Using working memory intervention to improve math performance: ill-conceived, poorly executed, or just not quite there yet?

Kerry Lee
The Education University of Hong Kong

Contents

Working Memory

- What is working memory?
  - The capacity to process and remember information at the same time
- When do we use our working memory?
  - Now!
    - Listening comprehension makes intensive use of working memory
      - Perceive what is said
      - Retain what is said and combine with previously spoken information
      - Deciphering the meaning of what is just said in the context of what was previously said

Working memory at work

- O
- 2
- A
- N
- S
- 1
- R
- M

Working memory at work

- N
- O
- R
- S
- M
- A
- 2
- 1

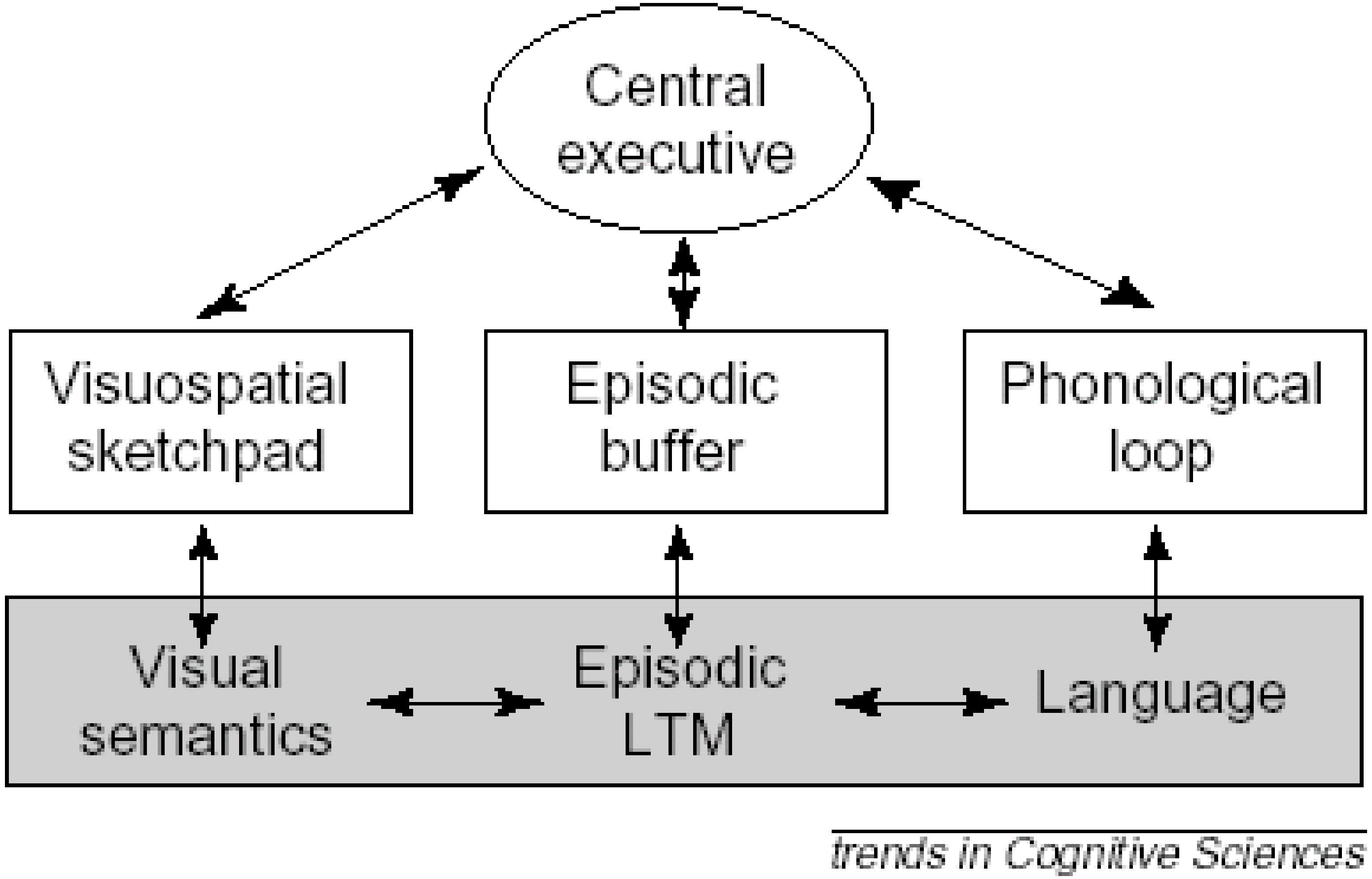
Working memory at work

- O
- 2
- A
- N
- S
- 1
- R
- M

- N
- O
- R
- S
- M
- A
- 2
- 1

Working memory

The capacity to process and remember information at the same time

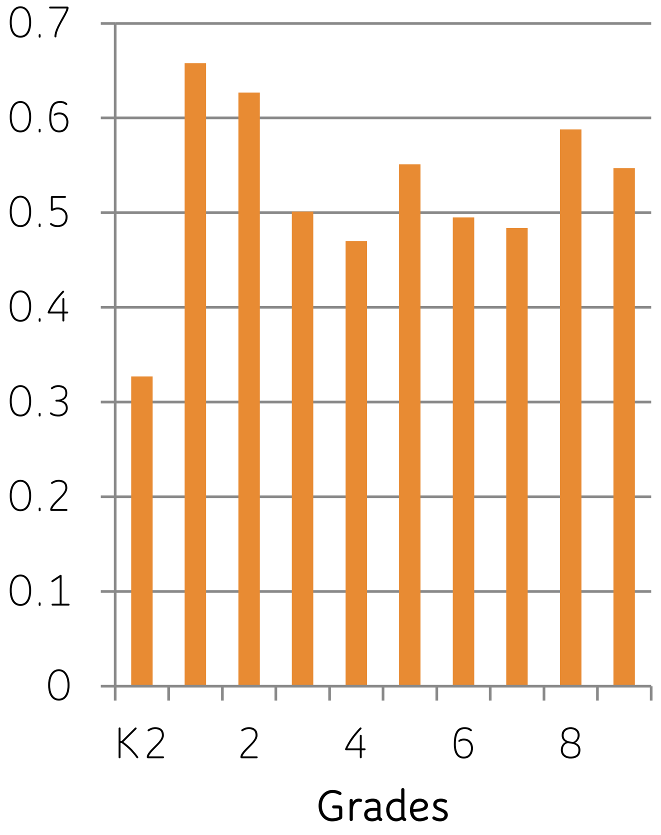


Baddeley (2000)

Why is working memory important?

