Supplementary Information

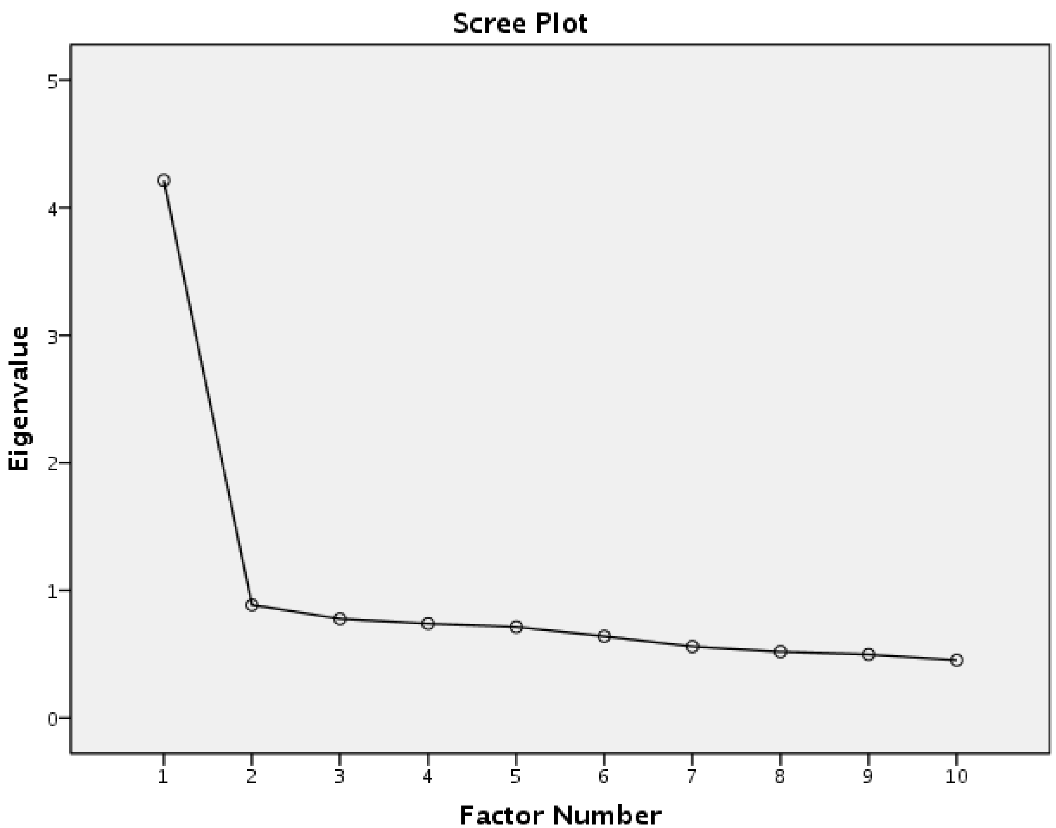
Spatial ability or spatial abilities? Investigating the phenotypic and genetic structure of spatial ability

Kaili Rimfeld, Nicholas G. Shakeshaft, Margherita Malanchini, Maja Rodic, Saskia Selzam, Kerry Schofield, Philip S. Dale, Yulia Kovas & Robert Plomin

Figures

Tables

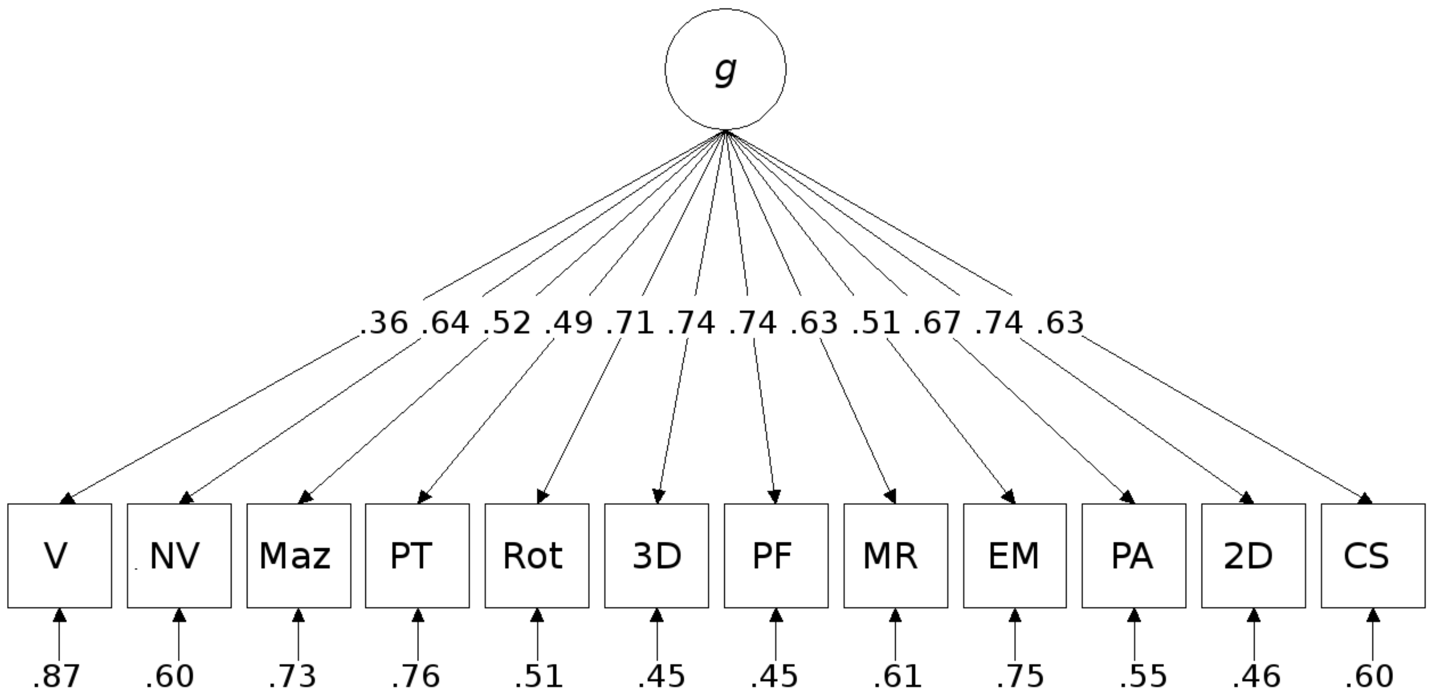
Figure S1. Scree plot illustrating the proportion of variance explained by the extracted factors from the ten spatial tests.



The scree plot taken from the exploratory factor analysis (EFA) illustrating the factor structure of spatial ability. Only a single factor emerges with an eigenvalue above 1, indicating that spatial ability is unifactorial.

Figure S2. Confirmatory factor analyses for spatial ability and *g*, with a) one factor and b) two factors.

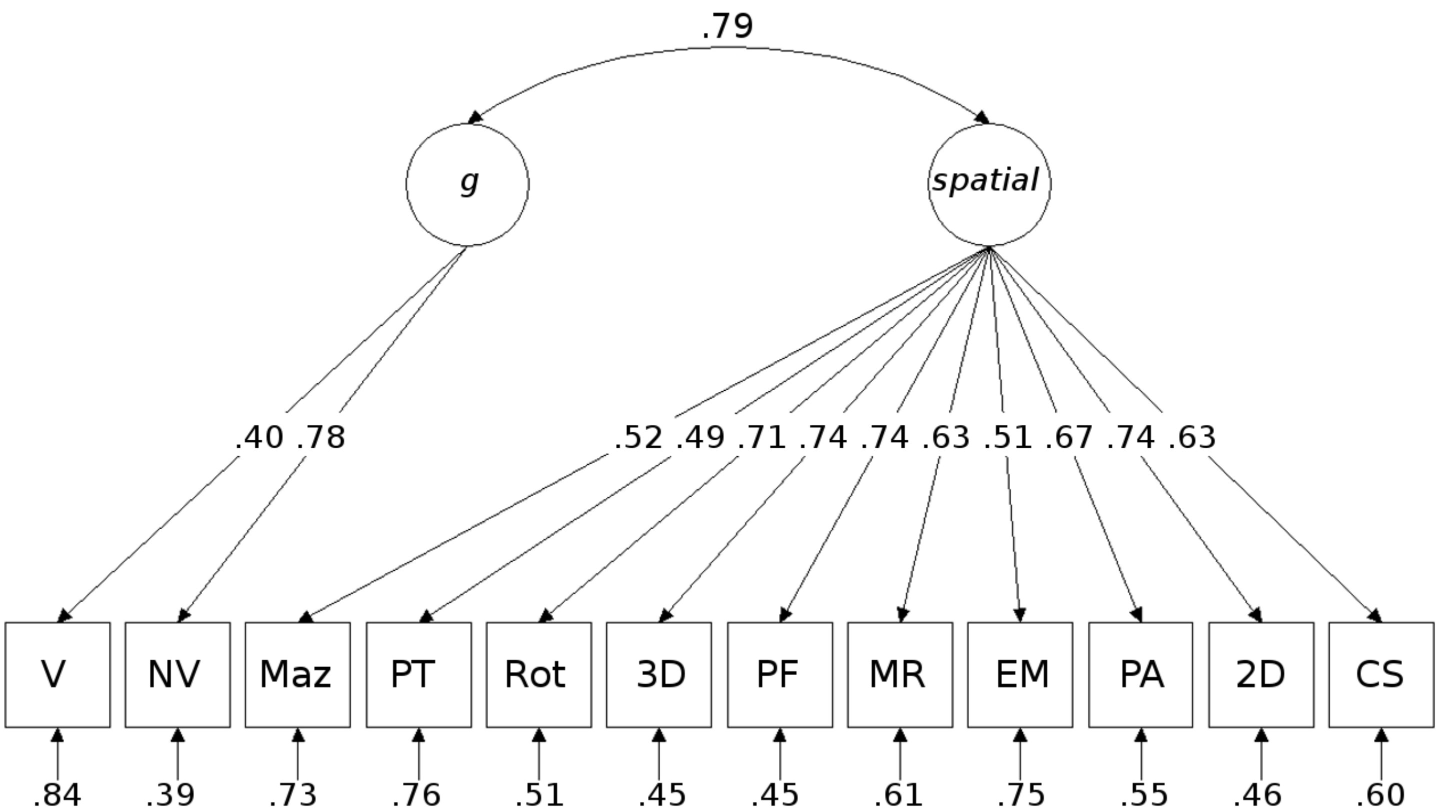
a) One factor



	AIC	BIC	X^2	RMSEA	CFI	TLI	SRMR
1 Factor Model	53273.19	53476.99	166.23 **	0.03	.98	.97	0.03

Confirmatory factor analysis loadings and fit indices with two non-spatial and ten spatial tests loading onto a single *g* factor. V=verbal ability (Mill Hill vocabulary); NV=non-verbal ability (Raven's matrices); Maz=mazes; PT=perspective-taking; Rot=mental rotation; 3D=3D drawing; PF=paper folding; MR=mechanical reasoning; EM=Elithorn mazes; PA=pattern assembly; 2D=2D drawing; CS=cross-sections; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; RMSEA = Root Mean Square Error of Approximation; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; SRMR = Standardized Root Mean Square Residual; **=p<0.01.

