Supplementary Tables

Table S1. All 18 game tasks: targeted ability, corresponding paradigm, and difficulty scaling.

Game task	Primary ability	Corresponding paradigm	Difficulty scaling
ColorJudgement	Selective attention / inhibitory control	Stroop (Stroop, 1935)	More color options, shorter time, more correct answers required
AlienLanguage	Visual search / processing speed	Visual search	More items within a fixed time limit
Singler	Inductive reasoning	Odd-one-out / Raven-type (Leavitt, 2011)	More animals, more distinguishing features, more complex combinations
MatchCard	Visuospatial working memory	N-back (Kirchner, 1958)	Shorter exposure, more image types, more matches required
CubesProjection	Mental rotation	Mental rotation (Shepard & Metzler, 1971)	More blocks along X/Y/Z axes
Mines	Deductive reasoning	Deductive reasoning (Gray & Veksler, 2004)	Larger board, more mines, time limit, more boards/lives
Train	Cognitive flexibility / task switching	Task switching (Monsell, 2003)	More frequent/varied targets, more destinations
Roadblocks	Visuospatial working memory	Visuospatial WM (Baddeley, 1992)	Larger board, more obstacles, shorter visibility/time
Pack	Spatial planning / mental rotation	Mental rotation + spatial planning (Shepard & Metzler, 1971)	Type/size/number of empty cells in the grid
ServeFood	Visual search / processing speed	Visual search	More customers, tighter time, more correct orders
SpaceStation	Visuospatial working memory	Route learning (Maguire, Frackowiak, & Frith, 1996)	Larger layout, more turns/crossings, longer path
Delivery	Working memory	Memory updating (Garavan, 1998)	More and more varied items
Toilet	Sustained attention	Continuous performance test (Rosvold et al., 1956)	Longer levels, shorter time per command
Lines	Visuospatial planning / problem-solving	Trail Making (Reitan, 1958)	Larger n×n grid, more dots
Shopping	Working memory	Free recall / short-term memory (Miller, 1956)	More list items, more shelf items, shorter list visibility
OneOrTwoSteps	Processing speed / response inhibition	Go/No-Go (Verbruggen & Logan, 2008)	More rows, tighter time, more special rule plots
InnerSpaceMap	Spatial navigation / mental rotation	Mental rotation + navigation (Shepard & Metzler, 1971)	Larger layout, more inaccessible areas

Game task	Primary ability	Corresponding paradigm	Difficulty scaling
Barbecue	Divided attention / response inhibition	Go/No-Go	More food items, varied cooking times

Table S2. Structure of the fixed daily training rotation.

Day type	Tasks administered
Set A days (days 1, 4, 7, …)	AlienLanguage, Mines, Train, CubesProjection, Roadblocks (+ the 5 in-game test tasks every third day)
Set B days (days 2, 5, 8, …)	ServeFood, Lines, Barbecue, SpaceStation, Delivery
Set C days (days 3, 6, 9, …)	AlienLanguage, Mines, OneOrTwoSteps, InnerSpaceMap, Shopping

Note. The full day-by-day schedule across 19 training days follows the study's daily training plan. The five in-game test tasks (ColorJudgement, Singler, MatchCard, Pack, Toilet) recurred on the every-third-day test schedule.

Table S3

Table S3. Descriptive statistics and signed within-group effect sizes by group.

Outcome	Group	Pre, M (SD)	Post, M (SD)	Paired n	Signed dz
MRT rotation RT	ADT	2.292 (0.265)	2.032 (0.385)	48	+0.57
MRT rotation RT	LDT	2.251 (0.311)	2.137 (0.431)	42	+0.21
MRT rotation RT	Recreational gaming	2.381 (0.178)	2.312 (0.207)	45	+0.36
Raven >=2-element ACC	ADT	0.832 (0.076)	0.856 (0.078)	47	+0.21
Raven >=2-element ACC	LDT	0.794 (0.104)	0.847 (0.072)	39	+0.44
Raven >=2-element ACC	Recreational gaming	0.901 (0.055)	0.903 (0.063)	49	+0.09
N-back 2-vs-0 RT	ADT	0.194 (0.102)	0.125 (0.066)	48	+0.73
N-back 2-vs-0 RT	LDT	0.193 (0.149)	0.182 (0.100)	45	+0.06
N-back 2-vs-0 RT	Recreational gaming	0.186 (0.098)	0.151 (0.077)	48	+0.35
ANT executive RT	ADT	0.075 (0.026)	0.055 (0.023)	47	+0.70
ANT executive RT	LDT	0.078 (0.032)	0.058 (0.023)	40	+0.62
ANT executive RT	Recreational gaming	0.064 (0.021)	0.053 (0.016)	43	+0.42
Go/No-Go d′	ADT	5.604 (1.083)	5.397 (1.133)	51	-0.18
Go/No-Go d′	LDT	5.811 (1.161)	5.403 (1.199)	46	-0.31

Outcome	Group	Pre, M (SD)	Post, M (SD)	Paired n	Signed dz
Go/No-Go d′	Recreational gaming	5.808 (1.156)	6.054 (0.962)	49	+0.18

Note. Signed dz: positive indicates improvement (RT decrease, ACC/d′ increase), negative indicates decline. Pre and post descriptive means are based on available observations; dz and paired n use paired observations. Two participants' pretest values on four tasks were treated as missing (Section 2.8).

