Personality and Individual Differences 69 (2014) 43–49

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

Gender and task confidence as predictors of the Domain-Masculine Intelligence Type (DMIQ)

Josephine Storek [a], Adrian Furnham [a,b,*]

[a] Research Department of Clinical, Educational and Health Psychology, University College London, United Kingdom
[b] Norwegian Business School, Oslo, Norway

ARTICLE INFO

Article history:
Received 24 August 2013
Received in revised form 14 April 2014
Accepted 8 May 2014
Available online 3 June 2014

Keywords:
Gender
Task
Self-estimated intelligence

ABSTRACT

This study investigated the relationship between gender, ability tests and task confidence as predictors of self-estimated intelligence. Participants estimated their mathematical and spatial intelligence before and after completing various ability tests which they also rated for their probable success. Males rated their intelligence higher than females on both before and after measures, as well on their task estimated probably success. There was no sex difference on the ability test. Gender and actual task success scores were predictors of self-estimates, with the former being twice as powerful as the latter. The results are discussed in terms of gender differences in confidence and expectations when approaching ability tests and their effect on self-estimated intelligence.

1. Introduction

There is overwhelming evidence that demonstrates stable and universally consistent gender differences in self-estimated intelligence (SEI) (Furnham & Shagabutdinova, 2012; von Stumm, Chamorro-Premuzic, & Furnham, 2009), also known as the '*hubris-humility effect*' (HHE) (Beloff, 1992; Storek & Furnham, 2012; Storek & Furnham, 2013) This study explores mechanisms that may influence the self-estimation process.

While the determinants of HHE are complex, the following causes have been suggested to play a role: diverse child rearing and socialisation practices; social and gender-role normative stereotyping and self-stereotyping (Guimond, Martinot, Chatard, Crisp, & Redersdorff, 2006), self-enhancement and self-derogatory evaluation biases (Beyer, 1999; Furnham, 2001; Kwan, John, Robin, & Kuang, 2008), lack of confidence and/or overconfidence (Sleeper & Nigro, 1987), gender differences in self-concept and inaccurate self-estimates (Pallier, 2003), personality traits and male superiority in certain areas of cognition (Chamorro-Premuzic & Furnham, 2005)

It is possible that '*humility*' is an erroneous 'label' for female ability to provide more accurate self-estimates of ability than males (Rammstedt & Rammsayer, 2002). It seems males are *overconfident* about their math performance, whilst females report low math confidence (Carr, Hettinger-Steiner, Kyser, & Biddlecomb, 2008). This study examined gender difference in test scores as well as self-assessments both before and after taking ability tests.

A meta-analytical study (Szymanowicz, & Furnham, 2011) assessing the extent of gender differences in mathematical/logical, spatial, overall and verbal SEI demonstrated, 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. 'Domain-Masculine Intelligence Type' (DMIQ) which is a composite of mathematical/logical and spatial intelligences, has been shown as the best predictor of HHE and gender differences in SEI (Storek & Furnham, 2012; Storek & Furnham, 2013).

This study was designed to establish the determinants of gender differences in self-estimates as well as the existence of the HHE effect by introducing a number of timed psychometric tasks (TCAP) and confidence assessments (TSP). Few studies have used psychometric measures to assess validity and accuracy of SEI (Batey, Chamorro-Premuzic, & Furnham, 2009; Chamorro-Premuzic, Moutafi, & Furnham, 2005; Holling & Preckel, 2005) but none included task-confidence, as well as before and after assessments.

Participants in this study were asked to undertake a gender-stereotype inducing task, i.e. numerical, reasoning and crystallised intelligence problems that were likely to increase hubris and humility (Betsworth, 1999; Dar-Nimrod, 2007; Ehrlinger & Dunning, 2003). Task-confidence probes (TSP) were introduced to

---

* Corresponding author at: Research Department of Clinical, Educational and Health Psychology, University College London, United Kingdom. Tel.: +44 2076795395.

E-mail address: a.furnham@ucl.ac.uk (A. Furnham).

http://dx.doi.org/10.1016/j.paid.2014.05.006

clarify the conflicting claims in current literature and investigate whether confidence plays a role in gender differences in self-assessments (Carr et al., 2008; Pallier, 2003). After each block of psychometric problems, the participants were asked to estimate their task-success confidence. Repeated measures were included to validate assertions that they influence behaviour and performance and as such change mood and confidence (Bartsch & Nesselroade, 1973).

