No Evidence of Intelligence Improvement After Working Memory Training: A Randomized, Placebo-Controlled Study

Thomas S. Redick, Zach Shipstead, Tyler L. Harrison, Kenny L. Hicks, Michael J. Kane, David E. Fred, David Z. Hambrick, Randall W. Eagle

- **Introduction** – Hypothesis, critical analysis towards prior studies, predictions, possible outcomes

- **Results** – How procedure took place, description of results and what they tell us

- **Discussion** – How results differ form other studies, what we can conclude more firmly about intelligence improvement after working memory training

Objective

- Purpose of experiment - to be able to answer:

"*Does repeated practice on an adaptive dual n-back task transfer to, and actually cause, improvements in intelligence, multitasking and WM capacity?*"

Introducing past study, and their findings

Susanne Jaeggi

Dual n-back accuracy positively correlates with performance on fluid intelligence tests (Jaeggi et al., 2008).

'Trained subjects exhibited **significantly larger gains** on an intelligence test in comparison to **no contact control subjects who did not perform any intervention practice**, between the pre and post mid sessions.'

Criticism from recent study on Jaeggi et al (2008):

- Data collapsed across different transfer tests administered under different time limits.
- There were procedural differences across the four studies.
- Patterns of transfer differed across four studies.
- Individual studies based on very small samples.

What Redick et al (2013) have changed as result:

- Included diverse samples.
- Several transfer measures.
- Placebo (active control) group.
- Transfer sessions increased to 3.

Predictions – What might happen

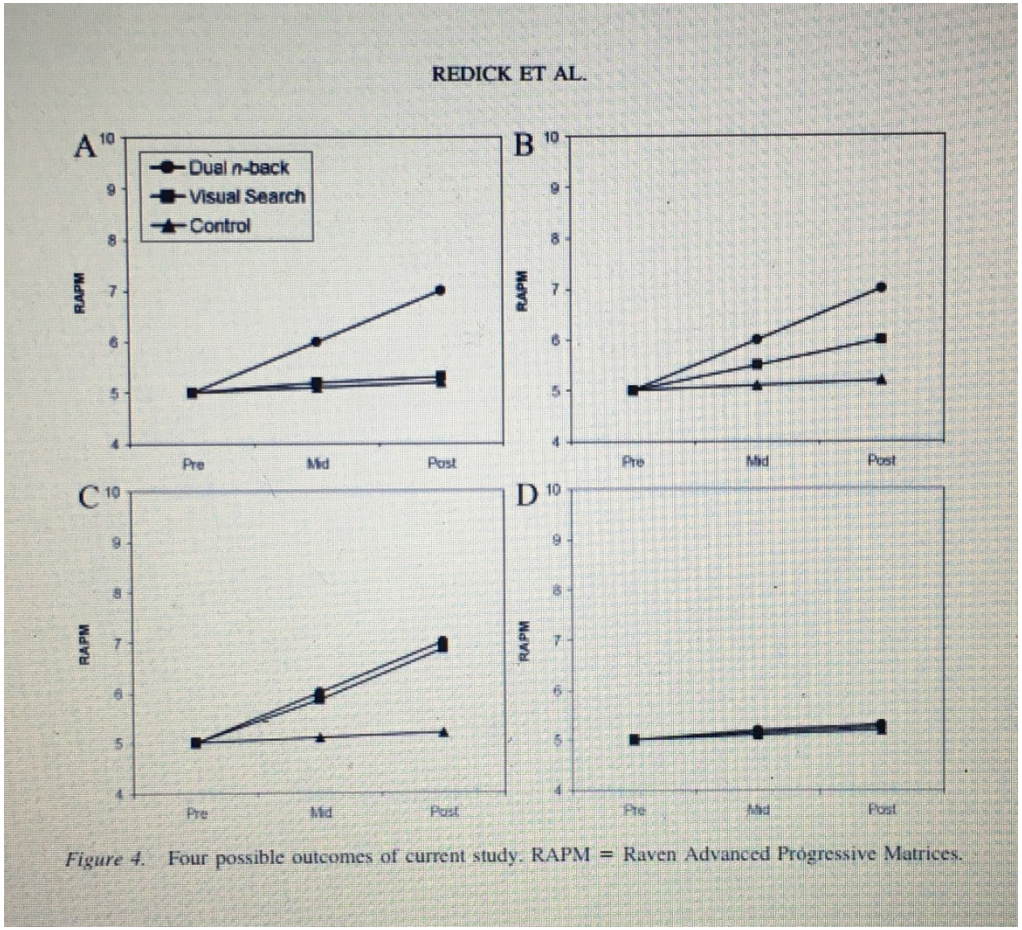


Figure 4. Four possible outcomes of current study. RAPM = Raven Advanced Progressive Matrices.