b) Two factors

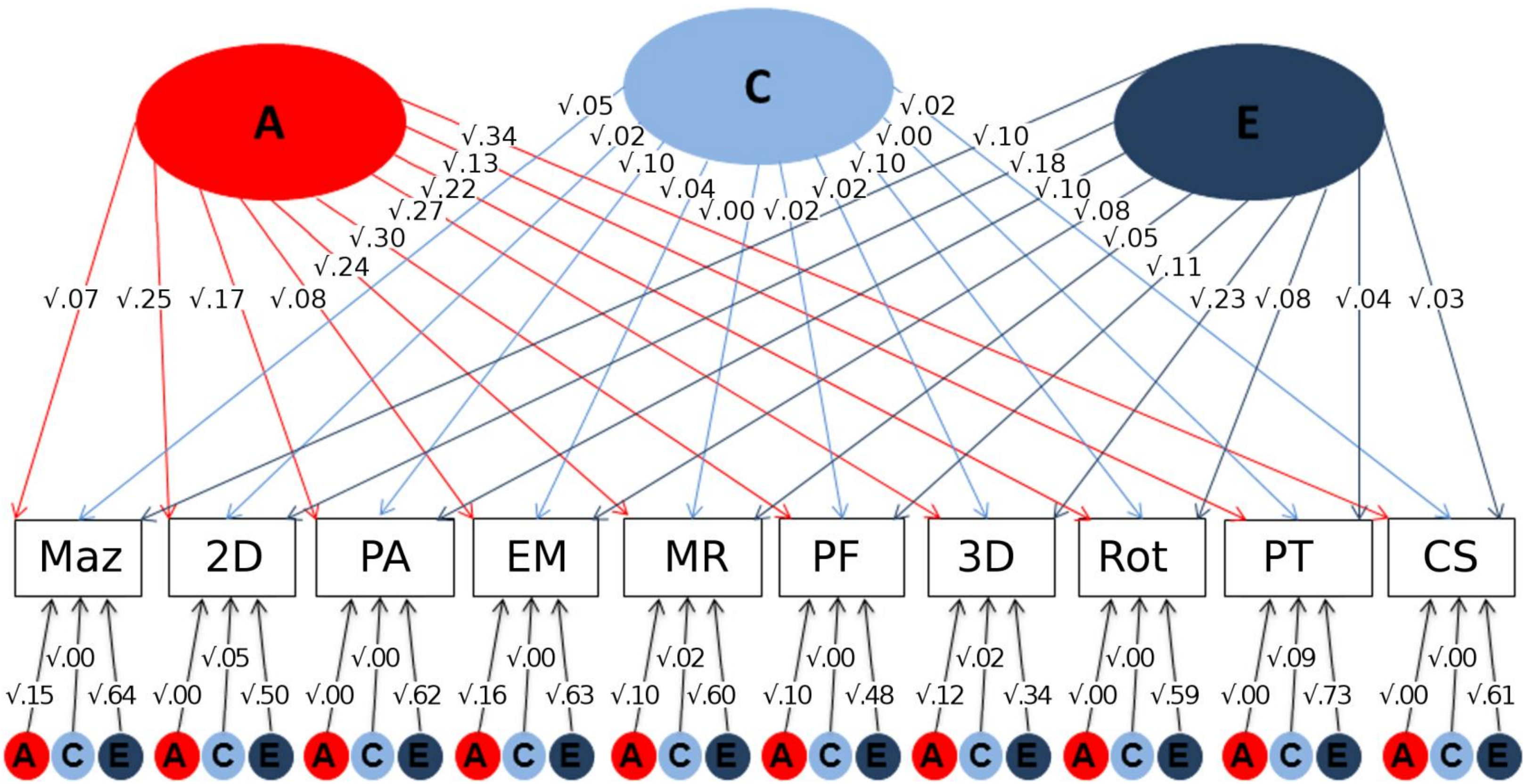


	AIC	BIC	X^2	RMSEA	CFI	TLI	SRMR	*r*
2 Factor Model	53241.72	53451.18	132.77 **	0.03	.98	.98	0.02	0.79

Confirmatory factor analysis loadings and fit indices with two non-spatial and ten spatial tests loading onto two factors. Factor one (*g*): V=verbal ability (Mill Hill vocabulary); NV=non-verbal ability (Raven's matrices). Factor two (spatial ability): Maz=mazes; PT=perspective-taking; Rot=mental rotation; 3D=3D drawing; PF=paper folding; MR=mechanical reasoning; EM=Elithorn mazes; PA=pattern assembly; 2D=2D drawing; CS=cross-sections; AIC = Akaike Information Criterion; BIC = Bayesian Information Criterion; RMSEA = Root Mean Square Error of Approximation; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; SRMR = Standardized Root Mean Square Residual; *r*=correlation between factors; **=p<0.01.

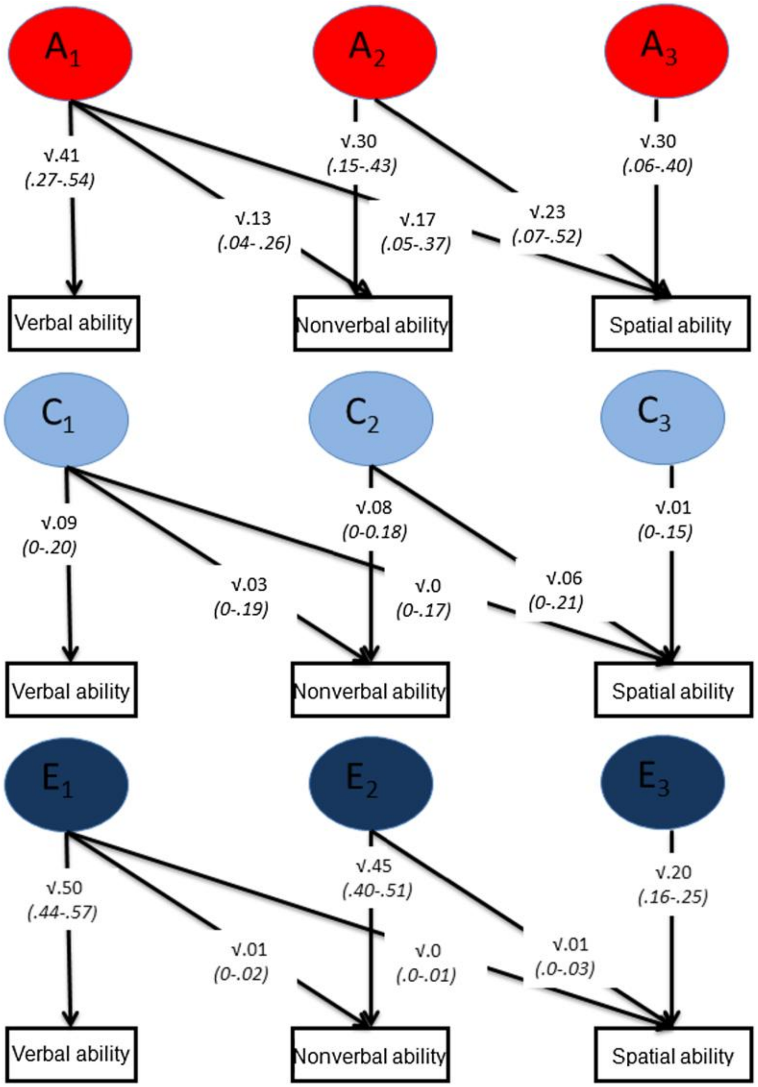
The two-factor model is a better fit (indicated by the lower AIC and BIC indices, higher CFI and TLI and lower RMSEA and SRMR).

Figure S3. Independent pathway model after correction for general intelligence.



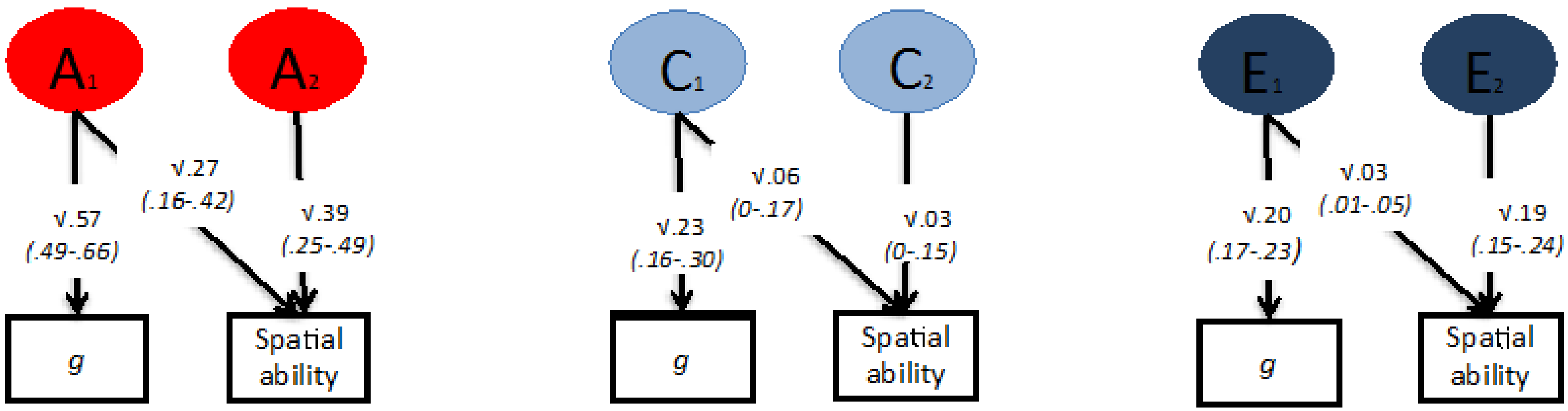
Independent pathway model presenting the standardized squared path estimates for the 10 spatial tests after correction for general intelligence using the regression method. A=additive genetic, C=shared environmental and E=non-shared environmental components of variance. Spa=overall spatial ability; Maz=mazes; 2D=2D drawing; PA=pattern assembly; EM=Elithorn mazes; MR=mechanical reasoning; PF=paper folding; 3D=3D drawing; Rot=mental rotation; PT=perspective-taking; CS=cross-sections.

Figure S4. Trivariate Cholesky decomposition for verbal ability, non-verbal ability and spatial ability.



Standardized path estimates (95% confidence intervals in parentheses) for genetic (A), shared environmental (C) and non-shared environmental influences (E) on spatial ability, shared with and independent from verbal and non-verbal ability.

Figure S5. Bivariate Cholesky decomposition for intelligence (ages 7-16) and spatial ability.



Standardized path estimates (95% confidence intervals in parentheses) for genetic (A), shared environmental (C) and non-shared environmental influences (E) on spatial ability, shared with and independent from *g* (intelligence).

Table S1. Mean scores (standard deviations) for ten spatial tests. N=sample size after exclusions (one randomly selected twin per pair); MZ=monozygotic; DZ=dizygotic; m=male; f=female; os=opposite sex. ANOVA analyses tested the effect of sex and zygosity: results = F statistic; ** = p<.01; R^2 = proportion of variance explained by sex, zygosity and their interaction.

Subject	N	Whole Sample	Male	Female	MZm	DZm	MZf	DZf	Dzos	Sex	Zyg	Sex x Zyg	R^2
Mazes	1213	5.77 (1.85)	6.23 (1.80)	5.49 (1.82)	6.39 (1.79)	6.19 (1.88)	5.42 (1.86)	5.50 (1.78)	5.83 (1.80)	40.10**	0.10	0.28	0.04
2D drawing	1345	3.58 (1.07)	3.90 (0.92)	3.40 (1.10)	3.98 (0.91)	3.95 (0.90)	3.33 (1.12)	3.47 (1.14)	3.60 (1.0)	56.49**	0.09	1.91	0.06
Pattern assembly	1300	6.62 (3.37)	7.43 (3.36)	6.15 (3.28)	7.33 (3.18)	7.33 (3.57)	5.91 (3.19)	6.14 (3.40)	7.03 (3.33)	25.12**	2.45	0.01	0.03
Elithorn maze	1160	7.68 (1.45)	8.26 (1.19)	7.32 (1.48)	8.34 (0.95)	8.21 (1.43)	7.19 (1.46)	7.30 (1.59)	7.87 (1.31)	108.07**	0.76	1.99	0.11
Mechanical reasoning	1314	9.28 (2.53)	10.28 (2.46)	8.67 (2.37)	10.38 (2.55)	10.43 (2.51)	8.66 (2.33)	8.57 (2.44)	9.38 (2.41)	123.71**	0.62	0.32	0.13
Paper folding	1262	8.02 (3.81)	8.70 (3.77)	7.63 (3.77)	8.64 (3.98)	8.86 (3.70)	7.37 (3.65)	7.69 (3.83)	8.25 (3.78)	20.99**	2.26	0.78	0.02
3D drawing	1211	2.95 (1.80)	3.52 (1.79)	2.62 (1.71)	3.70 (1.73)	3.51 (1.82)	2.57 (1.68)	2.64 (1.81)	2.97 (1.76)	66.49**	0.74	2.40	0.07
Mental rotation	1202	8.20 (4.03)	9.19 (3.77)	7.62 (4.06)	9.24 (3.54)	9.40 (4.10)	7.29 (4.06)	7.60 (4.41)	8.54 (3.84)	30.36**	2.10	0.01	0.04
Perspective taking	1222	4.41 (3.84)	5.87 (4.22)	3.54 (3.31)	6.11 (4.15)	5.86 (4.38)	3.56 (3.23)	3.30 (3.28)	4.61 (3.90)	81.66**	0.09	0.48	0.09
Cross section	1367	6.49 (3.58)	7.47 (3.61)	5.92 (3.44)	7.61 (3.68)	7.68 (3.46)	5.50 (3.59)	6.27 (3.23)	6.60 (3.50)	50.45**	1.80	2.94	0.05

Mean scores (standard deviations) from five sex and zygosity groups. ANOVA results indicate that sex and zygosity together explain between 2% and 13% of the variance in each spatial test.