It was predicted that HHE will be confirmed on DMIQ at the pre-task (*T*1) and post-task (*T*2) estimating conditions and that there will be a significant decrease in DMIQ estimates from *T*1 to *T*2 following the gender-stereotype inducing task (H1).

The 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 (Pallier, 2003). Consequently, males are expected to provide significantly higher task-success probability estimations than females (H2).

Given the ample evidence about sex differences in cognitive abilities (Halpern et al., 2007), sex differences are expected on the numerical and reasoning problems with males providing more correct answers than females (H3).

Gender is hypothesised to be the best predictor of DMIQ1 and DMIQ2 (H4), over and above task success and estimates. Gender is expected to influence the relationship between confidence and DMIQ1 and DMIQ2 (H5) and between task success and DMIQ1 and DMIQ2 (H6).

2. Method

2.1. Participants

There were 40 males and 40 females British adults recruited online. Participants' age ranged from 17 to 40 ($M = 26.65$, SD = 10.21) years. 10% of participants completed GSCE/O-levels/10th grade or similar level of education, 63% achieved A-levels/12th grade and non-university level of education, 11% achieved BA/BSc level, and 15% achieved MA/MSc/MBA or equivalent level of education.

2.2. Measures

2.2.1. Domain-Masculine Intelligence Type (DMIQ)

This is a self-estimated intelligence measure (Storek & Furnham, 2012) which included mathematical/logical and spatial intelligences that together form the Domain-Masculine Intelligence Type measured twice. The Alpha for DMIQ1 was .73 and for DMIQ2 .85.

2.2.2. Timed psychometric tasks

2.2.2.1. Numerical and reasoning problems (Bryon, 2006). Six numerical and reasoning problems that were taken from an intelligence test training book were presented in two 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 3 min was given. Participants were advised to leave unanswered problems blank, in order not to exceed the time limit, or face disqualification. Correct answers were available at the end of the survey. Alpha for the six items was .54 and the inter-item correlation was $r = .16$. (Range .07–.37)

2.2.2.2. General knowledge (GKT: Irwing, Cammock, & Lynn, 2001). Crystallised intelligence (Gc) were included in order to assess whether the Gc problems will produce previously observed sex differences (Irwing et al., 2001). Ten items (one from each section) from the 72-item questionnaire measuring general knowledge were selected, assessing knowledge of literature, general science, medicine, games, and finance. A time limit of 2 min was given. The alpha for the ten items was .81.

2.2.3. Task Success Probability Estimation Measure (TSP) (Storek & Furnham 2012)

Participants were asked to indicate how likely they felt they would succeed on a similar task but with increased difficulty, using a rating scale where 1 was *Very Unlikely* and 5 *Very Likely*, after each block. The three task success probability statements made up the task success probability measure, with individual scores computed for all participants. The alpha for the three-item measure was .66 and the inter-item correlation was = .39.

2.2. Procedure

Participants were members of public who were recruited to participate in an online experiment. They were recruited through an intensive email campaign by the five researchers. An email invitation, with an URL link (www.zoomerang.com) to the study and a background explanation of the study was sent to all participants. The snow-balling technique of participant recruitment was used. A total of one hundred and thirty-six participants completed the test in the set period. There were 96 females (71%) and 40 males (29%). However, 56 female subjects were dropped and 40 males (50%) and 40 females (50%) were randomly selected, bringing the total number of participants to 80.

Detailed scoring instructions were given at the beginning of each measure, including timing instructions for the psychometric problems. Participants were first asked to estimate their IQ prior (*T*1) and then post (*T*2) to completing a psychometric tasks (TCAP) and then assessing their task-success confidence (TSP). Debrief feedback, correct answers and an opportunity to leave comments about the survey was provided. Ethical permission was applied for and granted.