- Strong associations between WM & academic performances
  - Mathematics (for reviews, see Bull & Lee, 2014; Friso-van den Bos et al., 2013)
  - Reading (Follmer, 2018; Peng et al., 2018)
- Standardised working memory scores predicted children's academic standing in mathematics with 83% accuracy (Gathercole and Pickering, 2000)

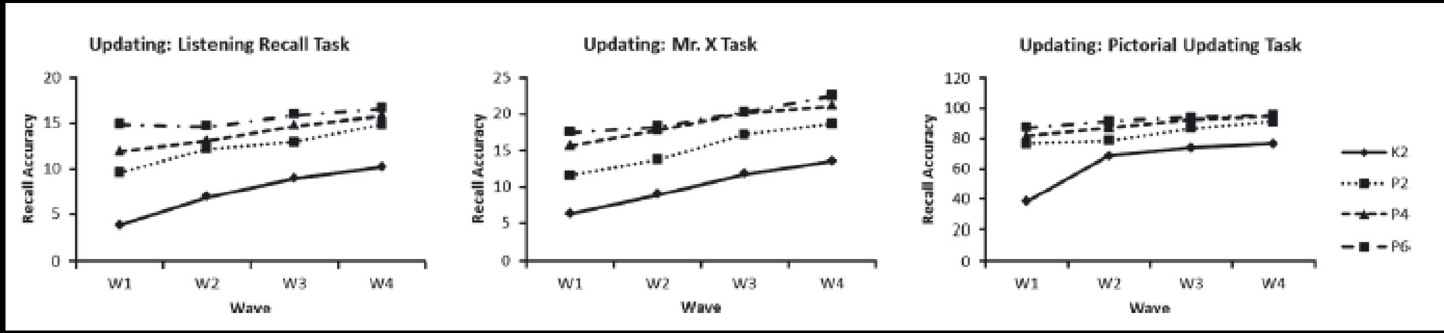
Correlation ratios between working memory capacity and math achievement



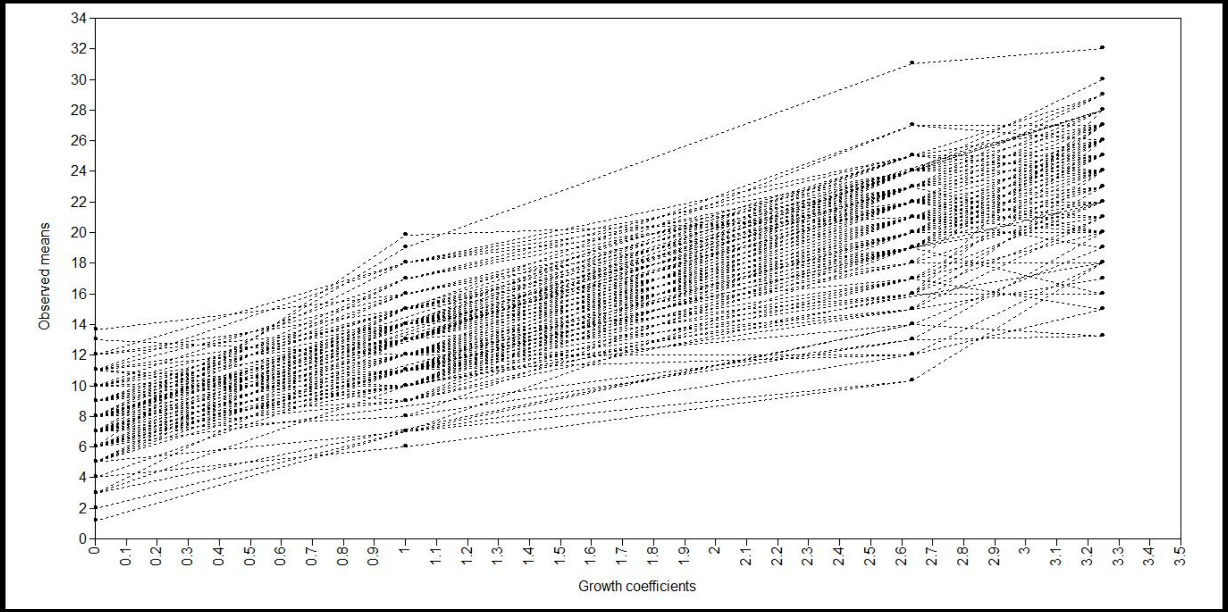
From Lee & Bull (2016)

Age & individual differences

- Capacity increases significantly from early childhood to the adolescent years (Gathercole et al., 2004; Lee et al., 2013)



From Lee et al., (2013)



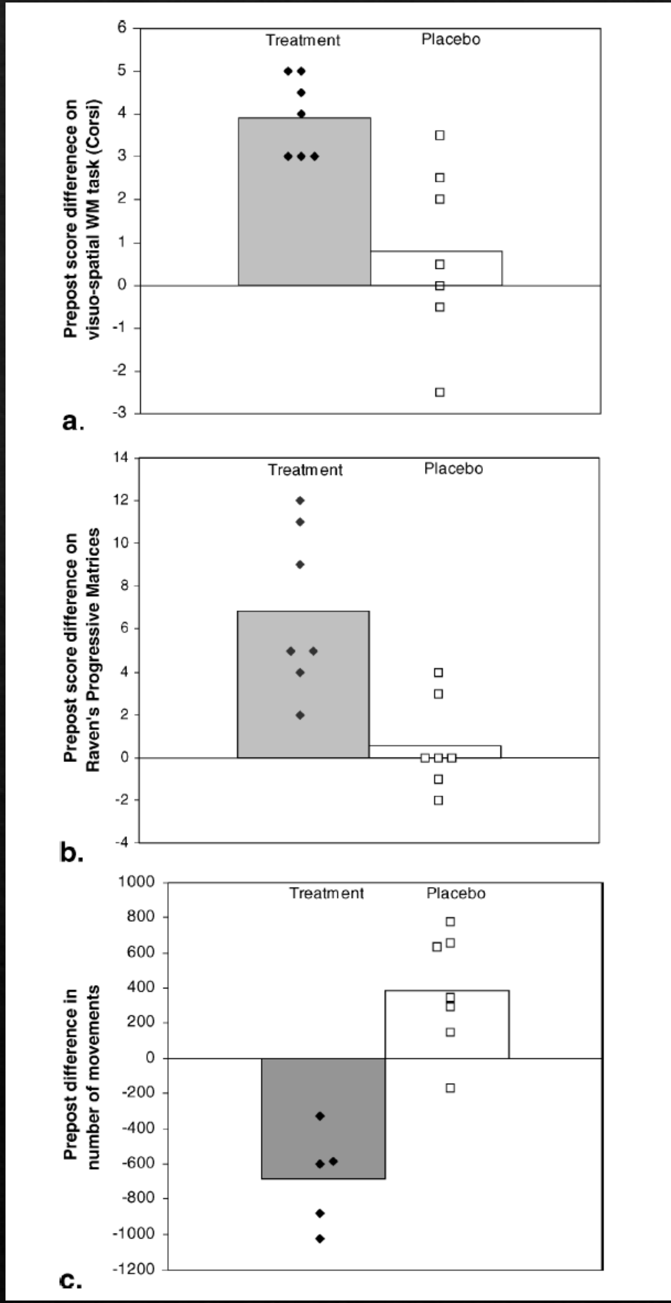
- Individual differences can be observed even amongst kindergarten children