Table S2. a) Correlation matrix and b) residual correlation matrix for ten spatial tests.

a) Correlation matrix

Correlations	Cross sections	2D drawing	Pattern assembly	Elithorn maze	Mechanical reasoning	Paper folding	3D drawing	Mental rotation	Perspective taking	Mazes
Cross sections	1									
2D drawing	0.41	1								
Pattern assembly	0.35	0.43	1							
Elithorn maze	0.25	0.34	0.30	1						
Mechanical reasoning	0.42	0.35	0.32	0.28	1					
Paper folding	0.44	0.48	0.44	0.31	0.43	1				
3D drawing	0.42	0.54	0.41	0.36	0.36	0.49	1			
Mental rotation	0.37	0.42	0.46	0.39	0.36	0.43	0.45	1		
Perspective taking	0.32	0.32	0.26	0.20	0.30	0.30	0.34	0.32	1	
Mazes	0.22	0.28	0.29	0.26	0.25	0.30	0.32	0.31	0.21	1

All correlations significant at $p < 0.01$.

b) Reproduced and residual correlation matrices

Reproduced correlations	Cross sections	2D drawing	Pattern assembly	Elithorn maze	Mechanical reasoning	Paper folding	3D drawing	Mental rotation	Perspective taking	Mazes
Cross sections	.36a									
2D drawing	0.41	.47a								
Pattern assembly	0.37	0.42	.38a							
Elithorn maze	0.30	0.34	0.31	.25a						
Mechanical reasoning	0.34	0.39	0.35	0.28	.32a					
Paper folding	0.42	0.47	0.43	0.34	0.39	.48a				
3D drawing	0.42	0.48	0.43	0.35	0.40	0.49	.50a			
Mental rotation	0.40	0.45	0.41	0.33	0.38	0.46	0.47	.44a		
Perspective taking	0.28	0.32	0.29	0.23	0.27	0.32	0.33	0.31	.22a	
Mazes	0.27	0.30	0.27	0.22	0.25	0.31	0.31	0.29	0.21	.20a
Residuals										
Cross sections										
2D drawing	0.00									
Pattern assembly	-0.02	0.01								
Elithorn maze	-0.05	0.01	-0.01							
Mechanical reasoning	0.08	-0.04	-0.03	0.00						
Paper folding	0.03	0.00	0.02	-0.03	0.03					
3D drawing	-0.01	0.05	-0.03	0.01	-0.04	0.01				
Mental rotation	-0.03	-0.03	0.05	0.07	-0.02	-0.03	-0.02			
Perspective taking	0.04	0.00	-0.03	-0.04	0.04	-0.03	0.01	0.01		
Mazes	-0.04	-0.02	0.02	0.04	0.00	-0.01	0.00	0.02	0.00	

a - reproduced communalities. Reproduced and residual correlation matrices extracted from exploratory factor analysis. Reproduced correlations are based on the extracted factors; they are similar to the original correlations, indicating that the factor extracted accounts for a large proportion of covariance in these tests. The residual correlations are calculated as the difference between the original and reproduced correlations. The residuals are small in magnitude, confirming that the first (and only) factor accounts for almost all covariance. There are several non-redundant residuals with values greater than 0.05 (presumably reflecting *g*, which was not controlled in this analysis).

Table S3. Sex-limitation model-fitting sub-model comparisons. FullHetACE=full genetic heterogeneity model, rG=Free; HetACE= quantitative heterogeneity model; cFullHetACE=full environmental heterogeneity model, rC=Free; HomACE= homogeneity model (no sex differences at all); ep=estimated parameters; minus2LL= minus 2 log-likelihood; df= degrees of freedom; AIC= Akaike information criterion; diffLL= change in log-likelihood; diffdf= change in degrees of freedom (significant differences are marked in bold).

Spatial ability							
Qualitative genetic differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
FullHetACE	9	4798.23	1785	1228.23	-	-	-
HetACE	8	4798.37	1786	1226.37	0.13	1	0.71
Qualitative environmental differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
cFullHetACE	9	4798.23	1785	1228.23	-	-	-
HetACE	8	4798.37	1786	1226.37	0.13	1	0.71
Quantitative differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
HetACE	8	4798.37	1786	1226.37	-	-	-
HomACE	5	4800.75	1789	1222.75	2.38	3	0.50

Mazes							
Qualitative genetic differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
FullHetACE	9	6739.17	2402	1935.17	-	-	-
HetACE	8	6739.3	2403	1933.3	0.13	1	0.72
Qualitative environmental differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
cFullHetACE	9	6739.19	2402	1935.19	-	-	-
HetACE	8	6739.3	2403	1933.3	0.11	1	0.74
Quantitative differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
HetACE	8	6739.3	2403	1933.3	-	-	-
HomACE	5	6744.91	2406	1932.91	5.61	3	0.13

2D Drawing							
Qualitative genetic differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
FullHetACE	9	7396.63	2674	2048.63	-	-	-
HetACE	8	7396.88	2675	2046.88	0.25	1	0.62
Qualitative environmental differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
cFullHetACE	9	7396.63	2674	2048.63	-	-	-
HetACE	8	7396.88	2675	2046.88	0.25	1	0.62
Quantitative differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
HetACE	8	7396.88	2675	2046.88	-	-	-
HomACE	5	7444.77	2678	2088.77	47.89	3	**0**

Pattern assembly							
Qualitative genetic differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
FullHetACE	9	7199.89	2573	2053.89	-	-	-
HetACE	8	7199.89	2574	2051.89	0	1	1
Qualitative environmental differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
cFullHetACE	9	7199.89	2573	2053.89	-	-	-
HetACE	8	7199.89	2574	2051.89	0	1	1
Quantitative differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
HetACE	8	7199.89	2574	2051.89	-	-	-
HomACE	5	7205.97	2577	2051.97	6.08	3	0.11

Elithorn maze							
Qualitative genetic differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
FullHetACE	9	6427.18	2293	1841.18	-	-	-
HetACE	8	6427.36	2294	1839.36	0.19	1	0.67
Qualitative environmental differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
cFullHetACE	9	6428.67	2293	1842.67	-	-	-
HetACE	8	6427.36	2294	1839.36	-1.31	1	1
Quantitative differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
HetACE	8	6427.36	2294	1839.36	-	-	-
HomACE	5	6440.54	2297	1846.54	13.17	3	**0**

Mechanical reasoning							
Qualitative genetic differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
FullHetACE	9	7229.19	2602	2025.19	-	-	-
HetACE	8	7229.19	2603	2023.19	0	1	1
Qualitative environmental differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
cFullHetACE	9	7229.19	2602	2025.19	-	-	-
HetACE	8	7229.19	2603	2023.19	0	1	1
Quantitative differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
HetACE	8	7229.19	2603	2023.19	-	-	-
HomACE	5	7237.08	2606	2025.08	7.9	3	**0.05**

Paper folding							
Qualitative genetic differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
FullHetACE	9	6949.91	2511	1927.91	-	-	-
HetACE	8	6949.91	2512	1925.91	0	1	1
Qualitative environmental differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
cFullHetACE	9	6949.91	2511	1927.91	-	-	-
HetACE	8	6949.91	2512	1925.91	0	1	1
Quantitative differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
HetACE	8	6949.91	2512	1925.91	-	-	-
HomACE	5	6958.43	2515	1928.43	8.52	3	**0.04**

3D drawing

Qualitative genetic differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
FullHetACE	9	6583.55	2389	1805.55	-	-	-
HetACE	8	6583.55	2390	1803.55	0	1	1
Qualitative environmental differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
cFullHetACE	9	6583.55	2389	1805.55	-	-	-
HetACE	8	6583.55	2390	1803.55	0	1	1
Quantitative differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
HetACE	8	6583.55	2390	1803.55	-	-	-
HomACE	5	6586.89	2393	1800.89	3.34	3	0.34

Mental rotation

Qualitative genetic differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
FullHetACE	9	6715.01	2409	1897.01	-	-	-
HetACE	8	6715.01	2410	1895.01	0	1	1
Qualitative environmental differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
cFullHetACE	9	6715.01	2409	1897.01	-	-	-
HetACE	8	6715.01	2410	1895.01	0	1	1
Quantitative differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
HetACE	8	6715.01	2410	1895.01	-	-	-
HomACE	5	6716.3	2413	1890.3	1.29	3	0.73

Perspective taking							
Qualitative genetic differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
FullHetACE	9	6726.35	2426	1874.35	-	-	-
HetACE	8	6726.35	2427	1872.35	0	1	0.99
Qualitative environmental differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
cFullHetACE	9	6726.35	2426	1874.35	-	-	-
HetACE	8	6726.35	2427	1872.35	0	1	1
Quantitative differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
HetACE	8	6726.35	2427	1872.35	-	-	-
HomACE	5	6826.25	2430	1966.25	99.9	3	**0**

Cross-sections							
Qualitative genetic differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
FullHetACE	9	7524.12	2699	2126.12	-	-	-
HetACE	8	7524.12	2700	2124.12	0	1	1
Qualitative environmental differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
cFullHetACE	9	7524.12	2699	2126.12	-	-	-
HetACE	8	7524.12	2700	2124.12	0	1	1
Quantitative differences:	ep	-2LL	df	AIC	diffLL	diffdf	p
HetACE	8	7524.12	2700	2124.12	-	-	-
HomACE	5	7528.59	2703	2122.59	4.47	3	0.21

Full sex limitation model results show that there were no significant qualitative sex differences in any of the spatial tests (i.e., no different genetic or environmental factors affecting males and females), but there were some significant quantitative sex differences (differences in the magnitude of ACE estimates for males and females). Significant results are indicated in bold. As noted in the text, little confidence can be placed in these differences, as the sex-limitation models are underpowered to detect differences of this small magnitude; nonetheless, separate ACE estimates for males and females are presented for reference in Table S4.

Table S4. Sex-limitation model-fitting results, showing A, C, E, estimates separately for males and females. A=additive genetic; C=shared environmental; E=non-shared environmental proportions of the variance (95% confidence intervals).

Overall spatial ability	**A**	**C**	**E**
Males	0.68 (0.32; 0.85)	0.12 (0; 0.45)	0.20 (0.15; 0.28)
Females	0.65 (0.39; 0.80)	0.09 (0; 0.33)	0.25 (0.20; 0.32)

Mazes	**A**	**C**	**E**
Males	0.39 (0.17; 0.58)	0.08 (0; 0.23)	0.53 (0.42; 0.67)
Females	0.16 (0; 0.41)	0.15 (0; 0.35)	0.69 (0.59; 0.80)

2D drawing	**A**	**C**	**E**
Males	0.40 (0.09; 0.65)	0.17 (0.00; 0.42)	0.43 (0.34; 0.55)
Females	0.29 (0.02; 0.50)	0.12 (0.00; 0.35)	0.58 (0.50; 0.68)

Pattern assembly	**A**	**C**	**E**
Males	0.35 (0.08; 0.57)	0.10 (0; 0.28)	0.55 (0.43; 0.70)
Females	0.40 (0.15; 0.49)	0.00 (0; 0.20)	0.60 (0.51; 0.70)

Elithorn maze	**A**	**C**	**E**
Males	0.38 (0.12; 0.54)	0.03 (0.00; 0.19)	0.59 (0.46; 0.77)
Females	0.34 (0.03; 0.50)	0.06 (0.00; 0.32)	0.59 (0.50; 0.71)

Mechanical reasoning	A	C	E
Males	0.03 (0.00; 0.35)	0.45 (015; 0.56)	0.52 (0.43; 0.61)
Females	0.41 (0.20; 0.52)	0.05 (0.00; 0.22)	0.54 (0.46; 0.64)

Paper folding	A	C	E
Males	0.04 (0; 0.41)	0.48 (0.13; 0.60)	0.48 (0.39; 0.59)
Females	0.53 (0.35; 0.62)	0.02 (0.00; 0.17)	0.45 (0.38; 0.53)

3D drawing	A	C	E
Males	0.40 (0.06; 0.68)	0.20 (0; 0.50)	0.40 (0.31; 0.51)
Females	0.58 (0.41; 0.65)	0.00 (0; 0.14)	0.42 (0.35; 0.50)

Mental rotation	A	C	E
Males	0.25 (0; 0.57)	0.21 (0; 0.22)	0.54 (0.42; 0.69)
Females	0.39 (0.09; 0.53)	0.07 (0; 0.32)	0.54 (0.46; 0.64)

Perspective taking	A	C	E
Males	0.28 (0; 0.42)	0.00 (0.00; 0.31)	0.72 (0.58; 0.88)
Females	0.32 (0; 0.46)	0.04 (0.00; 0.33)	0.64 (0.54; 0.76)

Cross-sections	A	C	E
Males	0.01 (0.00; 0.38)	0.40 (0.08; 0.50)	0.48 (0.48; 0.69)
Females	0.21 (0.00; 0.41)	0.17 (0.03; 0.37)	0.61 (0.53; 0.71)

A few quantitative sex differences emerged for individual spatial ability tests (Table S3); however, the differences were small when examining the ACE estimates for males and females separately. Even with over 1300 twin pairs, the sample size is not sufficiently large for sex-limitation models to reliably detect quantitative and qualitative sex differences of this small magnitude, so little confidence can be placed in these differences, as is evident from the large confidence intervals around the estimates when calculated for males and females separately.