Table S4. Full convergent-validity matrix: in-game test-task RT × standard test (Spearman ϱ).

In-game task (level)	Standard measure	ϱ	95% CI	p	pFDR	n
ColorJudgement (level 13)	Go/No-Go d′	−.105	[−.303, .101]	.315	.779	93
ColorJudgement (level 13)	N-back 2-vs-0 RT	.160	[−.046, .353]	.127	.779	92
ColorJudgement (level 13)	MRT rotation RT	.026	[−.180, .230]	.806	.905	92
ColorJudgement (level 13)	ANT executive control	.064	[−.150, .272]	.559	.779	86
ColorJudgement (level 13)	Raven ≥2-element ACC	−.118	[−.321, .095]	.275	.779	87
Singler (level 8)	Go/No-Go d′	−.085	[−.284, .120]	.416	.779	93
Singler (level 8)	N-back 2-vs-0 RT	−.064	[−.266, .142]	.542	.779	92
Singler (level 8)	MRT rotation RT	−.078	[−.279, .129]	.459	.779	92
Singler (level 8)	ANT executive control	−.018	[−.229, .195]	.869	.905	86
Singler (level 8)	Raven ≥2-element ACC	−.146	[−.346, .067]	.179	.779	87
MatchCard (level 12)	Go/No-Go d′	−.005	[−.209, .199]	.961	.961	93
MatchCard (level 12)	N-back 2-vs-0 RT	.308	[.111, .483]	.003	.070	92
MatchCard (level 12)	MRT rotation RT	.053	[−.153, .255]	.615	.779	92
MatchCard (level 12)	ANT executive control	−.119	[−.323, .095]	.273	.779	86
MatchCard (level 12)	Raven ≥2-element ACC	−.057	[−.265, .155]	.598	.779	87
Pack (level 8)	Go/No-Go d′	−.021	[−.224, .183]	.840	.905	93
Pack (level 8)	N-back 2-vs-0 RT	.089	[−.118, .289]	.398	.779	92
Pack (level 8)	MRT rotation RT	.189	[−.016, .379]	.071	.779	92
Pack (level 8)	ANT executive control	−.059	[−.268, .155]	.589	.779	86
Pack (level 8)	Raven ≥2-element ACC	.097	[−.116, .302]	.370	.779	87
Toilet (level 18)	Go/No-Go d′	−.065	[−.265, .141]	.536	.779	93
Toilet (level 18)	N-back 2-vs-0 RT	.171	[−.035, .364]	.102	.779	92
Toilet (level 18)	MRT rotation RT	−.052	[−.254, .155]	.623	.779	92
Toilet (level 18)	ANT executive control	−.020	[−.230, .193]	.858	.905	86
Toilet (level 18)	Raven ≥2-element ACC	−.118	[−.321, .095]	.276	.779	87

Note. Spearman rank correlations between in-game test-task response time (mean per-trial RT across the participant's non-passive sessions at the shared level) and posttest standardized measures, for the 94 participants with at least one non-passive session at the shared level of all five tasks; n varies by pair because of missing standardized scores. 95% CIs are Fisher z-transformed. pFDR = Benjamini–Hochberg correction across all 25 pairings. Only MatchCard–N-back 2-vs-0 RT is significant before correction; none survives FDR correction.

Table S5

Table S5. Experiment 2: matched subsample analyses.

Matching rule	Pairs	Outcome	$\chi^2(1)$	p	pFDR
primary_025	19	MRT rotation RT	4.00	0.046	0.228
primary_025	19	Raven >=2-element ACC	1.90	0.168	0.419

Matching rule	Pairs	Outcome	$\chi^2(1)$	p	pFDR
primary_025	19	N-back 2-vs-0 RT	0.44	0.507	0.507
primary_025	19	ANT executive RT	0.99	0.319	0.507
primary_025	19	Go/No-Go d′	0.49	0.486	0.507
age_010	13	MRT rotation RT	6.95	0.008	0.042
age_010	13	Raven >=2-element ACC	0.38	0.536	0.621
age_010	13	N-back 2-vs-0 RT	0.57	0.451	0.621
age_010	13	ANT executive RT	0.24	0.621	0.621
age_010	13	Go/No-Go d′	3.05	0.081	0.202
birth_distance_025	32	MRT rotation RT	1.05	0.307	0.383
birth_distance_025	32	Raven >=2-element ACC	3.03	0.082	0.383
birth_distance_025	32	N-back 2-vs-0 RT	0.34	0.559	0.559
birth_distance_025	32	ANT executive RT	1.05	0.304	0.383
birth_distance_025	32	Go/No-Go d′	1.17	0.280	0.383

