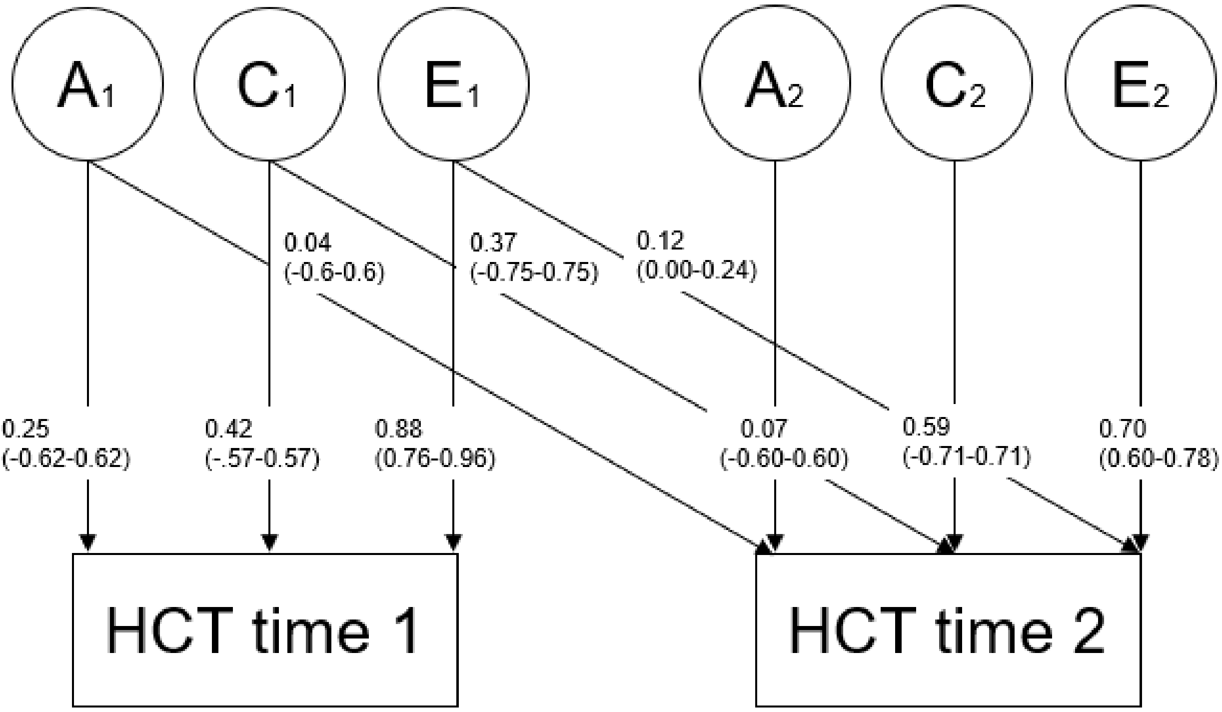
Supplementary Information

[S1] Supplementary Figure 1: unstandardised estimates derived from a Cholesky decomposition of heartbeat counting scores, after accounting for Body Mass Index and sex.



Body mass index (BMI) was calculated using the following equation: weight(kg)/height(m^2). When BMI was regressed out of heartbeat counting scores, sample sizes at time 1 and time 2 were 278 and 200, respectively. We controlled for selection bias by linking to the full TEDS sample, as described in the main text. Results are very similar to the findings derived from raw data in that the 95% confidence intervals are overlapping (see Supplementary Figure S10 above for direct comparison; estimates in the main text are obtained by squaring the unstandardised values). Since the sample size is larger without residualising the heartbeat counting task scores, we present the raw data analyses in the main text.

[S2] Supplementary Table 1. Correlations between heartbeat counting and subscales of the SCARED at Time 1 & Time 2.

	Somatic/ panic scale	General anxiety scale	Separation anxiety scale	Social phobia scale	School phobia scale
Time 1	-.127*	.008	.041	.008	.019
Time 2	.022	.094	.081	.150*	.126
Time 1 -> Time 2	-.024	.058	.069	.154*	-.002

**denotes significant at p<.05 before correction for multiple comparisons. Note that none of these correlations reached significance after correction for multiple comparisons. Time 1 -> Time 2 refers to phenotypic correlations between heartbeat counting error scores at Time 1 and questionnaire/cognition measures at Time 2*

[S3] Supplementary Table 2. Correlations between heartbeat counting and subscales of the CSHQ at Time 1 & Time 2.