Table S5. Model-fitting results for univariate analyses of spatial ability tests, with twin intraclass correlations (N=complete twin pairs). A=additive genetic; C=shared environmental; E=non-shared environmental proportions of the variance (95% confidence intervals).

				Twin intraclass correlations	
	A	**C**	**E**	**MZ**	**DZ**
Spatial ability	0.69 (0.50; 0.80)	0.08 (0.00; 0.25)	0.23 (0.20; 0.29)	0.77 (0.71; 0.82) (N=229)	0.41 (0.31; 0.50) (N=305)
Mazes	0.35 (0.11; 0.44)	0.01 (0.00; 0.20)	0.64 (0.56; 0.73)	0.36 (0.27; 0.44) (N=384)	0.17 (0.08; 0.26) (N=494)
2D drawing	0.33 (0.13; 0.51)	0.12 (0.00; 0.28)	0.55 (0.48; 0.62)	0.45 (0.38; 0.53) (N=432)	0.26 (0.18; 0.34) (N=574)
Pattern assembly	0.42 (0.22; 0.49)	0.00 (0.00; 0.15)	0.58 (0.51; 0.66)	0.40 (0.32; 0.48) N=412	0.20 (0.12; 0.28) (N=540)
Elithorn maze	0.39 (0.21; 0.47)	0.00 (0.00; 0.13)	0.61 (0.53; 0.70)	0.39 (0.29; 0.47) (N=342)	0.16 (0.07; 0.25) (N=456)
Mechanical reasoning	0.41 (0.20; 0.53)	0.06 (0.00; 0.22)	0.53 (0.47; 0.61)	0.48 (0.40; 0.55) (N=427)	0.26 (0.18; 0.33) (N=557)
Paper folding	0.53 (0.38; 0.59)	0.00 (0.00; 0.12)	0.47 (0.41; 0.53)	0.54 (0.46; 0.60) (N=396)	0.24 (0.16; 0.32) (N=525)
3D drawing	0.59 (0.43; ;0.65)	0.00 (0.00; 0.13)	0.42 (0.35; 0.47)	0.57 (0.50; 0.64) (N=385)	0.28 (0.19; 0.36) (N=479)
Mental rotation	0.36 (0.14; 0.53)	0.10 (0.00; 0.27)	0.54 (0.47; 0.63)	0.44 (0.36; 0.52) (N=376)	0.27 (0.18; 0.35) (N=492)
Perspective taking	0.33 (0.10; 0.41)	0.00 (0.00; 017)	0.67 (0.59; 0.76)	0.31 (0.21; 0.39) (N=391)	0.18 (0.09; 0.26) (N=501)
Cross sections	0.18 (0.00; 0.38)	0.22 (0.05; 0.37)	0.60 (0.53; 0.68)	0.40 (0.32; 0.48) (N=428)	0.30 (0.22; 0.38) (N=574)

General spatial ability was substantially heritable (69%), with a small proportion of variance explained by shared environmental factors (8%) and the rest of the variance explained by non-shared environmental factors (23%). Heritability was lower for the individual 10 tests, ranging from 18% to 59%.

Table S6. Model fit statistics a) comparing Cholesky decomposition to Independent pathway model and Common pathway model; b) comparing Independent pathway model to Common pathway model. CholACE= Cholesky model; IPACE= Independent pathway model; CPACE= Common pathway model; ep=estimated parameters; minus2LL= minus 2 log-likelihood; df= degrees of freedom; AIC= Akaike information criterion; diffLL= change in log-likelihood; diffdf= change in degrees of freedom.

a)

base	comparison	ep	minus2LL	df	AIC	diffLL	diffdf	p
CholACE	<NA>	175	62189.93	24893	12403.93	NA	NA	NA
CholACE	IPACE	70	62306.29	24998	12310.29	116.3653	105	0.21
CholACE	CPACE	53	62405.58	25016	12373.58	215.6527	123	0.00

b)

base	comparison	ep	minus2LL	df	AIC	diffLL	diffdf	p
IPACE	CPACE	53	62405.58	25016	12373.58	99.28742	18	0.00

a) Comparing the Cholesky ACE model and the independent pathway model shows that there is no significant deterioration in fit (indicated by the p-value). Comparing the Cholesky ACE model and the common pathway model shows a significant deterioration in fit.
b) Comparing the independent pathway model and common pathway model indicates that the former fits the data better than the latter (a significant deterioration of fit is indicated by the p-value).

Table S7. Independent pathway model presenting the standardized squared path estimates (95% CI). Cp=common path; SP= specific path; A=additive genetic; C=common environmental; E=non-shared environmental; 1=mazes; 2=2D drawing, 3=Pattern assembly, 4=Elithorn maze, 5=Mechanical reasoning, 6=Paper folding, 7=3D drawing, 8=Mental rotation, 9=Perspective taking, 10=Cross-sections. a) 10 spatial tests; b) 10 spatial tests after correction for general intelligence using the regression method.

a) 10 spatial tests

CpA2[1,1]	0.25 (0.15; 0.30)	CpC2[1,1]	0.01 (0.00; 0.05)	CpE2[1,1]	0.04 (0.01; 0.14)
CpA2[2,2]	0.45 (0.31; 0.52)	CpC2[2,2]	0.01 (0.00; 0.11)	CpE2[2,2]	0.07 (0.03; 0.18)
CpA2[3,3]	0.36 (0.30; 0.50)	CpC2[3,3]	0.00 (0.00; 0.08)	CpE2[3,3]	0.11 (0.01; 0.19)
CpA2[4,4]	0.26 (0.16; 0.32)	CpC2[4,4]	0.00 (0.00; 0.05)	CpE2[4,4]	0.01 (0.00; 0.09)
CpA2[5,5]	0.33 (0.20; 0.43)	CpC2[5,5]	0.05 (0.00; 0.17)	CpE2[5,5]	0.02 (0.00; 0.08)
CpA2[6,6]	0.40 (0.30; 0.53)	CpC2[6,6]	0.04 (0.00; 0.16)	CpE2[6,6]	0.10 (0.02; 0.16)
CpA2[7,7]	0.50 (0.32; 0.58)	CpC2[7,7]	0.01 (0.00; 0.10)	CpE2[7,7]	0.07 (0.03; 0.27)
CpA2[8,8]	0.41 (0.34; 0.55)	CpC2[8,8]	0.00 (0.00; 0.08)	CpE2[8,8]	0.10 (0.01; 0.17)
CpA2[9,9]	0.22 (0.12; 0.27)	CpC2[9,9]	0.01 (0.00; 0.08)	CpE2[9,9]	0.02 (0.00; 0.12)
CpA2[10,10]	0.27 (0.12; 0.40)	CpC2[10,10]	0.15 (0.05; 0.31)	CpE2[10,10]	0.05 (0.01; 0.12)

SpA2[1,1]	0.12 (0.00; 0.22)	SpC2[1,1]	0.00 (0.00; 0.15)	SpE2[1,1]	0.59 (0.49; 0.67)
SpA2[2,2]	0.00 (0.00; 0.09)	SpC2[2,2]	0.02 (0.00; 0.08)	SpE2[2,2]	0.45 (0.36; 0.49)
SpA2[3,3]	0.03 (0.00; 0.08)	SpC2[3,3]	0.00 (0.00; 0.05)	SpE2[3,3]	0.50 (0.43; 0.59)
SpA2[4,4]	0.14 (0.00; 0.23)	SpC2[4,4]	0.01 (0.00; 0.15)	SpE2[4,4]	0.57 (0.50; 0.66)
SpA2[5,5]	0.08 (0.00; 0.16)	SpC2[5,5]	0.02 (0.00; 0.11)	SpE2[5,5]	0.49 (0.44; 0.56)
SpA2[6,6]	0.09 (0.00; 0.13)	SpC2[6,6]	0.00 (0.00; 0.09)	SpE2[6,6]	0.38 (0.33; 0.48)
SpA2[7,7]	0.08 (0.00; 0.17)	SpC2[7,7]	0.00 (0.00; 0.11)	SpE2[7,7]	0.33 (0.19; 0.40)
SpA2[8,8]	0.01 (0.00; 0.10)	SpC2[8,8]	0.03 (0.00; 0.08)	SpE2[8,8]	0.45 (0.38; 0.54)
SpA2[9,9]	0.10 (0.00; 0.22)	SpC2[9,9]	0.02 (0.00; 0.13)	SpE2[9,9]	0.63 (0.53; 0.71)
SpA2[10,10]	0.00 (0.00; 0.04)	SpC2[10,10]	0.00 (0.00; 0.03)	SpE2[10,10]	0.53 (0.46; 0.57)

b) 10 spatial tests after correction for *g*

CpA2[1,1]	0.07 (0.02; 0.15)
CpA2[2,2]	0.25 (0.16; 0.32)
CpA2[3,3]	0.17 (0.06; 0.31)
CpA2[4,4]	0.08 (0.03; 0.16)
CpA2[5,5]	0.24 (0.18; 0.29)
CpA2[6,6]	0.30 (0.19; 0.37)
CpA2[7,7]	0.27 (0.17; 0.35)
CpA2[8,8]	0.22 (0.09; 0.36)
CpA2[9,9]	0.13 (0.08; 0.19)
CpA2[10,10]	0.34 (0.17; 0.41)

CpC2[1,1]	0.05 (0.00; 0.11)
CpC2[2,2]	0.02 (0.00; 0.10)
CpC2[3,3]	0.10 (0.00; 0.21)
CpC2[4,4]	0.04 (0.00; 0.10)
CpC2[5,5]	0.00 (0.00; 0.04)
CpC2[6,6]	0.02 (0.00; 0.10)
CpC2[7,7]	0.02 (0.00; 0.10)
CpC2[8,8]	0.10 (0.00; 0.23)
CpC2[9,9]	0.00 (0.00; 0.05)
CpC2[10,10]	0.02 (0.00; 0.16)

CpE2[1,1]	0.10 (0.05; 0.15)
CpE2[2,2]	0.18 (0.12; 0.25)
CpE2[3,3]	0.10 (0.05; 0.17)
CpE2[4,4]	0.08 (0.04; 0.14)
CpE2[5,5]	0.05 (0.02; 0.08)
CpE2[6,6]	0.11 (0.07; 0.17)
CpE2[7,7]	0.23 (0.16; 0.31)
CpE2[8,8]	0.08 (0.04; 0.14)
CpE2[9,9]	0.04 (0.01; 0.08)
CpE2[10,10]	0.03 (0.01; 0.07)

SpA2[1,1]	0.15 (0.00; 0.22)
SpA2[2,2]	0.00 (0.00; 0.10)
SpA2[3,3]	0.00 (0.00; 0.08)
SpA2[4,4]	0.16 (0.00; 0.25)
SpA2[5,5]	0.10 (0.00; 0.18)
SpA2[6,6]	0.10 (0.00; 0.15)
SpA2[7,7]	0.12 (0.00; 0.20)
SpA2[8,8]	0.00 (0.00; 0.06)
SpA2[9,9]	0.00 (0.00; 0.19)
SpA2[10,10]	0.00 (0.00; 0.05)

SpC2[1,1]	0.00 (0.00; 0.13)
SpC2[2,2]	0.05 (0.00; 0.09)
SpC2[3,3]	0.00 (0.00; 0.05)
SpC2[4,4]	0.00 (0.00; 0.16)
SpC2[5,5]	0.02 (0.00; 0.13)
SpC2[6,6]	0.00 (0.00; 0.09)
SpC2[7,7]	0.02 (0.00; 0.13)
SpC2[8,8]	0.00 (0.00; 0.06)
SpC2[9,9]	0.09 (0.00; 0.15)
SpC2[10,10]	0.00 (0.00; 0.05)

SpE2[1,1]	0.64 (0.56; 0.72)
SpE2[2,2]	0.50 (0.44; 0.56)
SpE2[3,3]	0.62 (0.55; 0.66)
SpE2[4,4]	0.63 (0.55; 0.73)
SpE2[5,5]	0.60 (0.53; 0.67)
SpE2[6,6]	0.48 (0.42; 0.55)
SpE2[7,7]	0.34 (0.27; 0.42)
SpE2[8,8]	0.59 (0.53; 0.63)
SpE2[9,9]	0.73 (0.63; 0.79)
SpE2[10,10]	0.61 (0.56; 0.67)

Standardized path estimates (following from Figure 4), with 95% confidence intervals, for the independent pathway model.

a) All spatial tests loaded substantially on the common A factor, with no significant specific genetic influence remaining after controlling for the common genetic factor. On average, the common A factor accounted for 85% of the heritabilities of the 10 spatial tests (for example the heritability of the Mazes task was 37% (the sum of common path, .25, and the specific path, .12), so the proportion of heritability accounted for by the common factor is .25/.37=68%). The spatial tests are differentiated by E factors, which indicate test-specific environmental influences and measurement error specific to each test.

b) These results show the same analysis after correcting the spatial scores for *g*. A common genetic factor still explained most of the heritability across the 10 tests, although loadings on the common A factor were reduced by about one third.

