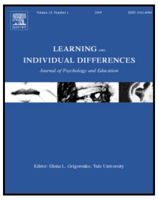
Learning and Individual Differences 50 (2016) 23–29

Contents lists available at ScienceDirect

Learning and Individual Differences

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



The role of gender, task success probability estimation and scores 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, London, United Kingdom
[b] Norwegian Business School, Oslo University College London, United Kingdom

ARTICLE INFO

Article history:
Received 21 April 2015
Received in revised form 1 June 2016
Accepted 20 June 2016
Available online xxxx

ABSTRACT

This paper reports a study aimed at understanding correlates of self-estimated intelligence. Participants twice estimated their mathematical and spatial intelligence (called domain-masculine intelligence type: DMIQ) on a normal distribution, before and after taking ability tests. They 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. As predicted 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 both before and after test estimates, over and above gender and test score, explaining 17% and 23% of variance, respectively. This is explained in terms of Dweck's (2007) mindset theory and Eccles and Wigfield's (2002) motivation theory. Results are discussed in terms of the literature on self-estimated intelligence and stereotype threat.

1. Introduction

While an extensive body of self-estimated intelligence (SEI) research is available, only few SEI studies have used psychometric measures to compare the accuracy and validity of SEI estimates (Batey, Chamorro-Premuzic, & Furnham, 2009; Chamorro-Premuzic, Moutafi, & Furnham, 2005; Holling & Preckel, 2005). To the best of 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 the general population (Furnham, 2001; Furnham & Shagabutdinova, 2012; Stieger et al., 2010; von Stumm, Chamorro-Premuzic, & Furnham, 2009). The strongest gender differences observed is on mathematical/logical and spatial intelligences, followed by overall (g) and verbal intelligence, with significantly higher self-estimates provided by males than females (Furnham, 2001; Rammstedt & Rammsayer, 2002a, 2002b). The magnitude of gender differences in mathematical/logical, spatial, overall and verbal self-assessed intelligences was further revealed in meta-analytical study (Szymanowicz & Furnham, 2011), with the biggest weighted mean effect sizes for mathematical/logical ($d = 0.44$), followed by spatial ($d = 0.43$), overall ($d = 0.37$) and verbal ($d = 0.07$) 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, 2013b, 2014). 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 unclear 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. Indeed, female self-estimates were shown to be significantly more accurate than were those of males.

Male self-estimates have been shown to be significantly inflated compared to their actual psychometric scores (Reilly & Mulhern, 1995). These findings were further substantiated by Carr, Hettinger-Steiner, Kyser, and Biddlecomb (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 poorer performance.

One possible explanation is that in these studies females experience stereotype threat which increases their performance anxiety and hence outcomes both on actual personality tests as well as those examining self-estimates. It seems possible that people develop, often inaccurate, general perceptions of their overall ability ("ability self") over time particularly as a results of schooling, which would impact on their self-rated ability and even test performance. There is a great deal of interest in the concept of stereotype threat as well as critique of its importance

---

[*] Corresponding author at: Research Department of Clinical, Educational and Health Psychology, University College London, London, United Kingdom.
E-mail address: a.furnham@ucl.ac.uk (A. Furnham).

http://dx.doi.org/10.1016/j.lindif.2016.06.011

(Flore & Wicherts, 2015). Nevertheless, gender differences in intelligence may be mediated by various social and cultural factors that impact on consistent and stable gender-based stereotypic threats when it comes to anything concerning intelligence and its measurement. Age has been shown to be related to self-estimated intelligence for various reasons: younger people are often better educated than older people, but older people are often more self-confident and have received more feedback about their intelligence. Hence we will control for age in this study.

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, 2013a, 2013b, 2014), was introduced.

This study was designed to ascertain the determinants of gender differences in the *domain-masculine intelligence* by introducing a number of timed psychometric tasks (TCAP) and confidence assessments (TSP). As in previous research (Storek & Furnham, 2012, 2013a, 2013b, 2014), 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 task confidence as well as actual scores. 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 issue of task confidence is important in this literature and relatively unexplored. Storek and Furnham (2013a) examined the relationship between fixed vs growth mindset ideas derived from Dweck (2007). The suggestion was that those with a growth/incremental/malleable mindset would have greater task confidence over time because of their belief in their ability to learn. However, they found little evidence of a significant relationship between SEI and mindset. However, this may have been because the mindset measure was too general, and neither task nor ability specific. Further, it took not account of specific feedback from task success or failure. In this study we asked participants to estimate how they would do in a similar task to the one they had done. Thus, we expected that entity theorists would tend to be more pessimistic with lower self-confidence than more optimistic incremental theorists. Equally, this self-confidence should relate to effort in subsequent tests which would be self-fulfilling. Indeed task confidence can also be understood in terms of Eccles expectancy value theory, which suggests task persistence is a function of subjective task values. That is, initial task confidence is probably related to subjective task values that are related to intrinsic motivation, interest and effort.