Improving Working Memory

The beginning

Klingberg et al., 2002

- Experiment 1
  - 7-15-year-olds with ADHD (*N* = 14) assigned to treatment and control
  - Pretest & posttest
    - Visuospatial WM (trained & non-trained), Stroop, Raven's, choice RT, head movement
  - Training
    - Visuospatial WM, backward digits, letter span, choice RT @30 trials/day for ~24 days, increasing difficulties
  - Control
    - Same tasks @10 trials/day for ~24 days, constant low level of difficulties



Klingberg et al., 2002

- Experiment 2
  - 20 to 29-year-olds university students (*N* = 4)
    - No control group
  - Pretest & posttest
    - Similar to Experiment 1
  - Training
    - Same as Experiment 1 for ~26 days

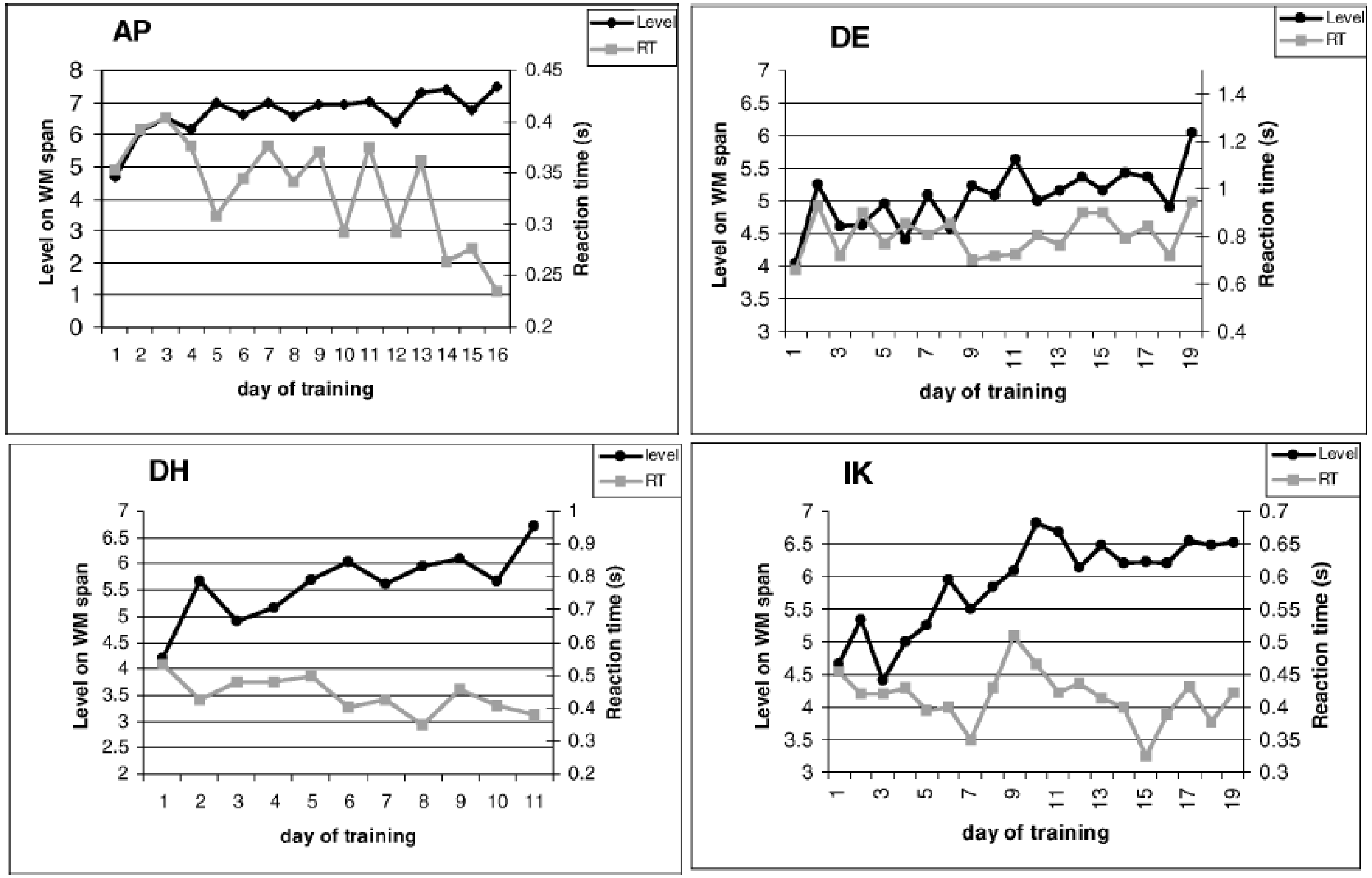
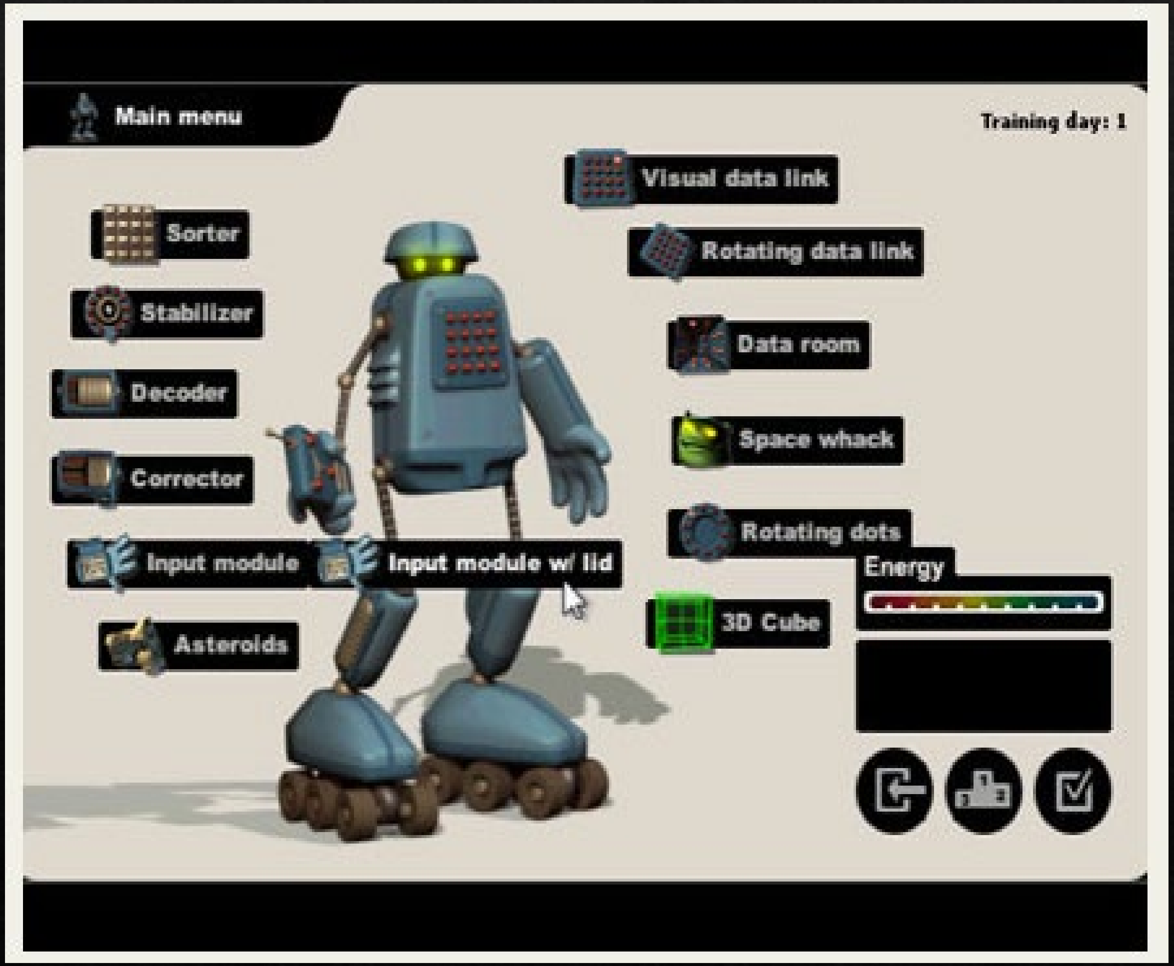