3. Results

Preliminary analysis of the data showed that the application of the tests appropriate. The two ability measures were significantly correlated ($r = .61$, $p < .001$)

3.1. Domain-masculine intelligence and the hubris-humility effect in T1 and T2

A Gender × DMIQ (pre/post) was significant ($F(1,70) = 4.97$, $p < .001$). Missing data explains the drop in *N* for all tasks. Then *t*-tests, with Bonfereoni corrections confirmed significant differences between males ($M = 114.29$, SD = 15.45) and females ($M = 98.50$, SD = 10.26) in the DMIQ at *T*1 ($t(68) = 5.39$, $p = .001$). The magnitude of differences in the means was large ($\eta^2 = .30$, Hedge's $d = 1.20$). Then *t*-tests, confirmed significant differences between male ($M = 113.06$, SD = 17.22) and females ($M = 94.10$, SD = 12.92) in the DMIQ at *T*2, ($t(70) = 5.53$, $p = .001$,). The magnitude of differences in the means was large ($\eta^2 = .30$, Hedge's $d = 1.25$). There was a statistically significant decrease in DMIQ from *T*1 ($M = 106.74$, SD = 15.92) to *T*2 ($M = 102.79$, SD = 17.70), $t(71) = 4.87$, $p = .001$. Hypothesis 1 was thus confirmed.

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

Table 1 shows that a *t*-test for Total TSP measure was significant, with males providing higher TSP estimates than females and medium effect size ($\eta^2 = .07$). Among the three individual

Table 1
Independent *t*-tests and effect sizes for task-success probability estimation and 3 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	3.22 (.72) 31	2.81 (.76) 39	2.25(68)*	.40	.05	.76	.07	.55
TSP1	3.22 (.94) 32	2.90 (.94) 39	1.43(69)	.32	−.13	.77	.03	.34
TSP2	3.27 (.98) 33	2.90 (.97) 39	1.63(70)	.38	−.08	.83	.04	.38
TSP 3	3.24 (.97) 33	2.64 (1.04) 39	2.52(70)*	.60	.13	1.08	.08	.60

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

Table 2
Correlations, means and standard deviations between DMIQ1 and DMIQ2, Gender, TSP, TCAP and Age.

	DMIQ1 112.86 (19.37)	DMIQ2 108.43 (21.20)	Gender 1.66 (.47)	TSP 3.00 (.82)	TCAP 4.34 (4.45)	Age 22.33 (6.86)
DMIQ1						
DMIQ2	.92***					
Gender	−.52***	−.54***				
TSP	.35**	.36**	−.26*			
TCAP	.29**	.37**	.04	.38**		
Age	−.09	−.15	−.03	.12	−.00	

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

TSP probes, only TSP3 that was asked after the crystallised intelligence items was significant, with medium effect size ($\eta^2 = .08$); males provided higher task-confidence than females.

The correlational results (see Table 2) revealed a negative correlation between gender and Total TSP, ($r = -.26$, $p < .05$), with males being more confident about their ability to resolve similar, yet more difficult, psychometric tasks than females, females ($M_{\text{Males}} = 3.22$, $SD_{\text{Males}} = .72$; $M_{\text{Females}} = 2.81$, $SD_{\text{Females}} = .76$). For males no significant relationships were observed. For females, a positive correlations were observed between TSP and DMIQ1 ($r = .32$, $p < .05$) and between TSP and DMIQ2 ($r = .34$, $p < .05$).

In order to investigate whether TCAP correlated differently in male and female subsamples, the data was split per gender and the correlations re-ran. For males the data revealed a positive relationship between TCAP and DMIQ1 ($r = .41$, $p < .01$). For females, a positive correlation was observed between TCAP and DMIQ2 ($r = .46$, $p < .01$) and between TCAP and TSP ($r = .46$, $p < .01$). Inspection of the correlational results (see Table 2) revealed no significant relationship between TCAP and gender ($r = .04$) and nor was a *t*-test for TCAP significant, $t(67)$ −.31, $p = .76$; $M_{\text{Males}} = 7.25$, $SD_{\text{Males}} = 4.30$; $M_{\text{Females}} = 7.50$, $SD_{\text{Females}} = 2.79$). A *t*-test for TCAP revealed significant gender differences $t(67) = 3.96$, $p = .001$, two-tailed between males ($M_{\text{Males}} = 3.18$, $SD_{\text{Males}} = .80$) and females ($M_{\text{Females}} = 2.88$, $SD_{\text{Females}} = .81$). Hypothesis 3 was confirmed.

2 × 2 χ^2 tests and effect sizes for six numerical and reasoning and ten crystallised knowledge problems. No significant gender differences were observed on the sixteen problems.