A: Groups trained via **dual n-back practice** seen improvements in fluid intel relative to both visual search practice and no contact.

B: Visual search practice produces improvement relative to no contact group – Dual n-back training builds greater improvement.

C: Both dual n-back and visual search training increases fluid intel relative to no contact.

D: null hypothesis – **none** of the groups show intel test improvement.

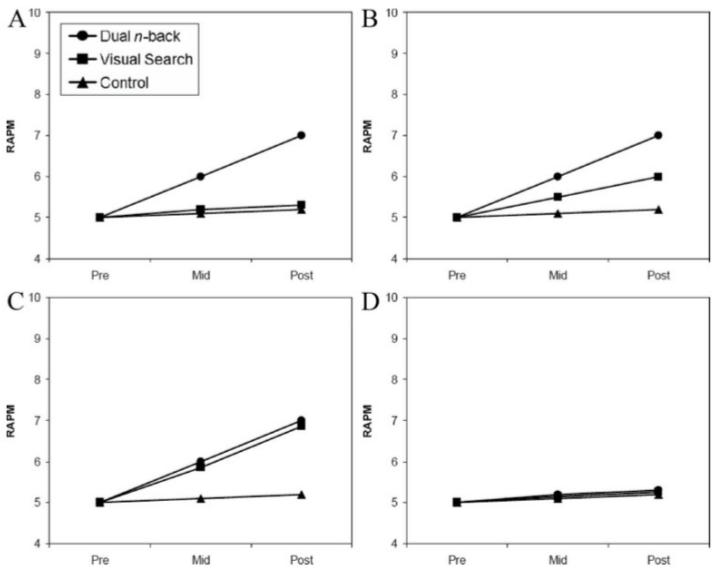


Figure 4. Four possible outcomes of current study. RAPM = Raven Advanced Progressive Matrices.

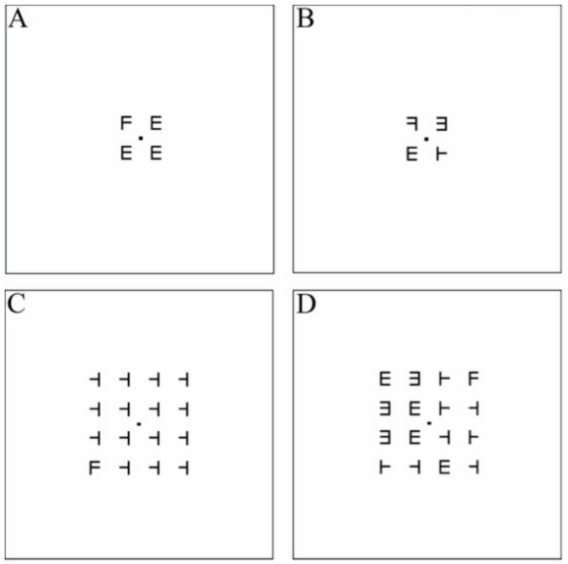
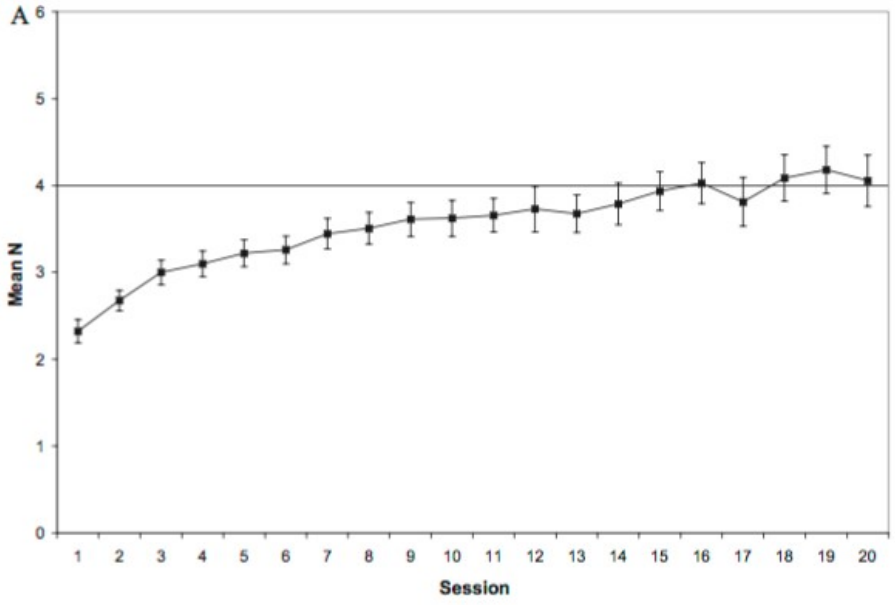


Figure 5. Example stimuli from different levels of the adaptive visual search task: 2 × 2 homogeneous (Level 1; A), 2 × 2 heterogeneous (Level 2; B), 4 × 4 homogeneous (Level 3; C), 4 × 4 heterogeneous (Level 4; D).