Fig. 2. Improvement during training of the visuo-spatial WM task in four adult subjects. Each graph shows data from a single individual.

CogMed


Main menu
Training day: 1
Visual data link
Sorter
Rotating data link
Stabilizer
Data room
Decoder
Space whack
Corrector
Rotating dots
Input module
Input module w/ lid
Energy
3D Cube
Asteroids


Level
5
4
3
2
1
Done
Cancel
Instructions


1 2 3
4 5 6
7 8 9
Level
5
4
3
2
1
Done
Cancel
Instructions

Successes & Disappointments

The search for transfer to math performance

Successes I

- Holmes et al., 2009
  - 10-year-olds (*N* = 42) with low WM assigned to treatment and control
  - Pretest, posttest, delayed
    - Automated Working Memory Assessment ; WASI; WORD; WOND
  - Training
    - CogMed; adaptive
    - 35 min/day ~ total 20 days
  - Control
    - Non-adaptive version of training
      @ same dosage as training group

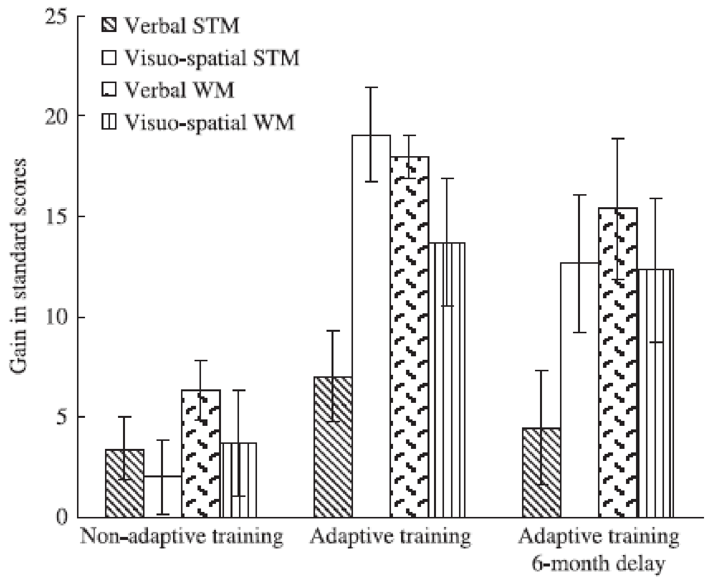


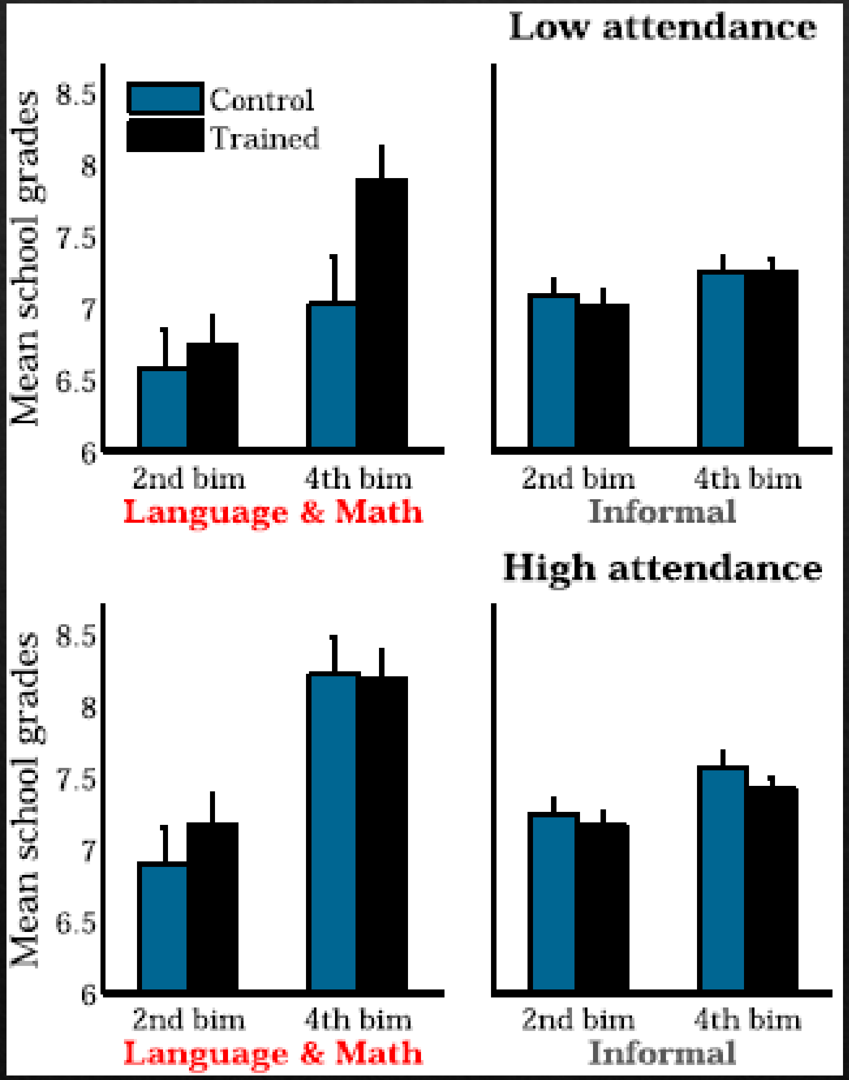
Figure 1 *Impact of training on working memory.*