Gender stereotypes, threats and self-confidence are all likely to play a role in HHE or the display of male hubris and female humility in estimation of abilities. Participants were asked to undertake a gender stereotype-inducing task, i.e., numerical and reasoning aptitude problems that are likely to increase hubris and humility as well as general stereotype threats (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, Larrick, & Klayman, 2006; Carr et al., 2008; Dunning, Griffin, Milojkovic, & Ross, 1990; Pallier, 2003). After each block, participants were asked to estimate their task success confidence.

Various hypotheses were tested some essentially replicating previous studies. It was predicted that HHE will be confirmed on DMIQ at the pre-task (T1) and post-task (T2) estimating conditions (H1). However, a more important study-specific hypothesis was that there will be a significant decrease in DMIQ estimates from T1 to T2 following the gender stereotype-inducing task (H2). This could be seen to be a manifestation of stereotype threat.

Existing literature suggests that males have higher self-confidence in general but particularly with respect to intelligence, despite being inaccurate about their (math) skills or underperforming, whereas females often lack confidence, while being accurate or outperforming males (Carr et al., 2008; Eccles-Parsons, Adler, & Meece, 1984; Pallier, 2003). Consequently replicating other research, males are expected to provide significantly higher task success probability estimations (TSP) (i.e., self-confidence) than females (H3).

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

The more important experimental hypotheses are these: gender is expected to be the best predictor of before and after self-estimates, namely, DMIQ T1 (H5) and DMIQ T2 (H6) over and above TSP and TCAP. Finally, gender is presumed to influence the relationship between TSP and DMIQ T1 (H7) and DMIQ T2 (H8). Gender is also expected to affect the relationship between TCAP and both DMIQ T1 (H9) and DMIQ T2 (H10).

2. Methods

2.1. Participants

A total of 488 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.

2.2. Measures

2.2.1. Repeated measure of domain-masculine intelligence type (DMIQ)

Based on the self-estimated intelligence measure (Furnham, 2001), 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. Alphas for DMIQ T1 and DMIQ T2 were 0.82 and 0.88, respectively.

2.2.2. Psychometric aptitude task—total correct aptitude problems (TCAP)

2.2.2.1. 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 s 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 min to complete. Correct answers were available at the end of the survey. Alpha for the fifteen items was 0.93.

2.2.2.2. Task success probability estimation measure (TSP) (Storek & Furnham, 2012). 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 0.82.

2.3. Procedure

Participants were members of public who were recruited to participate in an online experiment. We used university subject panel volunteers: individuals who volunteer to take part in studies. Many, but not all, are young and under- or post-graduates. We attempted to get as wide a range as possible, but supplemented them with undergraduates at our institution. The data were 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. Appropriate ethical permission was applied for and granted for the study.

3. Results

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

An independent samples *t*-tests, $t(385) = 6.16$, $p = 0.001$, 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 ($\eta^2 = 0.09$, Hedge's adjustment $d = 0.66$).

An independent samples *t*-tests, $t(227) = 4.68$, $p = 0.001$, 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 ($\eta^2 = 0.09$, Hedge's adjustment $d = 0.63$). Hypothesis 1 was thus confirmed.

A paired samples *t*-test 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 ($M = 108.21$, $SD = 22.04$), $t(224) = 5.66$, $p < 0.001$, two-tailed, $r = 0.78$, $p < 0.001$. 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 (0.38) indicated a small effect size. Hypothesis 2 was thus confirmed.

3.2. 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. Males providing higher TSP estimates than females. The observed effect sizes were small. Inspection of the correlational results (see Table 2) revealed a negative correlation between gender and TSP ($r = -0.18$, $p < 0.01$), with males providing higher TSP estimates than females ($M_{\text{Males}} = 3.18$, $SD_{\text{Males}} = 0.80$; $M_{\text{Females}} = 2.88$, $SD_{\text{Females}} = 0.81$). Hypothesis 3 was thus also confirmed.