Table S8. Common pathway model presenting the standardized path estimates. A- additive genetic, C- shared environmental and E- non-shared environmental components of variance. a) 10 spatial tests; b) 10 spatial tests when corrected for intelligence using the regression method.

a) 10 spatial tests

Spatial ability (Latent Factor)	
A	0.80 (0.64-0.89)
C	0.06 (0-0.21)
E	0.14 (0.11-0.18)

Loadings to Spatial ability factor	
Cross sections	0.61 (0.68-0.64)
2D drawing	0.73 (0.71-0.75)
Pattern assembly	0.66 (0.64-0.69)
Elithorn maze	0.50 (0.46-0.69)
Mechanical reasoning	0.62 (0.59-0.63)
Paper folding	0.72 (0.70-0.75)
3D drawing	0.77 (0.75-0.79)
Mental rotation	0.70 (0.68 -0.72)
Perspective taking	0.50 (0.47-0.54)
Mazes	0.51 (0.47-0.54)

Residual variance	**A**	**C**	**E**
Cross sections	0.00 (0-0.15)	0.09 (0-0.13)	0.54 (0.47-0.59)
2D drawing	0.00 (0-0.06)	0.02 (0-0.05)	0.44 (0.40 -0.49)
Pattern assembly	0.04 (0-0.09)	0.00 (0-0.06)	0.52 (0.46-0.58)
Elithorn maze	0.16 (0-0.25)	0.02 (0-0.16)	0.57 (0.50-0.66)
Mechanical reasoning	0.09 (0-0.18)	0.04 (0-0.14)	0.48 (0.43-0.55)
Paper folding	0.09 (0-0.13)	0.00 (0-0.07)	0.39 (0.34-0.44)
3D drawing	0.09 (0-0.13)	0.04 (0-0.08)	0.32 (0.28-0.38)
Mental rotation	0.01 (0-0.10)	0.04 (0-0.08)	0.46 (0.40-0.51)
Perspective taking	0.10 (0-0.19)	0.02 (0-0.13)	0.63 (0.55-0.71)
Mazes	0.15 (0-0.22)	0.00 (0-0.14)	0.59 (0.52-0.67)

b) 10 spatial tests after correction for *g*

Spatial ability (Latent factor)	
A	0.57 (0.35-0.75)
C	0.12 (0-0.31)
E	0.31 (0.25-0.38)

Loadings to Spatial ability factor	
Cross sections	0.52 (0.52-0.55)
2D drawing	0.67 (0.64-0.70)
Pattern assembly	0.58 (0.55-0.62)
Elithorn maze	0.44 (0.40-0.48)
Mechanical reasoning	0.51 (0.48 -0.55)
Paper folding	0.65 (0.62-0.68)
3D drawing	0.71 (0.68-0.74)
Mental rotation	0.61 (0.58-0.64)
Perspective taking	0.42 (0.38-0.46)
Mazes	0.44 (0.38-0.47)

Residual variance			
	A	**C**	**E**
Cross sections	0.02 (0-0.18)	0.09 (0-0.15)	0.63 (0.55-0.69)
2D drawing	0.00 (0-0.08)	0.03 (0-0.08)	0.51 (0.46-0.51)
Pattern assembly	0.06 (0-0.13)	0.00 (0-0.08)	0.60 (0.53-0.67)
Elithorn maze	0.16 (0-0.25)	0.01 (0-0.17)	0.64 (0.55-0.74)
Mechanical reasoning	0.12 (0-0.22)	0.03 (0-0.08)	0.59 (0.52-0.67)
Paper folding	0.11 (0-0.16)	0.00 (0-0.10)	0.47 (0.41-0.54)
3D drawing	0.10 (0-0.18)	0.02 (0-0.13)	0.37 (0.31-0.44)
Mental rotation	0.00 (0-0.10)	0.05 (0-0.10)	0.57 (0.51-0.63)
Perspective taking	0.01 (0-0.19)	0.09 (0-0.15)	0.72 (0.63-0.79)
Mazes	0.16 (0-0.24)	0.00 (0-0.15)	0.65 (0.57-0.73)

A, C and E influences on the common latent factor show that the spatial factor is highly heritable. The factor loadings on the latent factor are very substantial. There is some residual variance left after accounting for the latent factor, but this is very small in magnitude, with the A estimates for the residual variance not significant. It should be noted that the independent pathway model fitted the data better than the common pathway model (Supplementary Table S6), but the common pathway model results are presented here for completeness.

Table S9. Genetic, shared environmental and non-shared environmental correlations between 10 spatial tests.

Genetic correlations	Cross sections	2D drawing	Pattern assembly	Elithorn maze	Mechanical reasoning	Paper folding	3D drawing	Mental rotation	Perspective taking	Mazes
Cross sections	1.000									
2D drawing	0.883	1.000								
Pattern assembly	0.770	0.962	1.000							
Elithorn maze	0.860	0.805	0.732	1.000						
Mechanical reasoning	0.876	0.875	0.744	0.766	1.000					
Paper folding	0.966	0.933	0.885	0.862	0.824	1.000				
3D drawing	0.894	0.952	0.912	0.791	0.833	0.920	1.000			
Mental rotation	0.813	0.940	0.950	0.832	0.842	0.897	0.892	1.000		
Perspective taking	0.760	0.913	0.878	0.756	0.834	0.786	0.946	0.884	1.000	
Mazes	0.733	0.730	0.729	0.746	0.784	0.770	0.768	0.884	0.759	1.000

Shared environmental correlations	Cross sections	2D drawing	Pattern assembly	Elithorn maze	Mechanical reasoning	Paper folding	3D drawing	Mental rotation	Perspective taking	Mazes
Cross sections	1.000									
2D drawing	0.786	1.000								
Pattern assembly	0.676	0.864	1.000							
Elithorn maze	-0.059	0.164	0.261	1.000						
Mechanical reasoning	0.824	0.635	0.772	0.067	1.000					
Paper folding	0.795	0.606	0.728	-0.335	0.776	1.000				
3D drawing	0.683	0.826	0.551	0.415	0.428	0.213	1.000			
Mental rotation	0.710	0.740	0.868	0.294	0.738	0.689	0.612	1.000		
Perspective taking	0.537	0.194	-0.114	-0.605	0.067	0.412	0.234	0.056	1.000	
Mazes	0.191	0.702	0.730	0.126	0.282	0.313	0.335	0.382	-0.333	1.000

Non-shared environmental correlations	Cross sections	2D drawing	Pattern assembly	Elithorn maze	Mechanical reasoning	Paper folding	3D drawing	Mental rotation	Perspective taking	Mazes
Cross sections	1.000									
2D drawing	0.137	1.000								
Pattern assembly	0.137	0.113	1.000							
Elithorn maze	-0.005	0.113	0.103	1.000						
Mechanical reasoning	0.101	0.072	0.082	0.027	1.000					
Paper folding	0.095	0.161	0.168	0.033	0.105	1.000				
3D drawing	0.141	0.209	0.139	0.066	0.129	0.224	1.000			
Mental rotation	0.112	0.118	0.175	0.071	0.059	0.180	0.177	1.000		
Perspective taking	0.112	0.061	0.064	0.035	0.069	0.055	0.069	0.120	1.000	
Mazes	0.061	0.090	0.132	0.080	0.020	0.090	0.193	0.075	0.083	1.000

Genetic correlation is an index of pleiotropy: the extent to which the same genetic variants influence multiple traits. Importantly, the genetic correlation is estimated independently of the heritabilities of the traits; that is, the genetic correlation between the traits could be high even if the heritabilities of both traits were low. A shared environmental correlation of 1.0 indicates that the same environmental factors that make twins similar on one trait also make twins similar on another trait. Likewise, for non-shared environment (which is not shared between individuals, but may influence multiple traits for each individual), a correlation of zero indicates that completely different non-shared environmental influences affect the two traits. The results of the multivariate analyses shows that genetic correlations between spatial tests is very high, indicating that to a large extent the performance on these spatial tests is influence by the same genetic factors.

Table S10.. Summary of the development of the gamified battery (King's Challenge): a) Feasibility studies; b) TEDS pilot study.

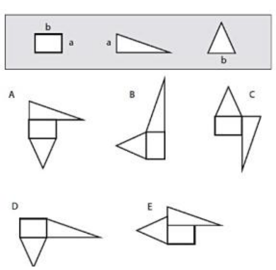
The "King's Challenge" game was constructed after conducting a literature review of the many measures used to test spatial ability, assembling a large variety of measures to test each of the putative components of this cognitive domain. We conducted several feasibility and pilot studies, modifying existing tests and developing some new ones as needed. We started with a paper-and-pencil battery including 27 different tests, and after multiple stages of feasibility and pilot testing (mostly conducted online) ultimately reduced the battery to 10 tests, selected according to the psychometric properties and test-retest reliability of each measure. Here we present:

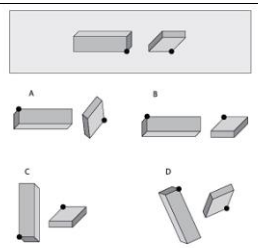
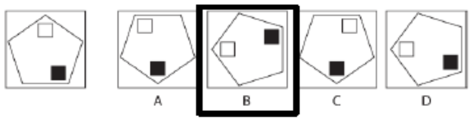
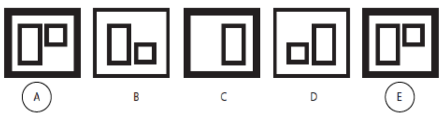
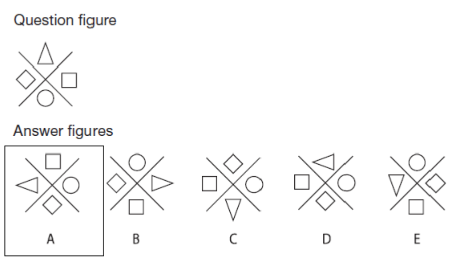
a) the results of two feasibility studies: feasibility 1- the initial paper-and-pencil battery, in which participants were tested in person and were subject to test-level time limits as described in the table; feasibility 2- the first battery administered online (with item-level time limits), from which initial test-retest correlations were obtained (with a 1-week interval between test and retest);

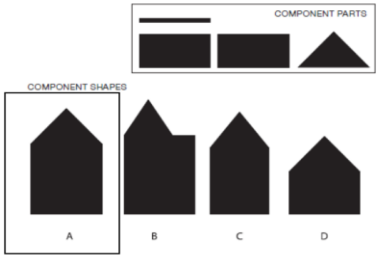
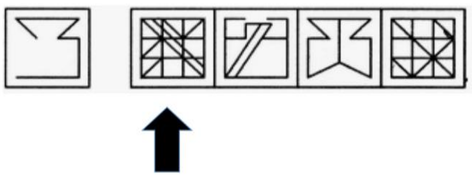
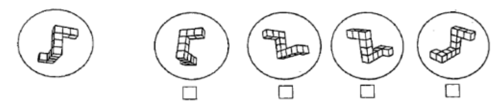
b) the results of the final stage prior to "gamification": a TEDS pilot study with the 10 selected tests. For the latter pilot study, siblings of the TEDS twins were recruited and final test-retest correlations obtained (with a 2-week interval between test and retest).

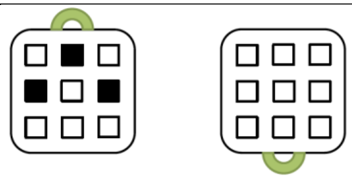
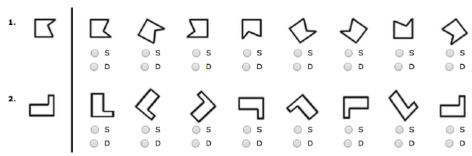
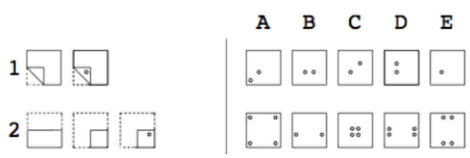
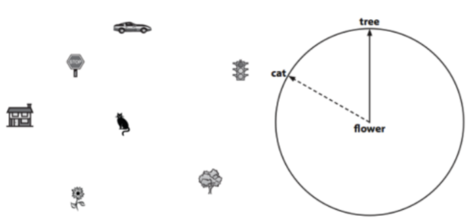
Following the final pilot study, the "gamified" battery was developed. The actual test items were administered in a format identical to those in the final pilot study, but the tests themselves were embedded into an overarching game narrative to encourage participation. This final battery was administered to a large twin sample as described in the manuscript.