	Bedtime resistance scale	Sleep onset delay scale	Sleep duration scale	Sleep anxiety scale	Night waking scale	Parasomnias scale	Sleep disordered breathing scale	Daytime sleepiness
Time 1	.037	-.009	-.049	.021	.004	.057	.094	-.109
Time 2	-.027	.081	-.017	-.005	.127	.122	.122	-.043
Time 1 -> Time 2	-.094	.028	-.067	-.123	-.085	-.026	.041	-.078

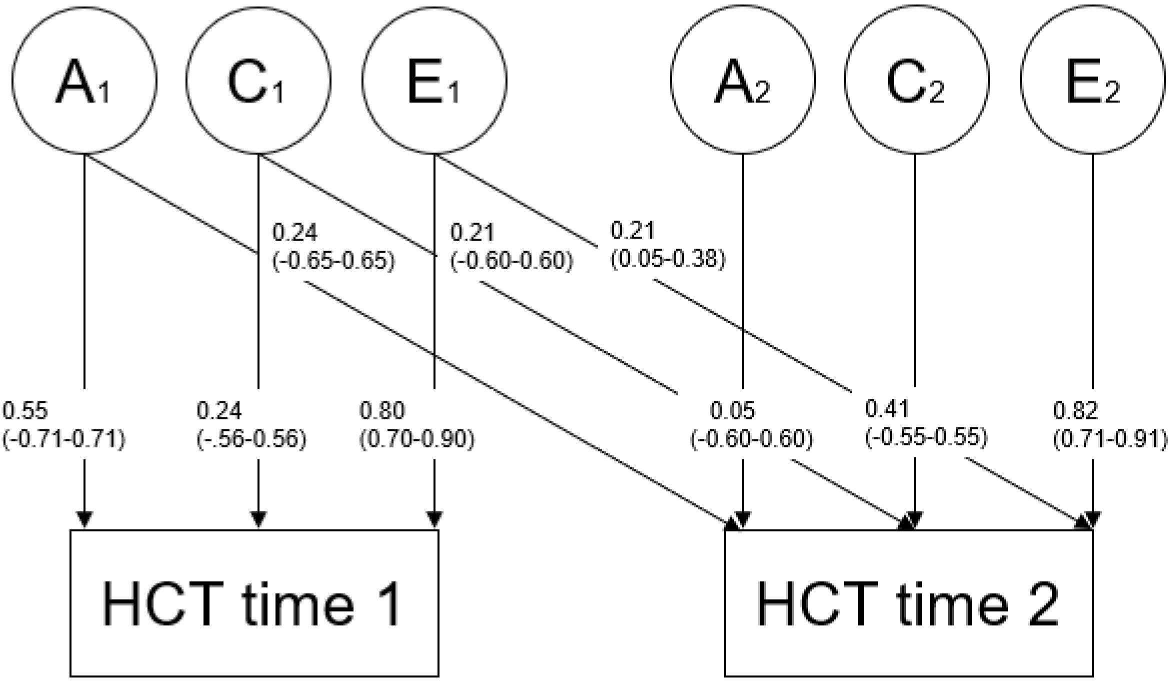
**denotes significant at p<.05 before correction for multiple comparisons. Note that none of these correlations reached significance after correction for multiple comparisons. Time 1 -> Time 2 refers to phenotypic correlations between heartbeat counting error scores at Time 1 and questionnaire/cognition measures at Time 2*

[S4] Supplementary Table 5. Correlations between heartbeat counting and expression recognition scores for each expression at Time 2.

	Anger	Disgust	Fear	Happiness	Sadness
Time 2	.098	.080	-.027	.057	.063
Time 1 -> Time 2	-.104	-.074	-.053	-.145*	-.145*

**denotes significant at p<.05 before correction for multiple comparisons. Note that none of these correlations reached significance after correction for multiple comparisons. Time 1 -> Time 2 refers to phenotypic correlations between heartbeat counting error scores at Time 1 and questionnaire/cognition measures at Time 2*

[S5] Supplementary Figure 2: Unstandardised estimates from the Cholesky decomposition model of performance at the heartbeat counting task, with 95% confidence intervals.



[S6] Supplementary Table 9: Model fit statistics for the Cholesky decomposition model of performance at the heartbeat counting task, and for other models for comparison.

base	comparison	ep	minus2LL	df	AIC	diffLL	diffdf	*p*
Sat	<NA>	54	31481.350	11332	8817.347	NA	NA	NA
Sat	ACE	21	31518.990	11365	8788.994	37.647	33	0.265
Sat	AE	20	31519.190	11366	8787.193	37.845	34	0.298
Sat	CE	20	31519.660	11366	8787.658	38.311	34	0.280
Sat	Env	19	31526.990	11367	8792.986	45.639	35	0.108
ACE	AE	20	31519.190	11366	8787.193	0.199	1	0.656
ACE	CE	20	31519.660	11366	8787.658	0.664	1	0.415
ACE	Env	19	31526.990	11367	8792.986	7.992	2	0.018

Note: Sat = saturated model; ACE = Full model; AE = model with no shared environmental influences across time; CE = model with no genetic influences across time; Env= model with no familial influences and only individual-specific environmental influences across time. Dropping shared environmental or genetic influences on stability individually does not significantly reduce model fit compared to the full model (p=0.656 and 0.415 for the AE and CE models, respectively). However, a model simultaneously removing both sources of familial influence (the 'Env' model) is significantly worse-fitting (p=0.018) than the full model. Therefore, although the paths are not significantly above zero, we cannot rule out genetic and shared environmental influences on stability in HCT scores and present the full model in the main text