Table 2
Correlations and partial correlations, means and standard deviations between before and after estimates (DMIQ T1 and DMIQ T2), gender, total success perceptions (TSP), actual scores (TCAP) and age.

	DMIQ T1	DMIQ T2	G	TSP	TCAP	A
	112.86	108.43	1.66	3.00	4.34	22.33
	(19.37)	(21.20)	(0.47)	T (0.82)	(4.45)	(6.86)
DMIQ T1						
DMIQ T2	0.78***					
Gender	−0.30***	−0.30***				
TSP	0.47***	0.62***	−0.18**			
TCAP	0.16**	0.40***	−0.18***	0.43***		
Age	0.08	0.01	−0.14**	−0.06	0.12*	
Controlled for age						
DMIQ T1						
DMIQ T2	0.78***					
Gender	−0.29***	−0.30***				
TSP	0.48***	0.63***	−0.19**			
TCAP	0.15**	0.40***	−0.17**	0.44***		

Note. *N* between 198 and 487.
* $p < 0.05$ (two-tailed).
** $p < 0.01$ (two-tailed).
*** $p < 0.001$ (two-tailed).

Equally, inspection of the correlational results (see Table 2) revealed a small negative correlation between gender and TCAP ($r = -0.18$, $p < 0.001$), with males correctly solving more problems than females. An independent samples *t*-test for TCAP revealed significant gender differences $t(307) = 3.96$, $p < 0.001$, two-tailed between males ($M_{\text{Males}} = 5.47$, $SD_{\text{Males}} = 4.60$) and females ($M_{\text{Females}} = 3.77$, $SD_{\text{Females}} = 4.27$). The magnitude of the differences in the means (mean difference = 0.43, 95% CI: 0.86 to 2.55) was small ($\eta^2 = 0.05$; Hedge's adjustment $d = 0.01$).

The 2×2 χ^2 tests and the effect sizes for the 5×3 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 therefore also confirmed.

3.3. Gender, task success probability (TSP) and total correct aptitude problems (TCAP) as predictors of DMIQ T1 and DMIQ T2

First, the relationships between the DMIQ T1 and DMIQ T2, gender, TSP and TCAP were explored. Table 2 shows the results of the

Table 1
Independent *t*-tests and effect sizes for task success probability (TSP) estimation and 5 individual TSP probes.

	Males	Females			95% CI		Effect size	
	M (*SD*) *n*	*M* (*SD*) *n*	*t*(*df*)	Mean difference	L	U	η^2	*d*
Total TSP	3.18 (0.80) 90	2.88 (0.81) 132	2.75(220)*****	0.30	0.09	0.52	0.03	0.37
TSP1	3.61 (1.09) 99	3.32 (1.04) 154	2.11(251)*	0.29	0.02	0.56	0.02	0.27
TSP2	2.81 (1.04) 110	2.54 (1.04) 150	2.01(248)*	0.27	0.01	0.54	0.02	0.48
TSP 3	3.43 (1.02) 98	2.97 (1.10) 143	3.27(237)**	0.46	0.18	0.73	0.04	0.43
TSP 4	3.40 (0.91) 99	3.20 (1.09) 143	1.51(240)	0.20	−0.06	0.46	0.01	0.20
TSP 5	2.67 (1.15) 96	2.31 (1.13) 140	2.38(234)*	0.36	0.06	0.66	0.02	0.31

Note. d = Hedge's adjustment or Cohen's d adjusted for sample size.
* $p < 0.05$ (two-tailed).
** $p < 0.01$ (two-tailed).
*** $p < 0.001$ (two-tailed).

correlational and partial correlational analyses. DMIQ T1 and DMIQ T2 were strongly intercorrelated ($r = 0.78$, $p < 0.001$). Gender correlated negatively ($r = -0.30$, $p < 0.001$), with DMIQ T1 as well as DMIQ T2 ($r = -0.30$, $p < 0.001$), with females providing lower scores than males. A positive relationship was observed between DMIQ T1 and TSP ($r = 0.47$, $p < 0.001$) and DMIQ T2 and TSP ($r = 0.62$, $p < 0.001$). DMIQ T1 also correlated positively with TCAP ($r = 0.16$, $p = 0.01$) as did DMIQ T2 ($r = 0.40$, $p < 0.001$). The correlations between TSP, TCAP and DMIQ T2 were stronger than with DMIQ T1. A medium positive correlation was observed between TSP and TCAP ($r = 0.43$, $p < 0.001$).