Table 1
Demographic Information

		Gender		Age (years)		No. of subjects			
Group	*N*	Male	Female	*M*	*SD*	GT	GSU	MSU	Other
Dual *n*-back	24	10	14	21.1	2.7	9	7	7	1
Visual search	29	12	17	20.7	2.5	9	11	8	1
Control	20	10	10	21.2	2.5	7	7	5	1

Note. GT = Georgia Tech student; GSU = Georgia State University student; MSU = Michigan State University student; other = not currently attending one of these three colleges.



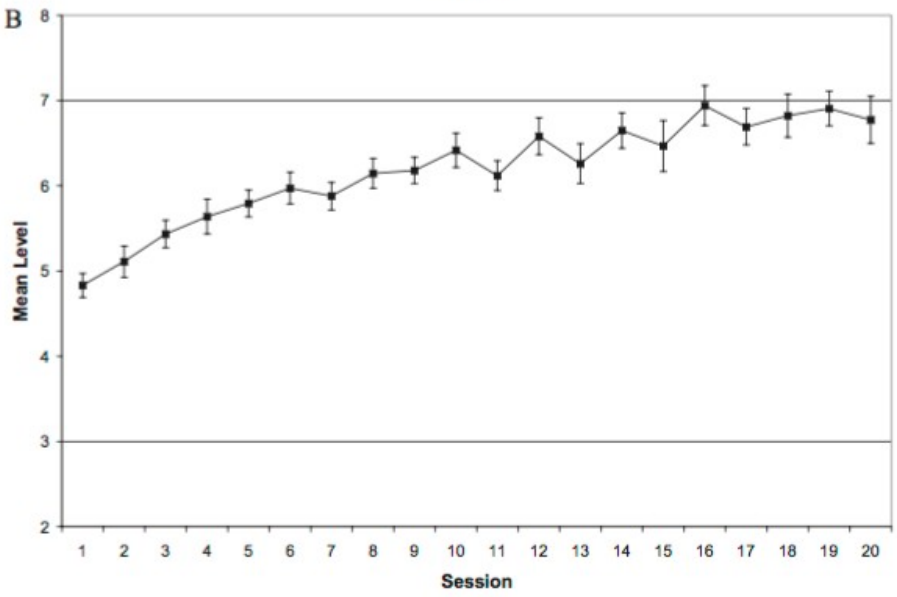


Figure 6. Practice data for the dual *n*-back (A) and visual search (B) tasks. Error bars represent ±1 standard error of the mean.

Table 2
Mean Performance for the Transfer Tasks as a Function of Training Group and Session