Table 1 *Impact of training on cognitive measures*

	Non-adaptive						Adaptive									
	Pre-training		Post-training		Pre- to post-training		Pre-training		Post-training		Pre- to post-training		6-mth follow-up		Pre- to follow-up	
	Mean	*SD*	Mean	*SD*	*p*	*d*	Mean	*SD*	Mean	*SD*	*p*	*d*	Mean	*SD*	*p*	*d*
Verbal STM	87.25	8.18	90.68	5.85	0.04	0.47	89.82	9.44	96.82	12.18	0.01	0.62	95	13.67	0.13	0.44
Visuo-spatial STM	92.78	10.47	94.8	8.79	0.29	0.21	83.36	13.98	102.4	11.49	<.01	1.2	96.05	14.12	<.01	0.83
Verbal WM	82.85	1.47	89.18	9.54	<.01	0.78	78.3	6.58	96.27	9.66	<.01	1.55	93.68	14.84	<.01	1.16
Visuo-spatial WM	84.73	13.42	88.43	10.97	0.18	0.3	80.2	9.15	93.91	13.55	<.01	1.03	92.63	16.86	<.01	0.85
Following instructions (max. 30)	16.85	4.57	16.25	4.3	0.52	0.14	14.45	4.02	18.27	4.37	<.01	0.83	16.5	3.59	0.04	0.52
Verbal IQ	94.7	14.9	92.85	10.9	0.36	0.14	88.73	11.14	90.86	11.52	0.2	0.19	92.78	9.1	0.08	0.39
Performance IQ	88.65	12.1	93.65	11.1	0.01	0.43	88.05	13.09	90.68	12.96	0.11	0.2	87.11	9.07	0.72	0.08
Basic word reading	83.3	14.2	83.9	13.3	0.68	0.04	83.68	12.35	83	15.06	0.64	0.05	82.83	14.14	0.36	0.07
Mathematical reasoning	93.4	15.8	96.4	16.2	0.18	0.19	84.27	12.28	85.68	12.7	0.31	0.11	89.94	9.88	0.01	0.49

Successes II

- Goldin et al., 2014
  - 6 to 7-year-olds (N = 111) assigned to treatment and control
  - Pretest & posttest
    - Attentional Network Test, Stroop, Tower of London (no WM), results on school administered tests
  - Training
    - Matamarote: separate working memory, planning, and inhibition games; adaptive
    - Children played one of the games for 15 min/day – total 27 days
  - Control
    - Commercially available computer games @ same dosage as training group



Disappointments I

- Alloway et al., 2013
  - 10 to 11-year-olds (*N* = 94) with learning difficulties assigned to treatment and control
  - Pretest, posttest, delayed
    - Automated Working Memory Assessment ; WASI; WORD; WOND
  - Training
    - Jungle Memory; adaptive
    - 15 min/day; four time/week ~ 8 weeks
  - Control
    - Passive
    - Active; same training game but @ once/week ~ 8 weeks

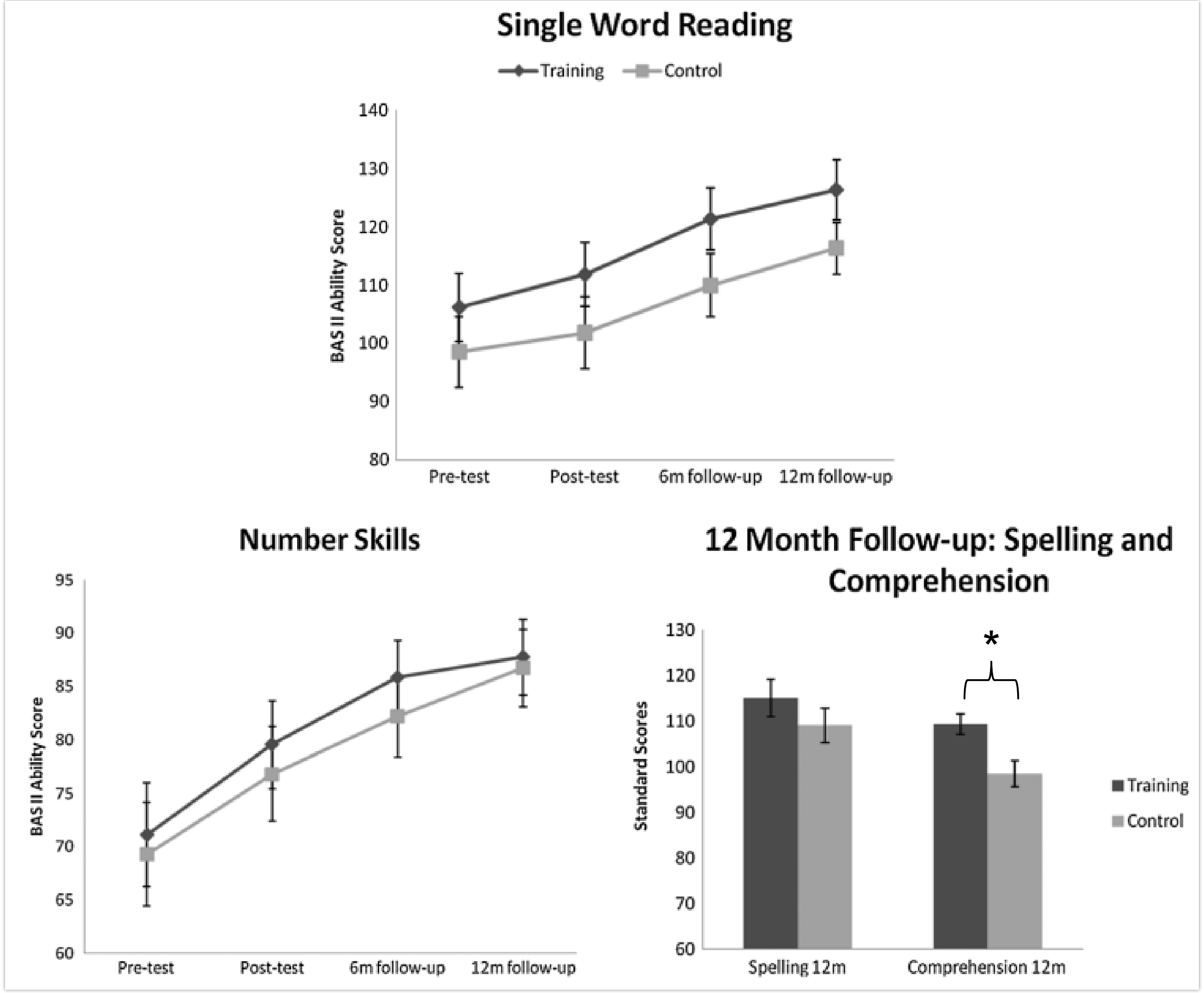
Table 2
Difference between pre-test scores and post-test scores (Time 2–Time 1) as a function of group.