As in previous studies (Storek & Furnham, 2012, 2013a, 2013b), 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 DMIQ T1 and DMIQ T2. A positive relationship between age and TCAP ($r = 0.12$, $p = 0.01$) indicated that older participants solved more TCAP problems which is contrary to assertions that fluid cognitive ability declines with age (Beier & Ackerman, 2001, 2003; Deary et al., 2003). The correlations were re-run, with age partialled out, but this had little impact on the strength of the observed relationships, with values slightly higher. It is likely however that social class and education is likely to have more of an effect on the variables used in this experiment, but which were not measured and therefore available for examination.

Subsequently, the data were split per gender and the correlational analysis recomputed. The results are presented in Table 3. TSP displayed a strong positive relationship with DMIQ T1 and DMIQ T2 for both genders, with stronger correlations between TSP and DMIQ T2 than between TSP and DMIQ T1. Medium positive correlations were observed between TCAP and DMIQ T2 for both genders, but no significant relationships were observed between TCAP and DMIQ T1. These findings indicate that the relationships between TSP and TCAP and DMIQ became stronger following the task.

3.4. Gender as the best predictor of DMIQ T1 and DMIQ T2

To determine the best predictor of DMIQ T1 and DMIQ T2, simultaneous multiple regressions were performed. Results are reported in Table 4. The first model predicting DMIQ T1 was significant with the overall model explaining 27% of total variance. Gender and TSP were significant predictors of DMIQ T1, with gender accounting for 5% and TSP for 17% of variance. TCAP did not significantly contribute to the prediction of DMIQ T1. Contrary to prediction, TSP and not gender was the best predictor of the DMIQ T1. Hypothesis 5 was not supported. The second model, predicting DMIQ T2, was also significant with the overall model explaining 44% of total variance. Gender, TSP and TCAP were significant predictors, explaining 3%, 23% and 1% of variance, respectively. As in DMIQ T1, TSP, and not gender, was the best predictor of DMIQ T2. Hypothesis 6 was also not supported.

Table 3
Correlations, means and standard deviations between before and after estimates (DMIQ T1 and DMIQ T2), total success estimates (TSP), actual scores (TCAP) and age—per gender.*, **

	Males		Females	
	DMIQ T1	DMIQ T2	DMIQ T1	DMIQ T2
M (*SD*)	120.64 (18.13)	116.02 (21.58)	108.55 (18.70)	102.57 (21.14)
DMIQ T1				
DMIQ T2	0.64***		0.83***	
TSP	0.49***	0.65***	0.41***	0.57***
TCAP	0.14	0.44***	0.10	0.31***
Age	0.01	0.08	0.07	−0.07

N between 47 and 321.
* $p < 0.05$ (two-tailed).
** $p < 0.01$ (two-tailed).
*** $p < 0.001$ (two-tailed).

Table 4
Beta coefficients for simultaneous multiple regressions of gender, total success estimates (TSP) and actual scores (TCAP) onto before and after estimates (DMIQ T1 and DMIQ T2).

	DMIQ T1		DMIQ T2	
Dependent variable	β	t	β	t
Gender	**−0.23**	**−3.83*****	**−0.18**	**−3.26****
TSP	**0.46**	**7.07*****	**0.54**	**9.17*****
TCAP	−0.08	−1.20	**0.14**	**2.34***
Regression model	**$F(3, 212) = 26.48$*****		**$F(3, 205) = 53.43$*****	
R^2	0.27		**0.44**	
R^2 Change	0.27		**0.44**	
Adj. R^2	0.26		**0.43**	
f^2	0.37		**0.79**	

Note. Significant values are in bold.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

3.5. Impact of gender on the relationship between TSP and DMIQ T1 and DMIQ T2

Two 2-way between-groups analysis of variance were conducted to explore whether gender influences the relationship between TSP and DMIQ T1 and DMIQ T2. Results are presented in Table 5. The interaction effect between gender and TSP estimation conditions was not significant, $F(2210) = 0.30$, $p = 0.74$, $\eta_p^2 = 0.00$. There was a statistically significant main effect for TSP, $F(2210) = 19.56$, $p < 0.001$, $\eta_p^2 = 0.16$ with large effect size. The main effect for gender was also significant, $F(1210) = 13.26$, $p < 0.001$, $\eta_p^2 = 0.06$, with medium effect size.

Planned contrasts revealed significant differences between Group 1 and Group 2 (contrast estimate −13.68, $p < 0.001$) and between Group 2 and Group 3 (contrast estimate −10.93, $p < 0.001$). 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. Hypothesis 7 was partially confirmed.