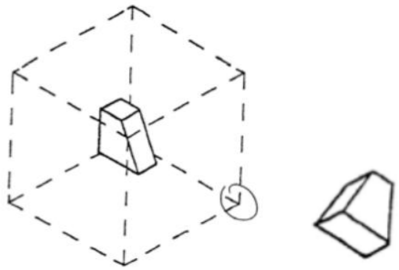
a) Feasibility studies

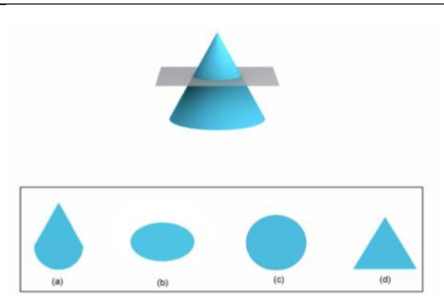
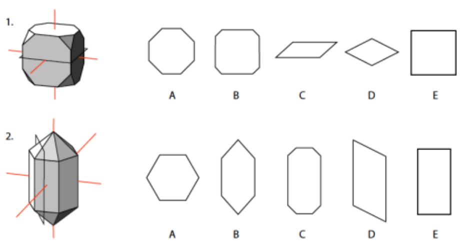
TEST	DESCRIPTION for administration in feasibility 1	REASON FOR KEEPING/DROPPING and ADJUSTMENTS during following stages / Feasibility 2 results	SOURCE
1. Pattern Assembly (1) 	Participants are asked to decide which option (A - E) is made up of the parts presented in the grey box at the top. The test includes **20 items** and participants are allowed **10 minutes** to complete the test.	Included in the second feasibility study (online). The test produced normal distribution and reasonable test-retest reliability. Test-retest (cleaned, standardised): r=0.59, N=40, p<0.001. **Adapted version included in the gamified test (the King's Challenge)**.	Spatial reasoning section 1 in the "How2become" booklet (https://www.how2become.com/testing/spatial-reasoning-tests/)(originally 40 items)
2. 3D rotation (1)	Participants are presented with a pair of three-dimensional objects; one of the corners of each object is marked with a black dot. Participants are asked to imagine which one of the 4 options (A - D) would reflect what the pair of objects would	This task produced a ceiling effect and was dropped after first feasibility study.	Spatial reasoning section 1 in the "How2become" booklet (https://www.how2

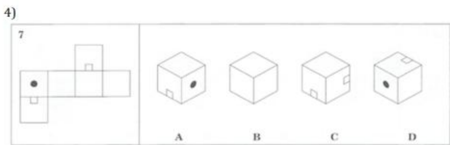
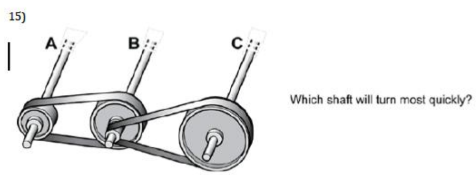
	look like if they were both rotated by the same amount. Participants have **10 minutes** to complete **20 questions**.		become.com/testing/spatial-reasoning-tests/)(originally 40 items)
3. 2D rotation (1) 	Participants are asked to identify which one of the 4 options (A - D) is the same 2D object as the question figure on the left, but rotated. The test includes **20 items** and participants are allowed **10 minutes** to complete them all.	Dropped after the first feasibility study as another task assessing 2D rotation (task 5 below) performed better in terms of distribution and internal reliability.	Spatial reasoning section 1 in the "How2become" booklet (https://www.how2become.com/testing/spatial-reasoning-tests/)
4. Identical shapes 	Participants are asked to identify which two 2D objects (A - E) are identical. The test includes **20 items** and participants have **4 minutes** to complete it.	Dropped after the first feasibility study: too easy, highly skewed distribution.	Spatial reasoning section 1 in the "How2become" booklet (https://www.how2become.com/testing/spatial-reasoning-tests/)
5. 2D rotation (2) 	Participants are asked to identify which one of the answer figures (A - E) is the same object as in the question figure, but rotated. Participants have **7 minutes** to complete **19 items**.	Included in the second feasibility study (online). Produced a normal distribution and good test-retest reliability. Test-retest (cleaned, standardised) *r*=0.73, N=43, p<0.001. **Adapted version included in the gamified test (the King's Challenge).**	Spatial reasoning section 1 in the "How2become" booklet (https://www.how2become.com/testing/spatial-reasoning-tests/)
6.Pattern assembly (2)	Participants are asked to identify which one of the component shapes (A - D) is made from the component parts displayed in the rectangular box at the top. The test includes **20 items** and participants have **7 minutes** to complete it.	Included in the second feasibility study (online). The test produced a good distribution but very poor test-retest reliability. Test-retest (cleaned, standardised): *r*=0.26, N=44, p=0.08. The other pattern assembly test (task 1), showed much higher reliability, so was retained instead.	Spatial reasoning section 1 in the "How2become" booklet (https://www.how2become.com/testing/spatial-reasoning-tests/)

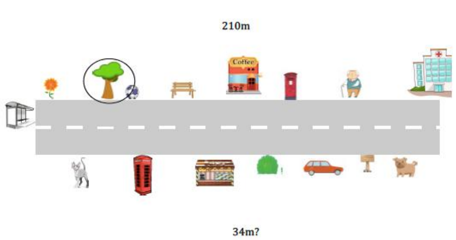
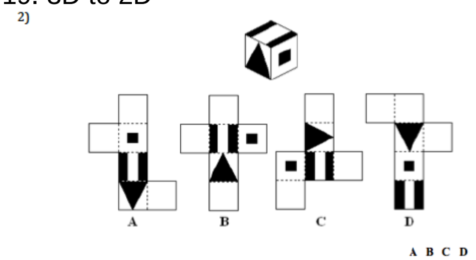
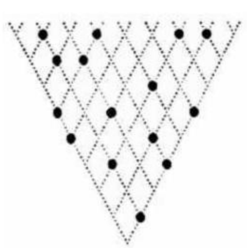
			
7. Embedded figures 	Participants are asked to identify which one of the 4 figures presented on the right-hand side of the page includes the question figure on the left-hand side embedded in its pattern. Participants are given **8 minutes** to complete **25 questions**.	Kept for the second feasibility study (online) including the same 25 items, but subsequently dropped due to relatively poor test-retest reliability and other psychometric properties (other scanning tasks had better psychometric properties): Test-retest (cleaned, standardised; r=0.50, N=48, p<0.001.	www.indiabix.com/non-verbal-reasoning/embedded-images
8. 3D mental rotation (2) 	Participants are asked to identify which 2 options (out of the 4 presented on the right-hand side) are rotated versions of the question figure. Only 2 options are correct at all times. The test is divided into **2 parts** and each part includes **10 questions**. Participants have **3 minutes** to complete each part.	An adapted version was retained for the second feasibility study (online), with only one correct answer per item and two incorrect options. Participants discontinued from the test after 4 consecutive incorrect responses. This was subsequently dropped due to very low test-retest reliability. Test-retest (cleaned, standardised): r=0.29, N=34, p=0.092.	Shepard &Metzler (Shepard, R and Metzler. J. "Mental rotation of three dimensional objects." *Science* 1971. 171(972):701-3 Adapted by S.G. Vanderberg, University of Colorado, July 15, 1971; Revised instructions by H. Crawford, University of Wyoming, September, 1979; Images digitalized and reprinted by Susanna Douglas, University of Texas, March 1996
9. 2D mental rotation (3) –AKA suitcase task	The task requires participants to mentally rotate the image on the left-hand side, and to colour in the corresponding pattern made up of squares in the figure on the right-hand side. Participants are	Dropped after the first feasibility study, as the task was much too easy –produced a very skewed negative distribution.	Adapted from Tzuriel, D. (1995). The Cognitive Modifiability

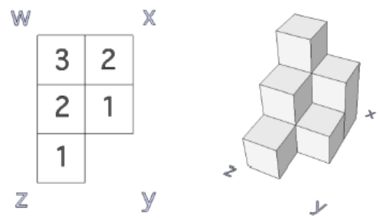
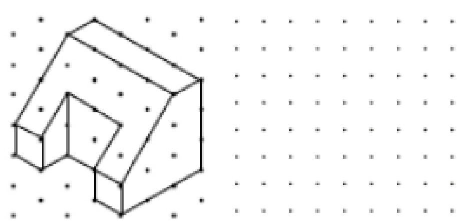
	given **3 minutes** to complete **11 items**.		Battery (CMB). Assessment and intervention: User’s manual. Tel Aviv, Israel: School of Education, Bar-Ilan University
10. Card rotation test 	Participants are asked to identify whether each one of the 8 options presented on the right-hand side of the page is the same shape as the one presented on the left-hand side. If participants think the shape is the same shape (but rotated) they should tick the option “s” at the bottom of the answer shape. If they think it’s a different shape, then they should select the option “d”. The test includes **10 items** to be completed in **3 minutes**.	Dropped after the first feasibility: produced a very skewed distribution, and several other 2D rotation tasks performed better.	French, J., Ekstrom, R., Price, L.: Manual for a kit of reference tests for cognitive factors. Princeton, New Jersey: Educational Testing Service 1963
11. Paper folding test 	On the left-hand side of the page participants are shown a sheet of paper folded following several stages. The last image of the sequence includes a dot. This dot represents a hole that is punched through all the thickness of the paper at that point. Participants are asked to identify which one of the 5 pictures on the right-hand side shows where the holes will be when the paper is completely unfolded again (by reversing the specific steps shown). Participants have **7 minutes** to complete **20 items**.	Included in the second feasibility study (online), with items re-ordered in progressively increasing difficulty (as indicated by scores in the first feasibility study). The resulting test produced a normal distribution and acceptable test-retest reliability. Test-retest (cleaned, standardised): *r*=0.59, N=44,p<0.001. **Adapted version included in the gamified test (the King’s Challenge).**	Adapted from University of Otago, New Zealand http://www.cs.otago.ac.nz/brace/resources/Paper%20Folding%20Test%20Vz-2-BRACE%20Version%2007.pdf
12. Perspective taking (1) –AKA “Point to the cat” 	Participants are first shown the picture on the left-hand side. They are asked to imagine that they are standing in a certain location (one of the shapes), facing another location, and they need to imagine pointing to a third location. They are then asked to draw the direction of their pointing, on the circular diagram shown on the right-hand side. For example: “Imagine you are standing at the flower and facing the tree. Now point to the cat”. Participants have **7 minutes** to complete **12 items**.	Dropped after the first feasibility study, due to poor distribution and internal reliability. Another perspective-taking task (task 13 below) performing better psychometrically and was retained instead.	Kozhevnikov, M. & Hegarty, M. (2001). A dissociation between object-manipulation and perspective-taking spatial abilities. *Memory & Cognition*, 29, 745-756. Hegarty. M. & Waller, D. (2004). A dissociation