Measures	Nonactive control	Active control (WMT-Low)	Training group (WMT-High)
Verbal WM	1	4	15*
Visuo-spatial WM	–2	–1	12*
IQ: vocab (Verbal)	0	–4	7*
IQ: matrix (Nonverbal)	–	–	10*
Spelling	–1	0	4*
Math	0	–1	2

Note: Data for the Matrix test are only available for the WMT-High group.
* Indicates a significant difference between pre- and post-training scores ($p < .01$).

Disappointments II

- Henry, Messer et al. 2014
  - 5- to 8-year-old typically developing (N = 36)
  - 6 weeks (thrice weekly @10 min) WM intervention based on adaptive practices on WM tasks
  - Active control with the same processing component as intervention
  - Near transfer to WM tasks but no far transfer to number skills



Disappointments III

- Roberts et al., 2016
  - 6 to 7-year-olds (*N* = 452) with low WM assigned to treatment and control
  - Pretest, posttest, delayed
    - Automated Working Memory Assessment ; Wide Range Achievement Test; WASI
  - Training
    - CogMed; adaptive
  - Control
    - Passive

Table 3. Outcome Comparisons At 6, 12, and 24 Months

Outcome Variable	Mean (SD) Score: Intervention Arm[a]	Mean (SD) Score: Control Arm[a]	Adjusted Mean Difference (95% CI)[b]	P Value
6-mo Outcomes				
AWMA				
Digit recall	103.2 (13.9)	102.2 (13.1)	0.19 (−2.02 to 2.38)	.87
Dot matrix	101.4 (15.4)	95.7 (14.8)	5.47 (2.87 to 8.07)	<.001
Mister X	105.1 (15.5)	107.0 (15.9)	−2.33 (−5.14 to 0.47)	.10
Backward Digit Recall	103.5 (16.8)	100.5 (13.0)	2.91 (0.02 to 5.79)	.04
WASI-2 IQ	98.4 (13.4)	97.4 (11.8)	0.78 (−1.57 to 3.13)	.51
12-mo Outcomes				
WRAT4 (primary outcome)				
Word reading	103.8 (14.7)	105.7 (13.4)	−1.81 (−3.78 to 0.15)	.07
Sentence comprehension	103.4 (15.9)	105.4 (16.4)	−2.02 (−4.79 to 0.73)	.15
Spelling	102.5 (17.0)	104.6 (16.7)	−1.92 (−4.42 to 0.57)	.13
Math computation	91.5 (14.4)	94.3 (16.4)	−2.64 (−5.48 to 0.20)	.07
AWMA				
Digit recall	103.6 (15.2)	103.3 (12.7)	−0.42 (−2.52 to 1.67)	.69
Dot matrix	102.7 (15.8)	96.03 (14.6)	7.78 (4.41 to 11.14)	<.001
Mister X	105.3 (15.2)	106.6 (15.5)	−0.98 (−4.43 to 2.48)	.57
Backward Digit Recall	103.3 (14.2)	102.0 (14.2)	1.80 (−0.85 to 4.46)	.18
Connors 3 ADHD Index *T* score	59.7 (17.1)	58.2 (16.4)	0.32 (−4.29 to 4.93)	.89
SDQ total difficulties	8.5 (5.4)	7.7 (5.0)	1.02 (−0.20 to 2.24)	.10
PedsQL				
Psychosocial health	75.7 (14.4)	77.2 (14.9)	−2.37 (−5.66 to 0.92)	.15
Physical health	81.1 (18.8)	83.9 (17.1)	−4.29 (−8.60 to 0.02)	.05
Teacher-reported measures				
ARS language and literacy	4.2 (0.9)	4.2 (0.8)	−0.04 (−0.17 to 0.09)	.56
ARS mathematical thinking	4.1 (0.9)	4.2 (0.9)	−0.05 (−0.20 to 0.09)	.48
Approach to learning	3.2 (0.8)	3.3 (0.7)	−0.08 (−0.20 to 0.04)	.20
24-mo Outcomes				
WRAT4 (primary outcome)				
Word reading	101.1 (14.7)	103.0 (13.4)	−1.97 (−4.27 to 0.32)	.09
Spelling	103.4 (17.0)	105.7 (15.8)	−2.43 (−5.45 to 0.60)	.11
Math computation	93.8 (15.6)	96.6 (15.5)	−3.03 (−5.39 to −0.67)	.01
AWMA				
Digit recall	101.6 (17.1)	103.9 (15.6)	−2.60 (−5.80 to 0.60)	.11
Dot matrix	99.3 (16.1)	97.0 (15.9)	2.96 (−1.03 to 6.95)	.14
Mister X	100.6 (13.2)	102.4 (13.3)	−1.52 (−4.81 to 1.78)	.36
Backward Digit Recall	101.5 (16.9)	101.7 (15.2)	0.29 (−2.12 to 2.70)	.81
CHU-9D	0.8 (0.1)	0.8 (0.1)	−0.01 (−0.04 to 0.02)	.39
SDQ total difficulties	9.2 (5.8)	8.3 (5.9)	0.28 (−1.35 to 1.90)	.73
PedsQL				
Psychosocial health	74.7 (15.4)	75.9 (15.4)	−0.39 (−4.05 to 3.26)	.83
Physical health	83.3 (17.2)	81.7 (19.5)	1.98 (−1.87 to 5.83)	.31

Updating/WM intervention

1st iteration

Disappointments III

- Ang et al., 2019
  - 6 to 7-year-olds (N = 70) with learning difficulties in math assigned to treatment and control
  - Pretest & posttest
    - Working Memory Test Battery for Children ; Updating; WISC; WIAT; Schonell; BLAB
  - Training
    - Four adaptive games based on the running span and keep track paradigms
    - 30 min/day; twice/week ~ 10 weeks
  - Control
    - Passive
    - Active; same dosage as training but no mnemonic component

Table 11.1 Means and standard deviations (in parentheses) of the outcome measures