For DMIQ T2, the interaction effect between gender and TSP was not significant, $F(2203) = 0.16$, $p = 0.86$, $\eta_p^2 = 0.00$. There was a statistically significant main effect for TSP, $F(2203) = 34.82$, $p = 0.00$, $\eta_p^2 = 0.26$, with large effect size, and for gender, $F(1203) = 11.10$, $p < 0.01$, $\eta_p^2 = 0.05$, with medium effect size. Planned contrasts revealed significant differences between Group 1 and Group 2 (contrast estimate −21.46, $p < 0.001$) and between Group 2 and Group 3 (contrast estimate −12.47, $p < 0.001$). 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. Hypothesis 8 was partially confirmed.

3.6. Impact of gender on the relationship between TCAP and DMIQ T1 and DMIQ T2

Two 2-way between-groups analyses of variance were conducted to explore whether gender influences the relationship between TCAP and DMIQ T1 and DMIQ T2. For DMIQ T1, the interaction effect between gender and TCAP was significant, $F(2381) = 3.26$, $p < 0.05$, $\eta^2 = 0.02$, with small effect size. The main effect for TCAP, $F(2381) = 19.56$, $p = 0.00$, $\eta^2 = 0.09$, was also significant, with medium effect size. The main effect for gender, $F(1381) = 26.49$, $p = 0.00$, $\eta^2 = 0.07$, was also significant, with medium effect size. Planned contrasts revealed significant differences between Group 2 and Group 3 (contrast estimate −14.73, $p < 0.001$).

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.

Table 5
Two-way ANOVA (TSP and gender) on before and after estimates (DMIQ T1 and DMIQ T2).*

		Mean score (*SD*)			*F*-score		
Variable	TSP groups	Total	Males	Females	TSP	Gender	TSP × Gender
DMIQ T1	G1 (L)	104.43 (20.17)	111.21 (23.80)	100.98 (17.28)	19.56***	13.26***	0.30
	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)			
DMIQ T2	G1 (L)	94.56 (23.04)	101.38 (27.69)	91.33 (19.97)	34.82***	11.10**	0.16
	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)			

Note. DMIQ T1 = domain-masculine intelligence type at pre-task estimation condition; DMIQ T2 = domain-masculine intelligence type at post-task estimation condition; TSP = task success probability estimation condition.

* $p < 0.05$ (two-tailed).

** $p < 0.01$ (two-tailed).

*** $p < 0.001$ (two-tailed).

As the main interaction effect was significant, further investigation of the relationship was warranted. Simple effects analysis was conducted. The data were 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 DMIQ T1 was significant, $F(2135) = 16.01$, $p < 0.001$, $\eta^2 = 0.19$, with large effect size. The robust tests of equality of means, Welch $(2, 72) = 12.83$, $p < 0.001$; Brown–Forsythe $(2, 97) = 14.67$, $p < 0.001$ 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 ($p < 0.05$) in the female sub-sample. As a result, a more stringent significance level, i.e., $p = 0.01$, was set for evaluating the results of the analysis. For females, the one-way analysis of variance was also significant, $F(2246) = 5.87$, $p < 0.01$, $\eta^2 = 0.05$, with medium effect size. The robust tests of equality of means, Welch $(2, 160) = 7.55$, $p < 0.01$; Brown–Forsythe $(2227) = 6.14$, $p < 0.01$ 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 DMIQ T2, the interaction effect between gender and TCAP was not significant, $F(1225) = 0.01$, $p = 0.94$, $\eta^2 = 0.00$. The main effect for TCAP, $F(1225) = 28.35$, $p < 0.001$, $\eta^2 = 0.11$, was significant, with medium effect size. The main effect for gender, $F(1225) = 12.99$, $p = 0.00$, $\eta^2 = 0.06$, was significant with medium effect size. Planned contrasts revealed significant differences between Group 1 and Group 2 (contrast estimate -15.18, $p < 0.001$). 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.

4. Discussion

The results confirmed the existence of gender differences on the numerical–spatial factor of SEI. A significant decrease in estimates was observed from the pre-task to post-task estimation condition ($d = 0.38$). 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 confidence and test scores were small.

The findings also revealed a stronger relationship between confidence, test score and later self-estimates compared to the initial estimate. This pattern was also observed when the data were split per gender. Interestingly, for both genders, test scores only correlated with the second self-estimate and not the first. These results appear to indicate that task confidence already played a role in the estimation process, indicating the individuals rely on their confidence before they are prompted to do so. In this sense, confidence may be self-fulfilling: those who believed they were good at the task took it more seriously and put in more effort than those who believed they would not do well and confirmed their beliefs.