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

First, the relationships between the DMIQ1 and DMIQ2, gender, TSP and TCAP were explored (see Table 2). DMIQ1 and DMIQ2 were significantly intercorrelated ($r = .92$, $p = .001$). Gender correlated negatively with DMIQ1 ($r = -.52$, $p = .001$) as well as DMIQ2 ($r = -.54$, $p = .001$), with females providing lower scores than males. A positive relationship was observed between DMIQ1 and TSP ($r = .35$, $p < .01$) and DMIQ2 and TSP ($r = .36$, $p < .01$) as well as between TCAP and DMIQ1 ($r = .29$, $p < .01$) and DMIQ2 ($r = .37$, $p < .01$). Gender correlated negatively with TSP ($r = -.26$, $p < .05$) and there was also a positive correlation between TSP and TCAP ($r = .38$, $p < .01$). Given participants' age range (43 years), age was included in the correlational analysis to explore whether it had an impact on DMIQ estimates. Age did not correlate with any of the variables.

3.4. Gender as the best predictor of DMIQ1 and DMIQ2

To determine the best predictor of DMIQ1 and DMIQ2 hierarchical multiple regressions were performed (Table 3).

The first model predicting DMIQ1 was significant $F(3,66) = 13.27$, $p = .001$, Adjusted $R^2 = .35$, $f^2 = .61$), with the overall model explaining 38% of total variance. Gender ($\beta = -.50$, $p = .00$, $r_{\text{part}} = -.48$) and TCAP ($\beta = .26$, $p < .05$, $r_{\text{part}} = .24$) were significant predictors of DMIQ1,with gender accounting for 23% and TCAP for 6% of variance. TSP did not significantly contribute to the prediction of DMIQ1. Gender was the best predictor of the DMIQ1.

The second model, predicting DMIQ2 was also significant $F(3,66) = 17.77$, $p = .001$, Adjusted $R^2 = .42$, $f^2 = .82$), with the overall model explaining 45% of total variance. Gender ($\beta = -.53$, $p = .00$, $r_{\text{part}} = -.50$) and TCAP ($\beta = .36$, $p < .01$, $r_{\text{part}} = .33$) were significant predictors of DMIQ2, explaining 25% and 11% of variance respectively. TSP did not significantly contribute to the prediction of DMIQ2. Thus, the results were identical to DMIQ1, with gender confirmed as the best predictor. Hypothesis 4 was confirmed.

Table 3
Beta coefficients for Simultaneous Multiple Regressions of Gender, TSP and TCAP onto DMIQ1 and DMIQ2.

Dependent Variable	DMIQ1		DMIQT2	
	β	t	β	t
Gender	**−.50**	**−4.90*****	**−.53**	**−5.49*****
TSP	.12	1.06	.09	.84
TCAP	**.26**	**2.46***	**.36**	**3.56****
Regression Model	**$F(3, 66) = 13.27$*****		**$F(3, 66) < 17.77$*****	
$R2$	.38		.45	
R^2 Change	.38		.45	
Adj. R^2	.35		.42	
f^2	.61		.82	

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

Table 4
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)	100.58 (13.33)	112.19 (11.76)	95.12 (10.35)	1.93	27.85***	.01
	G2 (M)	106.81 (16.50)	115.17 (17.25)	98.44 (11.11)			
	G3 (H)	111.33 (16.41)	119.82 (17.44)	102.19 (8.90)			
DMIQ2	G1 (L)	95.50 (16.41)	107.50 (14.58)	89.85 (14.32)	2.01	20.41***	.06
	G2 (M)	103.47 (17.27)	110.94 (19.67)	96.00 (11.01)			
	G3 (H)	107.72 (17.42)	116.43 (17.91)	98.35 (11.25)			

$^{*}p < .05$. $^{**}p < .01$. *** $p < .001$ (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.