Note. Primary 1:1 nearest-neighbour match: age at testing (caliper .25 years), sex and education stage exact; only 19 pairs. Sensitivity matches: .10-year age caliper (13 pairs), and birth-date matching (32 pairs; does not adjust age at testing). All matched analyses are descriptive and not used for inference. Primary-match demographics: mean ages 20.02 (ADT) vs. 20.07 (control) years; sex 12/7 in both groups; education stage identical; years of study 5/12/2 vs. 1/6/12.

Table S6

Table S6. Experiment 2: full-sample age × time terms (ADT vs. recreational gaming).

Outcome	Age × time β (95% CI)	$\chi^2(1)$	p	pFDR
MRT rotation RT	-0.003 (-0.087, 0.081)	0.00	0.946	0.946
Raven >=2-element ACC	-0.005 (-0.025, 0.015)	0.27	0.602	0.904
N-back 2-vs-0 RT	-0.006 (-0.028, 0.016)	0.25	0.617	0.904
ANT executive RT	0.005 (-0.001, 0.012)	2.77	0.096	0.481
Go/No-Go d′	0.048 (-0.219, 0.316)	0.13	0.723	0.904

Note. Age centred; Wald confidence intervals and likelihood-ratio p values. The group × time comparisons for models with and without age × time appear in main Table 4.

Panel B. Group × time estimates with and without age × time.

Outcome	Age-only β (95% CI)	Age + age × time β (95% CI)
MRT rotation RT	0.188 (0.049, 0.328)	0.191 (0.028, 0.355)
Raven >=2-element ACC	-0.020 (-0.053, 0.013)	-0.015 (-0.054, 0.024)
N-back 2-vs-0 RT	0.031 (-0.005, 0.068)	0.037 (-0.006, 0.080)
ANT executive RT	0.009 (-0.002, 0.019)	0.003 (-0.009, 0.016)
Go/No-Go d′	0.433 (-0.008, 0.874)	0.383 (-0.135, 0.902)

Table S7

Table S7. Training exposure and adherence by group (Experiment 1)

Measure	ADT (n = 51) Mdn [IQR]	LDT (n = 46) Mdn [IQR]	U	p
Total active training time (h)	22.30 [20.25, 23.89]	22.89 [21.48, 26.34]	952	.111
Sessions completed	872 [761, 953]	1,739 [1,392.75, 2,054.5]	122	< .001
Training days played	19 [19, 19]	19 [19, 19]	1,053.5	.171

Measure	ADT (n = 51) Mdn [IQR]	LDT (n = 46) Mdn [IQR]	U	p
Training days meeting the 1.5-h requirement (%)	94.74 [86.84, 100]	94.74 [79.21, 100]	1,274	.449
Passive sessions	147 [120.5, 183]	25 [13.25, 50.75]	2,204.5	< .001

Note. Medians with interquartile ranges; two-sided Mann–Whitney U tests. Total active training time is the summed duration of non-passive sessions. Training days exclude the installation (familiarisation) day; days after the 19th training day (permitted before post-testing, Section 2.3) are included. Session start = logged end time − duration; within each training day, a new sitting began when the next session started more than 15 min after the latest end time of the sessions already in that sitting; a day met the requirement when its sittings summed to at least 90 min (pooled: ADT 88.4% of 976 training days, LDT 86.6% of 887). Passive sessions: zero correct responses or fewer than 50% of the mean number of correct responses of other participants at the same difficulty level (pooled over all logged days, including the familiarisation day: ADT 8,260 of 44,390 sessions, 18.6%; LDT 2,005 of 78,764, 2.5%). Of the 48 LDT participants who completed training, one had no cognitive-outcome data and one had an empty game-log archive.

Table S8

Table S8. Correlations for the five domain-matched in-game task–standard measure pairs across difficulty levels.