This result may be explained in terms of Eccles expectancy value theory of motivation (Eccles & Wigfield, 2002). Thus, continued confidence is a function of expectation of doing reasonably well at the task, which has a mixture of attainment, intrinsic, utility and cost value. It is thus an index of sustained effort and motivation to do well in the experimental tasks.

As in previous studies, gender was expected to be the best predictor of self-estimated intelligence. The results failed to validate this claim, with test confidence confirmed as the best predictor of before and after estimates, over and above gender and actual scores, explaining 17% and 23% of variance, respectively. Thus, it appears that task confidence/motivational factor plays an important role in the prediction of the intelligence type. Interestingly, this variable as not been much examined in the self-estimated intelligence literature and may prove to be a most important mediating and moderating factor.

For the initial self-estimated intelligence type, results revealed significant task success effect, with significant differences between the lowest, average and high task success groups. The lowest estimates were provided in the lowest confidence group, average estimates in the average confidence group and the highest estimates in the highest confidence 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 the second estimate of intelligence.

For the initial self-estimate, the results revealed a significant interaction effect as well as significant test score and gender effects. Significant differences between the three scoring groups were observed; with lowest estimates provided by the group that solved an average number of psychometric problems, average 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 (Alicke, Klotz, Breitenbecher, Yurak, & Vredenburg, 1995; Kruger & Dunning, 1999).

Further, males provided higher initial estimates than females in all three groups, similar to all other studies (Furnham, 2001). Further analyses showed that males' initial self-estimates were significantly different in the lowest and medium test-scoring groups as well as between the medium and the highest test-scoring groups. Significant differences were also observed for females, with initial estimates significantly different in the lowest and highest as well as between medium and highest test-scoring groups.

For the latter, second self-estimates of this intelligence type the results revealed a significant test-scoring effect, with findings identical to the first estimation pattern. Equally, a significant gender effect revealed that males provided higher second estimate than females across the three groups, providing further support for the hubris-humility effect in self-estimated intelligence (Furnham, 2001).

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. It would be very interesting to see if this effect could be replicated on a female-favouring stereotypic task, namely, than females would have greater confidence than males.

The repeated measurement of self-estimated ability aimed to ascertain that hubris-humility effect 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, Gold, & Rodda, 1971). The results of the study confirmed the existence of this effect in the pre- and post-task domain-masculine intelligence estimates as well as significant reduction in the intelligence type estimates from pre- to post-task estimation condition. The effect sizes for effect's occurrence on, before and after self-estimates, ranged from medium to very large and the effect sizes for the intelligence estimate reduction ranged from small to medium. These results validated the findings of the previous studies (Storek & Furnham, 2012, 2013a, 2013b, 2014) as well as provided further support for the role gender plays in hubris-humility and self-estimated intelligence.

The main limitation of this study 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, 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, numerous studies about the accuracy of "subjective" assessments have shown that individuals are capable of accurate self-assessments of ability and that the current self-estimate measures are valid proxies of intellectual competence (Ackerman, Beier, & Bowen, 2002; Chamorro-Premuzic, Harlaar, Greven, & Plomin, 2010; Swim, 1994). Equally, the introduction of multiple measurements of domain-masculine intelligence estimates was intended to reduce the possible effects of "subjective" measurement. The experimental findings replicated the earlier correlation results (e.g., Storek & Furnham, 2012, 2013a, 2013b), providing further support for the observed results.

One important omission in this study, which is alas very common in experimental psychology, is to not explore the participants experience in a debriefing session. It may have been possible to explore further the whole issue of stereotypic threat and even the possible effect of the experimental manipulation even though it was explained. Further, studies in this area may well benefit from investigating participant's general beliefs about ability and intelligence as well as their specific experiences/motivations while taking part in these studies to understand fully the processes involved in self-estimated intelligence.

References

Ackerman, P. L., Beier, M. E., & Bowen, K. R. (2002). What we really know about our abilities 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.

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., 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.

Dreary, 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.

Dweck, C. (2007). *Mindset*. London: Constable and Robinson.

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*, 109–132.

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.

Flore, P., & Wicherts, J. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology*, *53*, 25–44.

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

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

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.

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.

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

Novell, A., & Hedges, L. V. (1998). Trends in gender differences in academic achievement from 1960 to 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.

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 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.

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.

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., & 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. *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.

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.