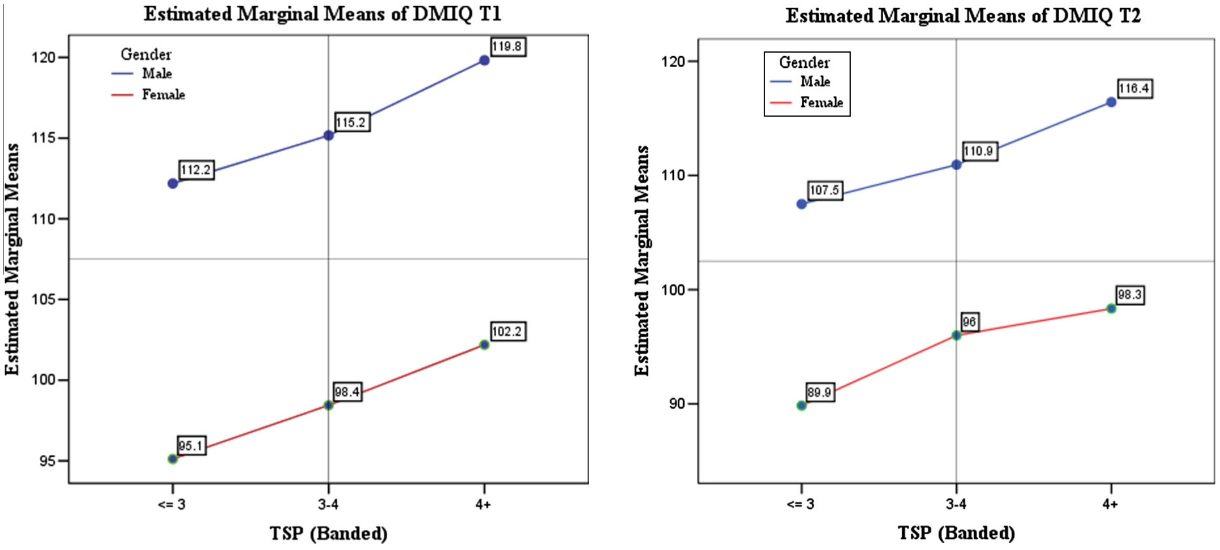

Fig. 1. Two 2-way ANOVAs (TSP and Gender) on DMIQ1 and DMIQ2.

3.5. Impact of gender on the relationship between TSP and DMIQ1 and DMIQ2

TSP was collapsed into a categorical variable with Group 1 containing individuals with lowest confidence in their ability to successfully solve similar tasks in future (TSP $\leqslant 3$, $n = 25$), Group 2 made of individuals that had an average confidence (TSP 3–4, $n = 18$), and Group 3 made of highly confident individuals (TSP 4+, $n = 27$).

Two 2-way between-groups ANOVA were conducted to explore whether gender influences the relationship between TSP and DMIQ1 and DMIQ2. Results are presented in Table 4. For DMIQ1, the Levene's Test of Equality of Error Variance was significant ($p < .05$), indicating the DMIQ2 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, $F(2,64) = .01$, $p = .99$, $\eta_p^2 = .001$. The main effect for TSP, $F(2,64) = 1.93$, $p = .15$, $\eta_p^2 = .06$, was also not significant.

The main effect for gender was significant, $F(1,64) = 27.85$, $p = .00$, $\eta_p^2 = .30$, with a large effect size. Planned contrasts revealed no significant differences between the groups. Post-hoc compari-

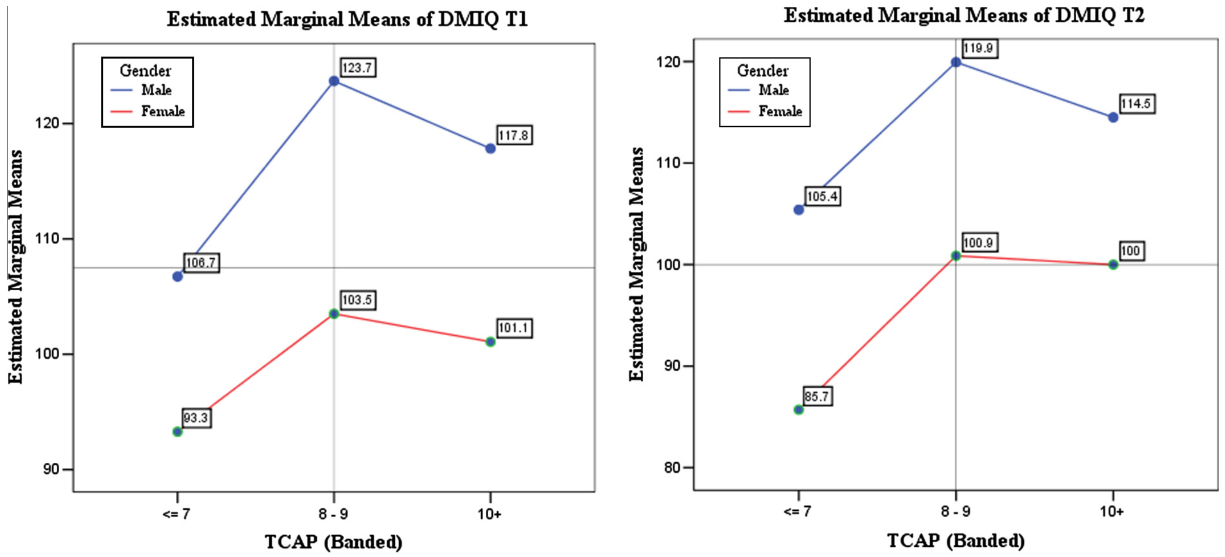

Fig. 2. Two 2-way ANOVAs (TCAP and Gender) on DMIQ1 and DMIQ2.