	Intervention		Active control		Passive control	
Task	Pre-test	Post-test	Pre-test	Post-test	Pre-test	Post-test
Pictorial Updating	47.12 (12.71)	55.64 (10.20)	47.92 (15.34)	49.25 (16.99)	51.05 (19.50)	54.81 (14.06)
Listening Recall	6.28 (2.61)	7.52 (2.567)	4.54 (3.41)	6.67 (2.57)	5.81 (3.93)	8.62 (3.04)
Backward Digit Recall	8.38 (2.34)	9.36 (3.01)	8.08 (4.20)	10.25 (3.45)	8.95 (4.20)	11.00 (3.87)
Block Recall	22.24 (5.00)	22.08 (4.723)	21.38 (3.90)	20.96 (4.52)	21.86 (3.68)	22.52 (3.37)
Digit Recall	26.60 (4.31)	28.85 (6.65)	25.38 (4.54)	27.75 (4.87)	25.95 (4.96)	27.95 (5.59)
Numerical Operations	13.52 (2.87)	17.58 (2.80)	12.71 (4.39)	16.89 (4.12)	14.38 (3.37)	18.52 (4.46)
Math Problem Solving	27.72 (3.84)	30.92 (3.81)	26.13 (4.45)	30.78 (4.84)	29.00 (4.52)	31.38 (5.11)
Fluency – Addition	8.84 (5.81)	13.00 (5.32)	7.75 (6.60)	13.35 (6.58)	7.81 (6.03)	13.38 (6.45)
Fluency – Subtraction	4.16 (3.02)	8.25 (4.50)	4.42 (5.69)	8.61 (4.58)	3.76 (4.39)	8.48 (4.09)
Block Design	12.60 (8.49)	19.76 (9.58)	16.42 (10.07)	21.39 (11.88)	18.00 (11.64)	24.10 (10.43)
Vocabulary	6.72 (4.112)	7.80 (4.35)	5.42 (4.28)	7.13 (4.96)	7.52 (4.86)	7.57 (3.97)

Scores in the table are raw scores

Getting slightly warmer

- Ang et al., 2015
  - 6 to 7-year-olds (N = 111) with learning difficulties in math assigned to treatment and control
  - Pretest, posttest, delayed
    - Updating; Letter Rotation; Block Recall; Backward Digit; WIAT; Raven's; Schonell;BLAB
  - Training
    - Seven adaptive games based on the running span and keep track paradigms
      - 30 min/day; 3-4/week ~ 8 weeks
    - CogMed
      - 45min/day; 3-4/week ~ 8 weeks
  - Control
    - Passive
    - Active; same dosage as training but no mnemonic component

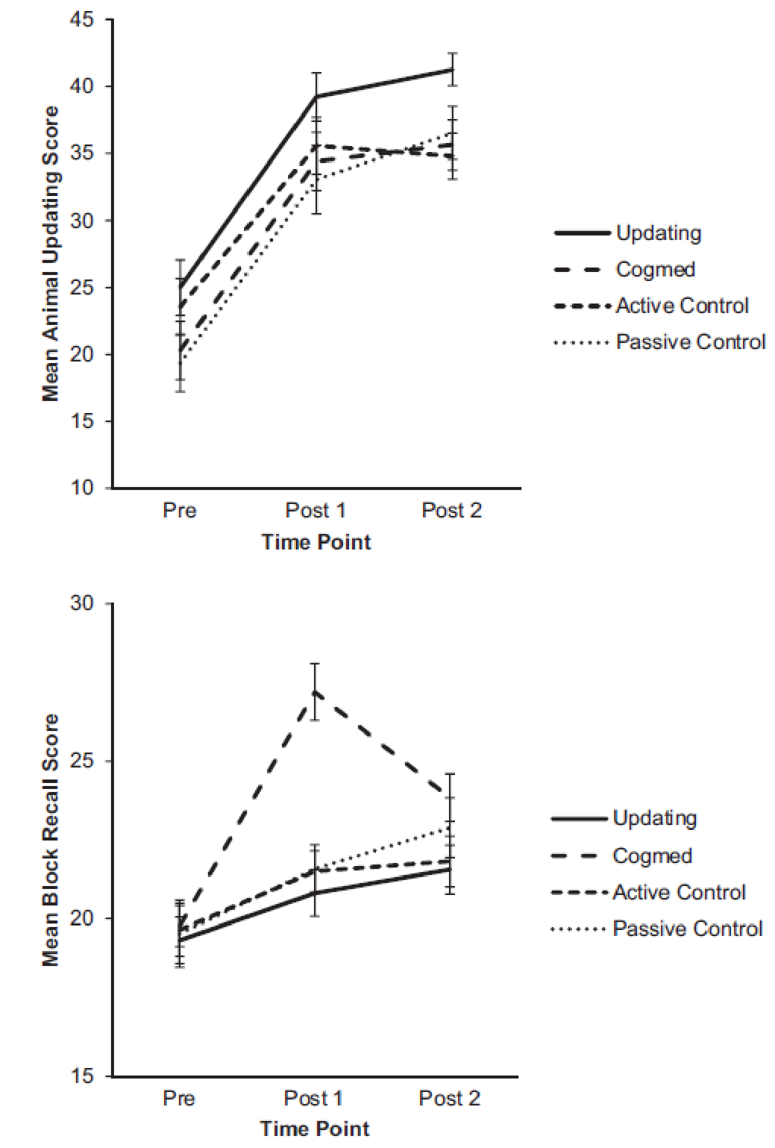
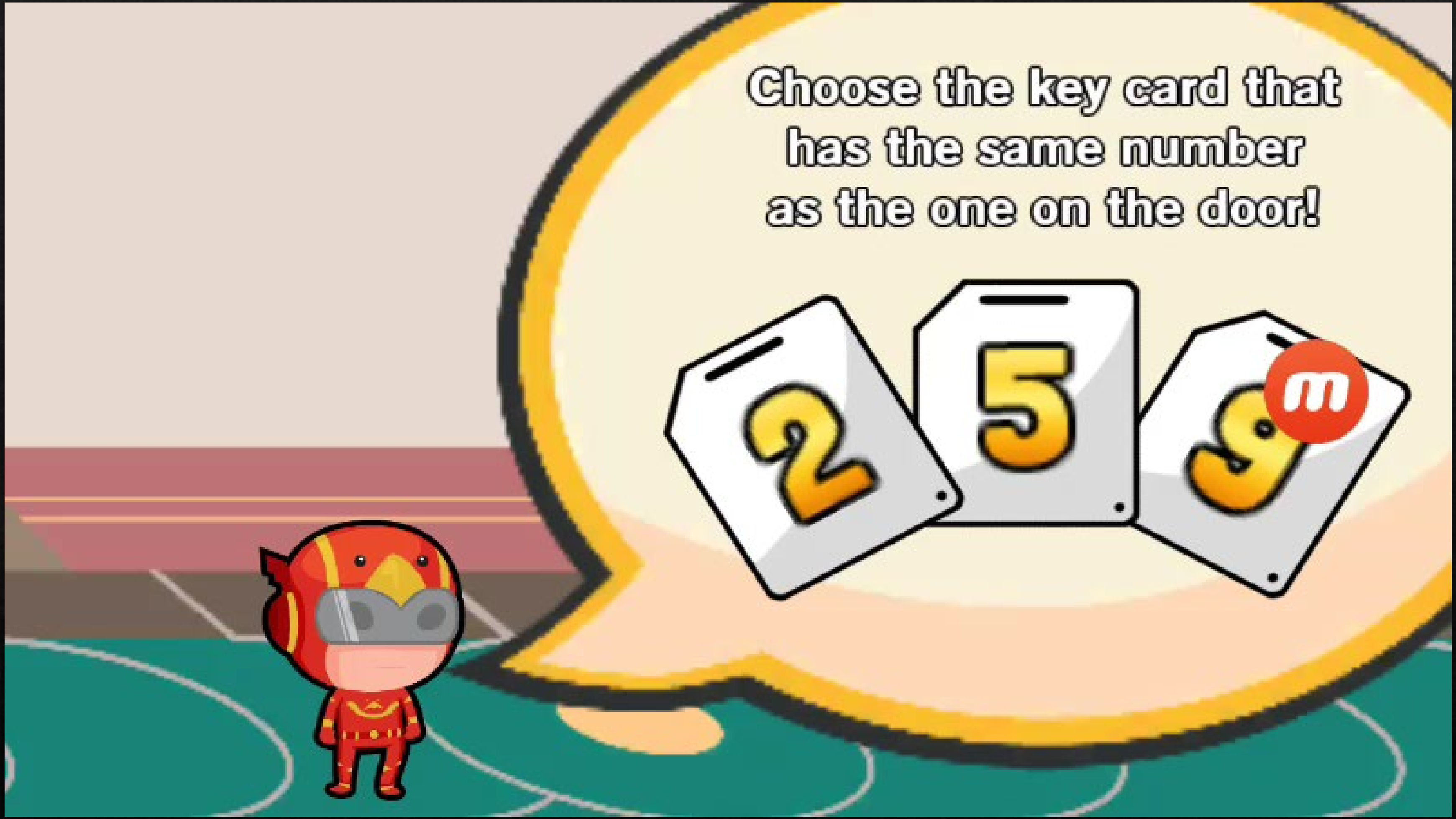