	Control			Visual search			Dual *n*-back		
Task	Pre	Mid	Post	Pre	Mid	Post	Pre	Mid	Post
RAPM	6.65 (2.18)	6.45 (2.50)	6.00 (3.00)	6.52 (3.04)	6.07 (2.87)	6.24 (3.34)	7.04 (2.48)	6.17 (2.28)	6.25 (3.08)
RSPM[a]	17.15 (2.39)	15.85 (2.50)	16.85 (2.35)	16.66 (2.53)	16.34 (2.30)	16.45 (2.47)	16.30 (2.67)	16.74 (2.54)	16.09 (2.61)
Cattell	11.95 (2.63)	11.75 (2.05)	11.45 (2.65)	10.72 (2.79)	11.07 (2.07)	11.24 (2.25)	12.00 (2.38)	11.71 (2.29)	11.38 (2.45)
Paper Folding	4.05 (1.70)	4.50 (1.43)	4.00 (1.26)	4.41 (1.38)	4.00 (1.60)	4.52 (1.34)	3.79 (1.47)	4.46 (1.69)	4.33 (1.34)
Letter Sets	6.85 (1.90)	6.75 (2.29)	6.80 (2.22)	7.79 (1.84)	6.90 (2.16)	6.83 (2.19)	7.08 (2.45)	7.17 (1.52)	7.04 (2.14)
Number Series	4.20 (0.83)	3.75 (0.85)	3.70 (1.22)	3.59 (1.32)	3.76 (1.35)	3.52 (1.24)	3.96 (0.96)	3.92 (1.18)	3.75 (1.19)
Inferences	4.35 (1.31)	4.30 (1.78)	4.45 (1.54)	4.41 (1.40)	4.03 (1.84)	4.24 (1.60)	3.67 (1.97)	4.04 (1.60)	4.04 (1.65)
Analogies	4.90 (1.59)	4.65 (1.42)	3.90 (1.71)	4.83 (1.65)	4.45 (1.64)	4.38 (1.66)	4.46 (1.62)	4.46 (1.53)	3.79 (1.50)
SynWin	352.40 (626.95)	682.50 (190.05)	701.50 (214.60)	461.14 (252.22)	625.76 (205.12)	729.14 (193.33)	480.28 (218.08)	581.88 (231.66)	655.08 (201.34)
Control Tower[b]	29.62 (10.45)	32.47 (11.22)	34.05 (11.02)	29.90 (11.38)	29.63 (13.87)	37.43 (15.63)	29.20 (8.10)	31.41 (11.54)	34.26 (11.29)
ATClab[c]	0.72 (0.09)	0.73 (0.12)	0.72 (0.12)	0.73 (0.14)	0.71 (0.12)	0.75 (0.12)	0.74 (0.13)	0.75 (0.09)	0.75 (0.12)
Symmetry Span	25.60 (9.30)	30.25 (9.34)	28.90 (12.14)	24.55 (11.15)	27.28 (12.53)	26.76 (10.99)	25.88 (8.75)	32.29 (9.92)	31.54 (11.80)
Running Span	38.50 (8.65)	40.90 (8.81)	43.00 (9.06)	39.52 (13.03)	39.52 (12.42)	42.34 (12.49)	37.96 (12.69)	40.13 (10.86)	42.21 (11.94)
Vocabulary	10.10 (1.21)	10.70 (1.08)	10.35 (1.79)	10.00 (1.79)	10.38 (1.43)	9.79 (1.74)	10.33 (1.13)	10.04 (1.40)	10.50 (1.41)
Knowledge	6.75 (1.83)	6.30 (2.18)	6.20 (1.61)	5.90 (1.99)	6.17 (1.91)	6.10 (1.78)	6.25 (2.03)	6.04 (1.23)	6.29 (1.81)
Letter Comparison	18.75 (3.77)	20.65 (3.50)	20.85 (3.72)	19.93 (4.08)	20.45 (5.68)	20.45 (5.38)	19.04 (4.84)	19.92 (4.03)	21.38 (3.61)
Number Comparison	28.90 (5.21)	31.15 (4.61)	31.00 (4.24)	29.14 (6.11)	29.52 (5.92)	29.93 (7.02)	28.83 (5.54)	28.58 (4.95)	29.00 (5.43)

Note. Standard deviations are shown in parentheses. Pre = pretest; mid = midtest; post = posttest; RAPM = Raven Advanced Progressive Matrices; RSPM = Raven Standard Progressive Matrices.
[a] $N = 23$ for dual *n*-back group due to experimenter error during midtest session. [b] $N = 19$ for dual *n*-back group due to computer problem during posttest session. [c] $N = 28$ for visual search group and $N = 19$ for control group due to computer problem during posttest session.

Table 3
Significance Testing Results for the Transfer Measures

Task	Group			Session			Group × Session		
	F	*p*	η_p^2	*F*	*p*	η_p^2	*F*	*p*	η_p^2
Fluid intelligence (spatial)									
Raven Advanced	0.05	.95	.001	2.00	.14	.028	0.33	.86	.009
Raven Standard	0.06	.94	.002	1.53	.22	.022	2.85	.03	.076
Cattell	1.08	.35	.030	0.28	.75	.004	0.92	.45	.026
Paper Folding	0.11	.90	.003	0.75	.47	.011	1.92	.11	.052
Fluid intelligence (verbal/numeric)									
Letter Sets	0.31	.74	.009	1.14	.32	.016	1.08	.37	.030
Number Series	0.70	.50	.020	1.56	.21	.022	0.78	.54	.022
Inferences	0.70	.50	.020	0.20	.82	.003	0.67	.61	.019
Analogies	0.46	.64	.013	*6.08*	*.00*	*.080*	0.69	.60	.019
Multitasking									
SynWin[a]	0.16	.85	.005	*29.95*	*.00*	*.300*	1.81	.13	.049
Control Tower	0.26	.98	.001	*17.28*	*.00*	*.200*	1.96	.10	.054
ATClab	0.23	.78	.007	0.32	.73	.005	0.80	.53	.023
Working memory capacity									
Symmetry Span	1.02	.37	.028	*10.30*	*.00*	*.128*	0.70	.59	.020
Running Letter Span	0.24	.98	.001	*8.64*	*.00*	*.110*	0.38	.82	.011
Crystallized intelligence									
Vocabulary	0.57	.57	.016	0.68	.51	.010	1.62	.17	.044
General Knowledge	0.37	.70	.010	0.17	.85	.002	0.68	.61	.019
Perceptual speed									
Letter Comparison	0.02	.98	.001	*5.55*	*.01*	*.073*	0.98	.42	.027
Number Comparison	0.58	.57	.016	1.54	.22	.022	0.75	.56	.021