Table 5
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 × gender
DMIQ1	G1 (L)	99.81 (14.00)	106.74 (14.71)	93.28 (9.75)	9.33***	34.28***	.52
	G2 (M)	110.52 (13.02)	123.69 (10.24)	103.50 (7.88)			
	G3 (H)	124.03 (14.29)	129.21 (13.80)	114.40 (9.55)			
DMIQ2	G1 (L)	93.00 (17.94)	105.40 (19.98)	85.71 (12.08)	7.40***	26.32***	.21
	G2 (M)	107.50 (13.69)	119.94 (13.75)	100.87 (8.03)			
	G3 (H)	109.50 (16.27)	114.50 (15.99)	100.00 (12.75)			

* $p < .05$ ** $p < .01$ *** $p < .001$ (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 Score.

sons using the Games-Howell test indicated that the mean score for Group 1 (<= 3) was significantly different from Group 3 (4+). No other differences were observed. Hypothesis 5 was partially confirmed (see Fig. 1).

For DMIQ2, the interaction effect between gender and TSP estimation conditions was not significant, $F(2,64) = .06$ $p = .94$, $\eta_p^2 = .001$. The main effect for TSP, $F(2,64) = 2.01$, $p = .14$, $\eta_p^2 = .06$, was also not significant. There was a statistically significant main effect for gender, $F(1,64) = 20.41$, $p = .001$, $\eta_p^2 = .24$, with a large effect size. Planned contrasts revealed no significant differences between the groups. Post-hoc comparisons using the Tukey HSD and Bonferroni tests indicated that the mean score for Group 1 (<=3) was significantly different from Group 3. No other significant differences were observed (see Fig. 2).

3.6. Impact of gender on the relationship between TCAP and DMIQ1 and DMIQ2

Individual scores for the sixteen correctly solved psychometric problems were computed, creating a new variable TCAP. TCAP was collapsed into a categorical variable, with Group 1 made of individuals who correctly solved fewest problems (TCAP $\leqslant 7$, $n = 35$), Group 2 of individuals who solved an average number of problems (TCAP 8–9, $n = 23$) Group 3 of individuals that correctly solved the most psychometric problems (TCAP 10+, $n = 22$).

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 5. For DMIQ1, the interaction effect between gender and TCAP was not significant, $F(2,74) = .52$, $p = .60$, $\eta^2 = .01$. The main effect for TCAP, $F(2,74) = 9.33$, $p = .00$, $\eta^2 = .20$, was significant, with large effect size. The main effect for gender $F(1,74) = 34.28$, $p = .00$, $\eta^2 = .32$ was also significant, with a very large effect size. Planned contrasts revealed significant differences between Group 1 and Group 2, (Contrast Estimate −11.52, $p = .001$). Post-hoc comparisons indicated that the mean score for Group 1 (<=7) was significantly different from Group 2 (8–9) as well as from Group 3 (10+).

The interaction effect between gender and TCAP was not significant, $F(2,66) = .21$, $p = .81$, $\eta^2 = .01$. The main effect for TCAP, $F(2,66) = 7.40$, $p = .001$, $\eta^2 = .18$ was significant, with large effect size. The main effect for gender, $F(1,66) = 26.32$, $p = .001$, $\eta^2 = .29$ was significant with a very large effect size. Planned contrasts revealed significant differences between Group 1 and Group 2, (Contrast Estimate −13.27, $p = .00$). Post-hoc comparisons using the Tukey HSD and Bonferroni tests indicated that the mean score for Group 1 (<=7) was significantly different from Group 2 (8–9) as well as from Group 3 (10+).