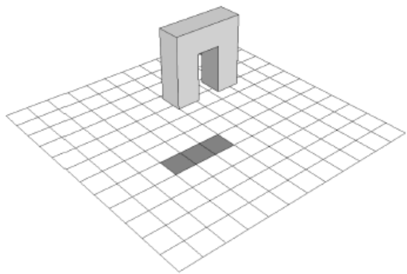
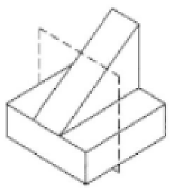
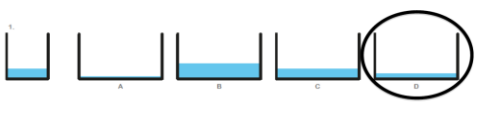
			between mental rotation and perspective-taking spatial abilities. *Intelligence*, 32, 175-191.
13. Perspective taking (2) 	Participants are presented with a transparent cube containing an irregular polygon suspended in the middle of the cube (see example figure). The same polygon is also presented outside the cube from a different viewpoint. Participants are asked to indicate on which corner of the cube they would have to stand in order to see the polygon from the new viewpoint (e.g. the bottom right corner in the example figure). Participants were allowed **8 minutes** to go through **24 questions**.	Included in the second feasibility study (online), with items re-ordered in progressively increasing difficulty. The test produced a normal distribution and very good test-retest reliability. Test-retest (cleaned, standardised): *r*=0.83, N=40, p<0.001. **Adapted version included in the gamified test (the King's Challenge).**	Adapted from Hegarty, M., Keehner, M., Khooshabeh, P., & Montello, D. R. (2009). How spatial abilities enhance, and are enhanced by, dental education. *Learning and Individual Differences*, 19(1), 61-70. Keehner, M., Hegarty, M., Cohen, C. A., Khooshabeh, P., & Montello, D. R. (2008). Spatial reasoning with external visualizations: What matters is what you see, not whether you interact. *Cognitive Science*, 32(7), 1099–1132.
14. Cut the cross-section (1)	Participants are asked to identify the cross-section of three types of figures: single objects (like the example figure), attached objects, and nested objects (where one object is inside the other). The plane cutting the figure can be vertical, horizontal (like the example) or oblique. Participants are given **7 minutes** to complete **15 items**.	Dropped after the first feasibility study, as its correlation with the other cross-sections test, task 15 (*r* = .76) was so high as to render it redundant. Participants also preferred the other cross-sections test.	Cohen, C. A. & Hegarty, M. (2007). Sources of difficulty in imagining cross sections of 3D objects. In D. S. McNamara & J. G. Trafton (Eds.),

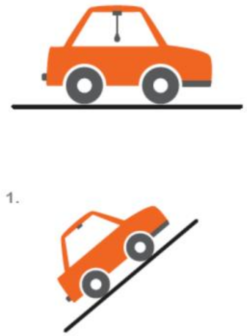
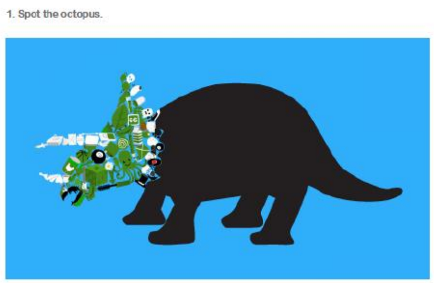
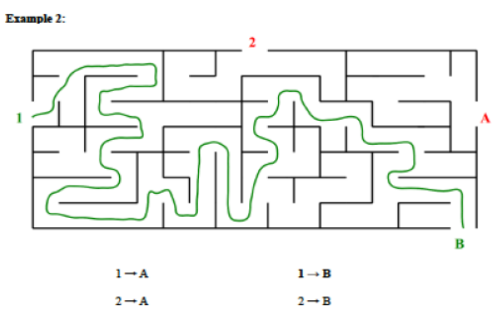
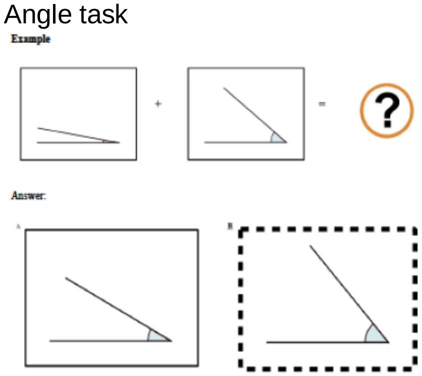
			*Proceedings of the Twenty-Ninth Annual Conference of the Cognitive Science Society* (pp.179-184). Austin TX: Cognitive Science Society. Cohen, C. A. & Hegarty, M. (2012). Inferring cross sections of 3D objects: A new spatial thinking test. *Learning and Individual Differences*, 22(6), 868-874.
15. Cross-section (2) 	Participants are asked to identify the shape that the cutting plane will produce when cutting through several symmetrical solids (see example figures). The plane can cut the solid vertically, horizontally or obliquely. Participants are given **7 minutes** to go through **15 questions**.	Included in the second feasibility study (online) with items re-ordered for progressively increasing difficulty. The test produced a normal distribution and good test-retest reliability. Test-retest (cleaned, standardised): *r*=0.75, N=43, p<0.001. **Included in the gamified test (the King's Challenge).**	Adapted from Ormand, C. J., Shipley, T. F., Tikoff, B., Manduca, C. A., Dutrow, B., Goodwin, L., Hickson, T., Atit, K., Gagnier, K. M., & Resnick, I. (2013). Improving Spatial Reasoning Skills in the Undergraduate Geoscience Classroom Through Interventions Based on Cognitive Science Research. Talk presented at the AAPG Hedberg Conference on 3D Structural Geologic

			Interpretation.
16. 2D to 3D visualization 	Participants are asked to identify which one of 4 3D shapes could be built from the 2D pattern presented on the left-hand side of the picture. Only one shape out of the 4 is the correct answer. Participants are given **8 minutes** to complete **25 items**.	Kept for the second feasibility study (online) but subsequently dropped due to a high positive skew (i.e., it was too difficult), and very poor test-retest reliability. Test-retest (cleaned, standardised): r=0.16, N=30, p=0.41.	Harcourt Assessment (1995), DAT for Selection-Technical Abilities Battery. Pearson Assessment: London
17. Mechanical reasoning 	Participants have **5 minutes** to complete **15 questions** revolving around a common theme: mechanical reasoning. Examples of questions are: "Which shaft will turn more quickly?" (See example picture) and "If only the right oar of the boat is pulled, in which direction will the boat go?"	Included in the second feasibility study (online), with 6 extra items added to the original 15. The test produced a normal distribution and good test-retest reliability: Test-retest (cleaned, standardised): r=0.69, N=46, p<0.001.). In addition to the overall score, the 21 items were grouped thematically into subtests: 5 'pulley' items, 4 'gear' items, and 12 'miscellaneous' items, each with their own subtest score. Following the second feasibility study, the 5 'pulley' items were removed, as this subtest produced poor test re-test reliability (r = 0.39, N=46, p=0.006). **Included in the gamified test (the King's Challenge).**	Adapted from Harcourt Assessment (1995), DAT for Selection-Technical Abilities Battery. Pearson Assessment: London Wiesen, J. (2009), Barron's Mechanical Aptitude and Spatial Relations Test, 2nd edition, Barron's Educational Series Wiesen, J. (2009), Barron's Mechanical Aptitude and Spatial Relations Test, 2nd edition, Barron's Educational Series
18. Spatial number line	Participants are shown a strip of street with a number at the top indicating the length of the street. At the bottom of each picture is a number followed by a question mark indicating a specific distance. Participants are asked to decide which landmark is situated at that specific distance. E.g. in the example picture the total length of the street	Kept for the second feasibility study (online) but subsequently dropped despite good test-retest reliability and distribution of scores: Test-retest (cleaned, standardised): r=0.67, N=50, p<0.001. This task was included in the initial battery	Adapted from the number line test (Siegler, R. S. and Opfer, J. E. (2006). Representational change and

	is 210 meters and participants are asked to identify which landmark is situated at a distance 34 meters from the beginning of the street located on the left-hand side of the page. The tree is the correct answer in this case. The numerical proportions are taken from those in the number line test (Siegler & Opfer, 2006). Participants have **2 minutes** to complete **9 items**.	experimentally as a 'number line' measure, to assess the relationship with mathematical abilities. Its low correlations with other measures appeared to confirm that this was not a spatial task, and it was dropped accordingly.	children's numerical estimation. *Cognitive Psychology*. doi:10.1016/j.cogpsych.2006.09.002
19. 3D to 2D 	Participants are presented with a cube and 4 unfolded 2D patterns. Participants have to decide which one of the 4 unfolded patterns makes the 3D cube. There is only one correct option. Participants have **9 minutes** to complete **13 items**.	Test kept for the second feasibility study (online). Subsequently dropped as it produced a positively skewed distribution and very poor test-retest reliability. Test-retest (cleaned, standardised): r=0.19, N=35, p=0.27.	http://www.psychometric-success.com/aptitude-tests/spatial-ability-tests-cubes.htm
20. Elithorn Maze 	Participants are asked to trace their route on each one of the grids presented (both triangular and rectangular grids are included in this version of the test). The aim of the task is to trace the route passing through the largest possible number of black dots. Participants are asked to start from the bottom part of the shape (the point of the triangle in this case) and move upwards; they can only move left or right on the grid and cannot go backwards; it is not possible to collect all the dots in the grid. **9 items** should be completed in **4 minutes**.	Included in the second feasibility study (online). This test was a computerised version of the original paper and pencil task, in which (in each item) a line moved upwards at a constant speed through a triangular grid, and the participant could change direction (left/right) at each intersection, in an attempt to collect the largest possible number of dots. This test produced a normal distribution and good test-retest reliability: (cleaned, standardised) r=0.76, N=51, p<0.001. **This adapted version was included in the gamified test (the King's Challenge).**	Adapted from Test of spatial planning ability included as a process subtest of the WISC-IV Integrated. ELITHORN, A. (1955). A preliminary report on a perceptual maze test sensitive to brain damage. J. *neurol. neurosurg. Psychiat*, 18, 287-292.
21. Drawing task	This task is divided into 5 subsections each asking participants to draw (see a description and examples for each subsection below). Participants	See subsections below	Adapted from Engage Students in Engineering

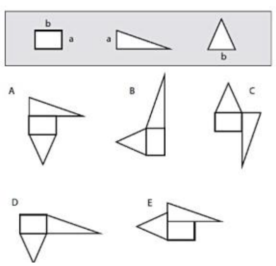
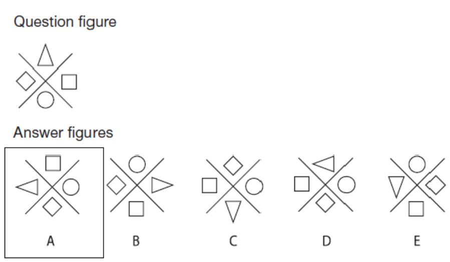
	were given **20 minutes** to complete all **5 subsections**.		site: http://www.wskc.org/documents/281621/307751/ENGAGE_SV_Sample_quiz_on_modules_3_4_and_5.pdf/c3df8086-b535-4ad3-a669-e55be8168820?version=1.0
21.1 2D to 3D drawing 	Participants are presented with the coded plan (see left-hand side of the example picture) and are asked to draw the 3D object corresponding to the plan (like the diagram in the right-hand side of the example picture). This subsection includes **5 items**.	Included in the second feasibility study (online). This test was a computerised version of the original paper-and-pencil task, with participants clicking on dots arranged in an isometric grid to draw lines between them. Showed good distribution and high test-retest reliability. Test-retest (cleaned, standardised): r=0.79, N=37, p<0.001. **Included in the gamified test (the King's Challenge).**	
21.2 3D to 2D viewpoints 	Participants are asked to draw the viewpoint indicated as the 'front' of the picture of the 3D solid (see example figure). The drawing that participants should produce is a 2D viewpoint of the 3D shape. This subsection includes **5 items**.	Included in the second feasibility study (online). This test was a computerised version of the original paper and pencil task, exactly the same as task 21.1, but with the dots arranged in a square rather than an isometric pattern. Showed good distribution and high test-retest reliability. Test-retest (cleaned, standardised): r=0.78, N=47, p<0.001. **Included in the gamified test (the King's Challenge).**	
21.3 Sketch the front and top views 	Participants are asked to sketch the front and top views of the shapes shown on the left-hand side of the grid (see example picture). This subsection included **5 items**.	Dropped after the first feasibility study, due to a highly positively skewed distribution.	
21.4 Draw the reflection	Participants are shown drawings of 3D floating objects and asked to draw the reflection of each	Dropped after the first feasibility study, due to a highly positively skewed	