Note. Entries in italics indicate values significant at $\alpha = .01$.
[a] At the Michigan State University testing location, 21 subjects were administered the same test version of SynWin at pre-, mid-, and posttest. Data were reanalyzed with subjects who only performed unique versions of SynWin at all three transfer sessions ($N = 14$, 21, and 17 for control, visual search, and dual *n*-back, respectively). The interpretation of the significance tests was the same as listed above with the full data.

Table 4
Inferential Results of the Transfer Composite Standardized Gain Scores

Construct	Midtest	Posttest
Fluid intelligence (spatial)	$F(2, 69) = 0.83, p = .44$	$F(2, 70) = 0.39, p = .68$
Fluid intelligence (verbal)	$F(2, 70) = 1.51, p = .23$	$F(2, 70) = 0.62, p = .54$
Multitasking	$F(2, 70) = 3.09, p = .05$	$F(2, 67) = 1.44, p = .24$
Working memory	$F(2, 70) = 1.89, p = .16$	$F(2, 70) = 0.89, p = .41$
Crystallized intelligence	$F(2, 70) = 1.44, p = .24$	$F(2, 70) = 0.16, p = .86$
Perceptual speed	$F(2, 70) = 1.15, p = .32$	$F(2, 70) = 0.86, p = .43$

Table 5
Posttest Survey Data

Topic	Dual *n*-back (%)	Visual search (%)	Control (%)	$\chi^2(2)$	*p*
Attention	52	72	50	3.27	.20
Intelligence	65	41	30	5.73	.06
Language	4	3	10	1.06	.59
Memory	78	45	40	8.01	.02
Perception	35	59	45	2.98	.23
Daily activities	43	10	10	10.51	<.01

Note. Due to experimenter error, survey data were not available for one dual *n*-back subject and two control subjects. However, survey data were included for the two control subjects who received the same transfer test items at pretest and posttest. The format of the question for each topic was "Do you feel that your participation in this study has changed your ___ ?"

Discussion

- Performance improvements on dual n-back and visual search tasks, but **no positive transfer** to intelligence, multitasking, working memory (WM) capacity & perceptual speed tasks.
- No evidence of dose-dependent relationship between **amount** of dual n-back training and fluid intelligence gains - compared to Jaeggi et al. (2008) findings of big improvement from pre- to post-test.

WM Training and Transfer to Fluid Intelligence

- Review by Morrison & Chein (2011) – 4 studies reporting significant transfer to reasoning and intelligence, 3 studies reporting no significant transfer, one study reporting significant transfer to some intelligence measures.
- But small sample sizes (n=3/4), some subjects reused, unwitting bias from file-drawer problem.
- Age – older adults/developmental ages/patients ✉ most show no transfer.
- Meta-analysis by Melby-Lervag & Hulme (2012) – few studies show evidence of transfer from WT training to fluid intelligence, age not significant.
- Data from this study congruous – no transfer when compared to no-contact control or active-control group.

Limitations

- 3 of fluid intelligence tasks – mean pretest scores close to maximum.
- Did not affect interpretation of other 14 transfer measures – no ANOVAs after removing significant.
- Difficult to assess reliability of shortened intelligence measures.
- Illusory placebo effect – trained subjects reported subjective improvement (questionnaire) in absence of objective improvement.

Variables That Affect Transfer

- Variables – amount of n-back improvement, pre-training ability level, sample size, number/duration of training sessions, transfer tests used and method of administration, session spacing, subject motivation, experimenter influences.
- Recent study (Jaeggi et al., 2011) – amount of n-back improvement is critical variable determining transfer to intelligence.
- Children with biggest gain showed transfer relative to active-control – own results do no not show this for high improvement group.

Future Work

- Understand what different WM processes affected by n-back.
- Why only some individuals benefit from training intervention.
- Whether amount of training improvement affects amount of intelligence transfer.
- Whether certain training methods are more effective for certain individuals (based on differences in pre-training fluid intelligence and WM capacity).
- Optimal number of training sessions.
- Underlying mechanisms responsible for transfer.