4. Discussion

The repeated measurement of self-assessed intelligence (DMIQ) aimed to ascertain that humility/hubris effect HHE can be reduced following the psychometric task and task-confidence probes, based on the assertions that repeated measurement affects mood, confidence and behaviour (Bartsch & Nesselroade, 1973). The results confirmed the existence of the Hubris-Humility Effect on DMIQ1 ($\eta^2 = .30$, $d = 1.20$ and DMIQ2 ($\eta^2 = .30$, $d = 1.25$). These results validated the findings of the previous studies (Storek & Furnham, 2012; Storek & Furnham, 2013) as well as provided further support for the role gender plays in HHE and DMIQ. Equally, a significant decrease in DMIQ estimates was observed from the pre-task to post-task estimation condition ($d = .57$), following the psychometric task and task-confidence probes. This has implications for when DMIQ is measured in any study, particularly how recently participants did a test.

Gender differences were observed on the confidence measure with males significantly more confident about their future performance than females. All studies show that females hold more negative self-assessments and stereotypical self-beliefs and have lower performance expectations than do males.

The gender-stereotype literature has provided abundant evidence for female underperformance on domain-masculine tasks (Dar-Nimrod, 2007; Ehrlinger & Dunning, 2003; Hyde, Fennema, & Lamon, 1990). The results established that the psychometric and task-confidence task caused both genders to lower their post-task estimates, although female estimates were lower than the male self-estimates. 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 (Meece, Glienke, & Burg, 2006; Meelissen & Luyten, 2008). Thus, the task combined with the repeated measurement of DMIQ seems to have affected both genders similarly.

Contrary to expectations, no gender differences were observed on the psychometric problems. It is possible that the composition of the psychometric task, i.e. ten crystallised problems and six numerical and reasoning problems, influenced the results. These results challenge the male advantage in crystallised intelligence claims and in particular, in the General Knowledge Test (Irwing et al., 2001). However these were short tests with barely acceptable alphas and the results need replicating with a bigger test battery.

As in previous studies, gender was expected to be the best predictor of DMIQ. Results confirmed gender as the best predictor of DMIQ1 and DMIQ2, explaining 23% and 25% of variance respectively. The psychometric task was the only other significant predictor. It is likely that the inclusion of Gc problems and reduction of the numerical and reasoning problems had positive impact on the perceived ability and future performance on both males and females, reducing the importance of task confidence. Equally, it is possible that more task-confidence probes were necessary in order for confidence to play role in the prediction of DMIQ.

The results showed males being more confident than females about their ability to succeed on a similar task, across the three task-confidence groups. It is possible that male hubris is less vulnerable to (negative) feedback that female humility.

The role gender plays in the relationship between TCAP and DMIQ1 and DMIQ2 was investigated. Significant gender effects revealed that males provided higher DMIQ1 estimates than females across the three groups. These results provide support for the claim that individuals are capable of accurate self-assessments of ability (e.g. Ackerman, Beier, & Bowen, 2002; Chamorro-Premuzic, Harlaar, Greven, & Plomin, 2010; Swim, 1994) as well as further evidence for the male confidence in the domain-masculine intelligence. It seems plausible that the psychometric task, combined with the confidence probes and repeated measurement of DMIQ initiated gender-stereotypical biases in both genders.

This study had various limitations. One 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. 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). Equally, the introduction of multiple measurements of DMIQ estimates was intended to reduce the possible affects of 'subjective' measurement.

Another issue was the size and representativeness of this sample which means it is very desirable to repeat the study on a much larger group to ensure the replicability of these results.

This study was done online with various well known problems and it would be desirable to replicate with a "class-room" based test particularly with respect to ability tests.

Most importantly there were limitations associated with the ability measures themselves which were short and had not particularly good alphas. To have a more comprehensive test battery measuring both fluid and crystallised intelligence is recommended in further studies.

References

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

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.

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. (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.

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.

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

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

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.

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.

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. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*, 139–155.

Irwing, P., Cammock, T., & Lynn, R. (2001). Some evidence for the existence of a general factor of semantic memory and its components. *Personality and Individual Differences, 30*, 857–871.

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.

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.

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

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

Sleeper, L. A., & Nigro, G. N. (1987). It's not who you are but who you're with: Self-confidence in achievement settings. *Sex Roles, 16*, 57–69.

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. (2013). Gender, 'g', and fixed versus growth intelligence mindsets as predictors of self-estimated Domain Masculine Intelligence (DMIQ). *Learning and Individual Differences, 25*, 93–98.

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.

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.