In-game task	Standard measure	Level	n	ϱ	p
ColorJudgement	Go/No-Go d′	8	96	−.093	.365
ColorJudgement	Go/No-Go d′	11	96	−.092	.372
ColorJudgement	Go/No-Go d′	13 a	96	−.115	.263
Singler	Raven ≥2-element ACC	4	89	−.093	.384
Singler	Raven ≥2-element ACC	6	90	−.063	.554
Singler	Raven ≥2-element ACC	8 a	90	−.169	.112
MatchCard	N-back 2-vs-0 RT	4	94	.266	.009
MatchCard	N-back 2-vs-0 RT	6	95	.180	.081
MatchCard	N-back 2-vs-0 RT	8	95	.312	.002
MatchCard	N-back 2-vs-0 RT	10	95	.116	.264
MatchCard	N-back 2-vs-0 RT	12 a	95	.279	.006
Pack	MRT rotation RT	4	93	.251	.015
Pack	MRT rotation RT	6	95	.071	.496
Pack	MRT rotation RT	8 a	94	.171	.099
Toilet	ANT executive control	8	89	.106	.323
Toilet	ANT executive control	10	89	−.065	.546
Toilet	ANT executive control	12	89	.106	.322
Toilet	ANT executive control	14	88	.169	.115
Toilet	ANT executive control	16	87	−.089	.410
Toilet	ANT executive control	18 a	87	−.009	.933

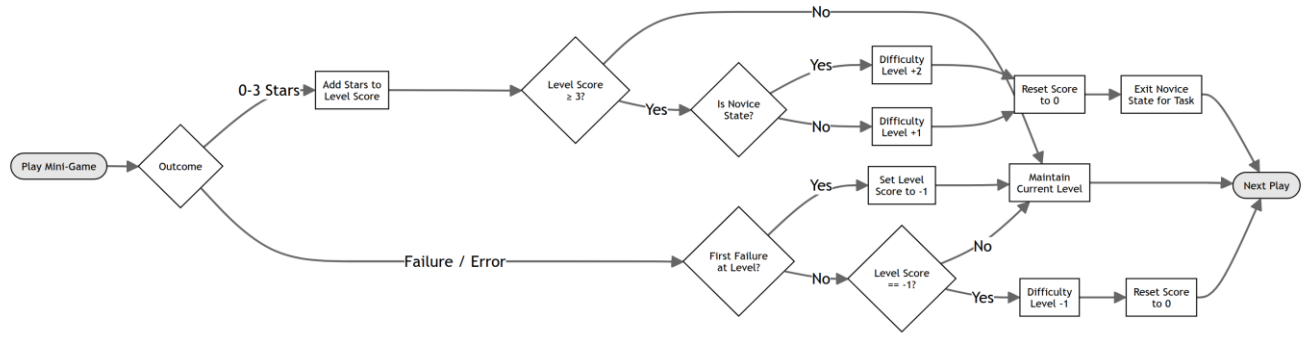
Note. Descriptive Spearman rank correlations between in-game test-task response time at each difficulty level and the corresponding posttest standardized measure, for all participants with data at that level (n = 87–96); p values are uncorrected. a Shared level used in Figure 3 and Table S4. MatchCard–N-back correlations are positive at every level (ϱ = .116 to .312),

and ColorJudgement–Go/No-Go correlations are negative at every level (ϱ = −.092 to −.115).

Figure S1

Figure S1. Decision logic of the adaptive-difficulty algorithm and a worked numerical example. (A) After each mini-game play, the algorithm evaluates the outcome (stars earned or failure) and updates a hidden performance score maintained separately for each task at its current difficulty level. A score reaching the promotion threshold (≥3) triggers a difficulty increase — +2 levels if the participant is encountering this difficulty for the first time ("novice" state), or +1 level otherwise — after which the score resets to 0. A failure sets the score to −1; a second consecutive failure at the same difficulty level (the "two-strike" rule) triggers a one-level demotion and resets the score to 0. (B) A worked example tracing a hypothetical participant's score, novice/normal state, and resulting difficulty level across eight consecutive plays of a single mini-game, illustrating how the rules in (A) combine over successive trials.

(A) Decision logic



(B) Worked numerical example

Play	Initial State	Stars	Algorithm Action & Decision Logic	Resulting State
1	L1, Score 0, Novice	2	0 + 2 = 2; < 3 → no promotion	L1, Score 2, Novice
2	L1, Score 2, Novice	2	2 + 2 = 4; ≥ 3 → promote; Novice → +2; reset; exit Novice	L3, Score 0, Normal
3	L3, Score 0, Normal	Fail	First failure at level → set score to -1	L3, Score -1, Normal
4	L3, Score -1, Normal	Fail	Score already -1 (two-strike) → demote 1 level; reset	L2, Score 0, Normal
5	L2, Score 0, Normal	1	0 + 1 = 1; < 3 → no promotion	L2, Score 1, Normal
6	L2, Score 1, Normal	3	1 + 3 = 4; ≥ 3 → promote; Normal → +1; reset	L3, Score 0, Normal
7	L3, Score 0, Normal	0	0 + 0 = 0; < 3 → no promotion (not a failure)	L3, Score 0, Normal
8	L3, Score 0, Normal	Fail	First failure at level → set score to -1	L3, Score -1, Normal