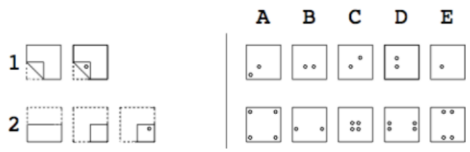
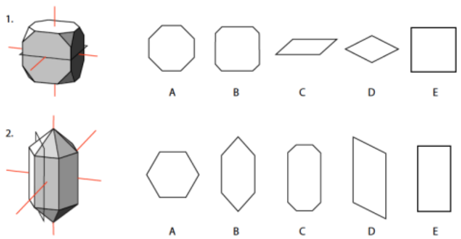
	object on the grid provided (see example). Each grid is like a mirror. This subsection includes **5 items**.	distribution.	
21.5 Sketch the cross-section 	Participants are provided with a grid onto which they need to sketch the cross-section of the objects cut by an imaginary plane shown on the left-hand side of the page (see example figure). This subsection includes **5 items**.	Kept for the second feasibility study (online), including only the easier items from the set. This test was a computerised version of the original paper-and-pencil task, conducted the same way as task 21.2. Dropped after the second feasibility study: it was normally distributed and reliable (test-retest r=0.76, N=47, p<0.001), but highly correlated with task 15 above (cross-sections 2) (r=0.65, N=70, p<0.001), so added little to the battery to justify its long duration compared to other tests.	
22. Water level task 	Participants are presented with water containers of different sizes drawn on the left side of the page. They need to decide which one of the 4 containers on the right side of the page (A, B, C, or D) has the exact same amount of water as that of the first container on the left hand side of the page. Participants are allowed **3 minutes** to complete **9 questions.**	Kept for the second feasibility study (online) as it produced a good distribution in the paper-pencil version. The test was subsequently dropped due to very poor test-retest reliability (cleaned, standardised): *r*=0.13, N=35, p=0.45.	Adapted from Piaget's water level task. Piaget, J., & Inhelder, B. (1956). The child's conception of space. London: Routledge & Kegan Paul.
23. Light bulb task	Participants are presented with a drawing of a car moving on a plane (flat) surface. Inside this car there is a hanging light bulb attached to a string. Participants are then presented with 8 drawings of the same car proceeding on different slopes (uphill and downhill). Their task is to draw the string and the light bulb in the correct inclination for each car, with reference to the angle at which the car is moving uphill or downhill. Participants have **3 minutes** to complete **8 questions**.	Dropped after the paper-pencil feasibility study, as it was much too easy, producing a highly negatively skewed distribution.	Developed by the team

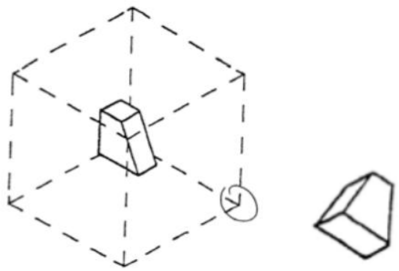
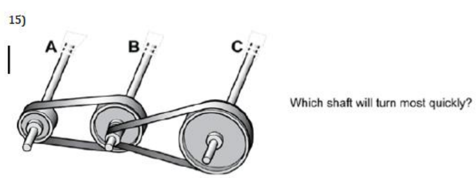
			
24. Scanning task (aka 'little things') 	Participants are presented with several drawings made of small icons. Their aim is to spot the item indicated at the top of each drawing, hidden within the larger figure (see example figure on the left). In order to test quick scanning skills, parts of the original drawings have been blackened out. In this way participants could focus on a restricted area and proceed as fast as possible. Participants are allowed **4 minutes** to complete **10 questions.**	Dropped after the first feasibility study, as another task assessing spatial scanning (task 25 – the mazes task) performed much better in terms of distribution and reliability.	Taken from an IPhone App "Little Things".
25. Mazes task 	Participants are presented with a series of mazes, each with multiple ways in and out, but with only one valid route connecting one of the entrances to one of the exits. Participants are asked to look at the map (see example picture on the left) and choose from the options available the valid route between a single entrance and exit. The test includes **10 items** with increasing difficulty to be completed in **4 minutes**.	Included in the second feasibility study (online). The test produced a normal distribution and good test-retest reliability: Test-retest (cleaned, standardised): r=0.74, N=42, p< 0.001. **Included in the gamified test (the King's Challenge).**	Developed by the team
26. Angle task 	Participants are presented with a series of angles and a mathematical operation to be performed on those angles (adding or subtracting). From four possible options, participants are asked to choose the angle that most closely represents the correct answer (see the example figure on the left). Participants have **2 minutes** to complete **10 questions** of increasing difficulty.	Kept for the second feasibility study (online) but subsequently dropped due to poor test re-test reliability (cleaned, standardised; *r*=0.41, N=41, p=0.009).	Developed by the team

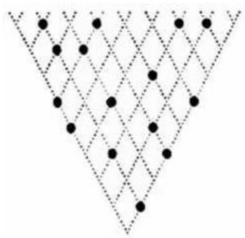
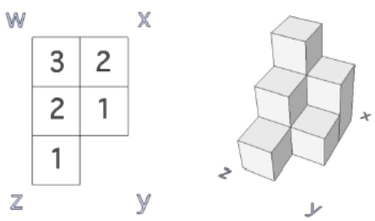
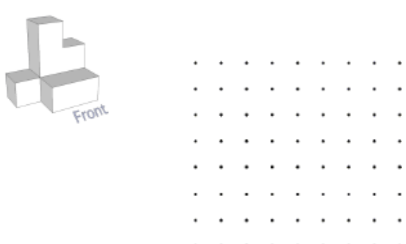
27. Water level task (2) 	Participants are presented with a series of bottles containing some water laying on a plane (flat) surface. Next to each bottle are four empty tilted bottles. Participants are asked to draw a line showing the water level for each tilted bottle as if they were filled with the same amount of water as that in the bottle on the left-hand side. The task includes **5 items** and participants have **2 minutes** to complete it.	Dropped after the paper-pencil feasibility study as it was too easy – negatively skewed distribution.	Adapted from Piaget's water level task Piaget, J., & Inhelder, B. (1956). The child's conception of space. London: Routledge & Kegan Paul.

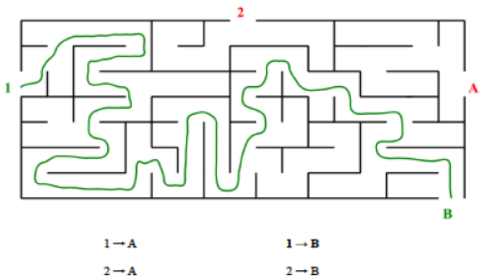
b) The King's Challenge TEDS sibling pilot analyses of 10 tests:

TEST (numbered as above, for reference)	DESCRIPTION	RESULTS	REFERENCE
1. Pattern Assembly (1) 	Participants are asked to decide which option (A - E) is made up of the parts presented in the grey box at the top. The test includes **15 items** each to be completed within a **20 seconds** time frame. Participants are discontinued if they provide 4 consecutive incorrect answers.	**TEDS sibling pilot results:** normally distributed, no floor/ceiling effects, with a mean score of 8.14, SD 2.4, N = 168; test-retest correlation *r*=.56, N = 101, p< .001	Adapted from Spatial reasoning section 1 in the "How2become" booklet (https://www.how2become.com/testing/spatial-reasoning-tests/)(originally 40 items)
5. Shapes rotation (mental rotation) 	Participants are asked to identify which one of the answer figures (A - E) is the same object as in the question figure, but rotated. The test included **15 items** each with a **20 seconds** time limit. Participants are discontinued if they provide 4 consecutive incorrect answers.	**TEDS siblings pilot**: reasonably normally distributed, M = 9.01, SD = 3.30, N = 154. Test-retest *r*= .56, N = 98, p< .001.	Adapted from Spatial reasoning section 1 in the "How2become" booklet (https://www.how2become.com/testing/spatial-reasoning-tests/)
11. Paper-folding test	On the left-hand side of the page participants are	**TEDS sibling pilot:** Normally distributed,	Adapted from

	shown a sheet of paper folded following several stages. The last image of the sequence includes a dot. This dot represents a hole that is punched through all the thickness of the paper at that point. Participants are asked to identify which one of the 5 pictures on the right-hand side shows where the holes will be when the paper is completely unfolded again (by reversing the specific steps shown). The test included **15 items** each to be completed within a **20 second** time limit. Participants are discontinued if they provide 4 consecutive incorrect answers.	N = 166, M = 8.83, SD = 3.3. Test-retest correlation *r* = .58, N = 104, p< .001.	University of Otago, New Zealand http://www.cs.otago.ac.nz/brace/resources/Paper%20Folding%20Test%20Vz-2-BRACE%20Version%2007.pdf
15. Cross-section (2) 	Participants are asked to identify the shape that the cutting plane will produce when cutting through several symmetrical solids (see example figures). The plane can cut the solid vertically, horizontally or obliquely. The test included **15 items** each to be completed within a **20 second** time limit. Participants are discontinued if they provide 4 consecutive incorrect answers.	**TEDS sibling pilot:** normally distributed, M = 7.67, SD = 2.8, N = 159. Test-retest r= .64, N = 91, p< .001.	Adapted from Ormand, C. J., Shipley, T. F., Tikoff, B., Manduca, C. A., Dutrow, B., Goodwin, L., Hickson, T., Atit, K., Gagnier, K. M., & Resnick, I. (2013). Improving Spatial Reasoning Skills in the Undergraduate Geoscience Classroom Through Interventions Based on Cognitive Science Research. Talk presented at the AAPG Hedberg Conference on 3D Structural Geologic Interpretation.
13. Perspective taking (2) –AKA "The cube"	Participants are presented with a transparent cube containing an irregular polygon suspended in the middle of the cube (see example figure). The same polygon is also presented outside the cube from a different viewpoint. Participants are asked to indicate on which corner of the cube they would have to stand in order to see the polygon	**TEDS sibling pilot:** normally distributed, M = 6.61, SD 3.34, N = 147. Test-retest *r*= .56, N = 92, p< .001.	Adapted from Hegarty, M., Keehner, M., Khooshabeh, P., & Montello, D. R. (2009). How spatial abilities

	from the new viewpoint (e.g. the bottom right corner in the example figure). The test included **15 items** each to be completed within a **20 second** time limit. Participants are discontinued if they provide 5 consecutive incorrect answers.		enhance, and are enhanced by, dental education. *Learning and Individual Differences*, 19(1), 61-70. Keehner, M., Hegarty, M., Cohen, C. A., Khooshabeh, P., & Montello, D. R. (2008). Spatial reasoning with external visualizations: What matters is what you see, not whether you interact. *Cognitive Science*, 32(7), 1099–1132.
17. Mechanical reasoning 	Examples of questions are: "Which shaft will turn more quickly?" (See example picture) and "If only the right oar of the boat is pulled, in which direction will the boat go?". The test included **16 items** each to be completed within a **25 second** time limit. Participants are required to complete every item.	**TEDS sibling pilot:** close to normally distributed, M = 9.53, SD 2.25, N = 180. Test retest *r*= .65, N = 113, p< .001.	Adapted from Harcourt Assessment (1995), DAT for Selection-Technical Abilities Battery. Pearson Assessment: London Wiesen, J. (2009), Barron's Mechanical Aptitude and Spatial Relations Test, 2nd edition, Barron's Educational Series Wiesen, J. (2009), Barron's Mechanical

			Aptitude and Spatial Relations Test, 2^{nd} edition, Barron's Educational Series
20. Elithorn Maze 	Participants are asked to trace their route on each one of the triangular grids presented. The aim of the task is to trace the route passing through the largest possible number of black dots. Participants start from the bottom of the triangle and move upwards at a fixed speed; they can only move left or right on the grid, changing direction as desired at each intersectionI, It is not possible to collect all the dots in the grid. The test included **10 items**, each to be completed within **7 second**s. Participants are required to complete every item.	**TEDS sibling pilot**: fairly normally distributed, M = 7.31, SD = 1.94, N = 184. Test-retest *r*= .69, N = 117, p< .001.	Adapted from Test of spatial planning ability included as a process subtest of the WISC-IV Integrated. ELITHORN, A. (1955). A preliminary report on a perceptual maze test sensitive to brain damage. J*. neurol. neurosurg. Psychiat*, 18, 287-292.
21.1 2D to 3D drawing 	Participants are presented with the coded plan (see left-hand side of the example picture) and are asked to draw the 3D object corresponding to the plan (see right-hand side of the example picture), by clicking on dots arranged in an isometric grid. This test included **5 items**, each with a time limit of **70 seconds**. Participants are required to complete every item.	**TEDS sibling pilot:** fairly normal distribution, M = 3.46, SD = 1.69, N = 155. Test-retest *r* = .63, N = 99, p< .001.	Developed by the team
21.2 3D to 2D viewpoints 	Participants are asked to draw the viewpoint indicated in the picture of the 3D solid as the 'front' (see example figure), by clicking on dots arranged in a square grid. The drawing that participants should produce is a 2D viewpoint of the 3D shape. This test included **5 items**, each had a time limit of **45 seconds**. Participants are required to complete every item.	**TEDS sibling pilot:** slightly negatively skewed distribution M = 3.73, SD .99, N = 186. Test-retest correlation *r*= .68, N= 117, p< .001.	

25. Mazes task **Example 2:**  	Participants are presented with a series of mazes, each with multiple ways in and out, but with only one valid route connecting one of the entrances to one of the exits. Participants are asked to look at the map (see example picture on the left) and choose from the options available the valid route between a single entrance and exit. The test includes **10 items** with increasing difficulty, each with a **25 second** time limit, discontinuing after 4 consecutive incorrect responses	**TEDS sibling pilot:** Normal distribution, M = 5.92, SD = 1.76, N = 167. Test-retest correlation *r*= .48, N = 106, p< .001.	Developed by the team