimation condition; DMIQ2 = Domain-Masculine Intelligence Type at post-task estimation condition. TCAP = Total Correct Aptitude Problems.

Figure 4 Two 2-way ANOVAs (TCAP and Gender) on DMIQ1 and DMIQ2