Fig. 2. Mean Animal Updating and Block Recall scores by condition and time of test. The error bars depict standard errors.

Updating/WM intervention

3rd iteration



A lot warmer

- Munez, et al. (under review)
  - 6 to 7-year-olds (N = 428) with learning difficulties in math assigned to treatment (x3) and control
  - Pretest, posttest, delayed
    - Running Letters, Keep Track, N-Back, and Complex Span task
    - Number line, numerical discrimination, WIAT
  - Training
    - Non-numeric working memory based on four paradigms as used in tests
    - Numeracy
    - Numeric working memory
      - 10-15 min/day; 2-3/week ~ 40 weeks
  - Control
    - Active; same dosage as training, similar to non-numeric working memory training but no mnemonic component

Table 2: Parameter estimates of Model_2 (and Hedges' g)

	WM-Updt	Math Pr Solving	Num Ops	Math Fluency	N-Line (0-10)	N-Line (0-100)	Num Discr
Means							
Normative Intercept	-.01	26.00***	13.82***	14.70***	14.17***	18.85***	64.41***
Normative Linear Slope	.89***	4.18***	3.66***	4.33***	-2.65***	-1.86**	6.89***
Normative Quadr Slope	-.07***	-.44***	-.40***	-.01	.37***	.01	-1.13***
Variances							
Normative Intercept	.35***	14.10***	10.83***	97.50***	12.79**	40.98***	126.02***
Normative Linear Slope	.02	1.13	.75	2.87**	3.73*	.89	11.29
Normative Quadr Slope	-[a]	.06	.05	-[a]	.11*	-[a]	.29
Covariances							
Intercept with L-Slope	.04**	-1.86	.69	4.21**	-2.99	-3.77**	-19.08
Treatment effects (means)							
T1_WM	.03 (.05)	.09 (.02)	.05 (.02)	.47 (.05)	.13 (.04)	-.22 (.03)	-.95 (.08)
T1_NWM	-.01 (.01)	-.05 (.01)	-.12 (.03)	.62 (.06)	-.48* (.14)	-.33 (.05)	-.71 (.06)
T1_NUM	-.02 (.03)	.14 (.04)	.03 (.01)	.74 (.07)	-.46* (.13)	-.58 (.09)	-1.02 (.09)
T2_WM	-.07 (.12)	-.21 (.05)	-.16 (.05)	-.57 (.06)	-.72 (.20)	.45 (.07)	1.91 (.17)
T2_NWM	.02 (.04)	.02 (.01)	-.10 (.03)	.52 (.05)	-.12 (.03)	1.43* (.22)	2.42* (.20)
T2_NUM	.01 (.02)	-.13 (.04)	.37 (.11)	-.07 (.01)	-.13 (.04)	1.14* (.18)	2.47* (.22)
Treatment effects(variances)							
T1_WM	.02	.30	.53	1.56	.87	1.08	5.37
T1_NWM	.06**	.04	.27	3.28	.28	2.82*	1.70
T1_NUM	.01	.05	.11	.191	.09	.20	7.97
T2_WM	.00	1.40	.15	.86	4.88	.37	51.75
T2_NWM	.01	.46	1.85	2.03	.13	1.44	1.16
T2_NUM	.02	.14	.18	.51	.53	.15	44.24

Note: *** $p < .001$; ** $p < .01$; * $p < .05$; T1 and T2 correspond to immediate effect (pre- to post-test) and delayed or long-term effect (post-test to delayed post-test), respectively.

[a] Indicates that the variance of the quadratic factor was not estimated due to convergence problems.

The reviews

Shipstead et al., 2012

- Critiques
  - Use of single tasks to measure WM capacity or other outcome constructs
  - Use of transfer tests that are very similar to the method of training
  - Conflation of STM with WM
  - Lack of proper control – from no control group to passive and no-adaptive control
  - Use of subjective reports as outcome measures

The Reviews – negative findings

- Hulme and Melby-Lervåg 2012; Melby-Lervåg and Hulme 2013; Melby-Lervåg et al., 2016
  - Evidence of short-term near and intermediate transfer, some evidence of long-term facilitation for visuospatial working memory, but no far transfer
- Redick et al. 2015
  - Provided additional criteria of rigour
    - Need to show near transfer as bases for far transfer
    - Transfer as the result of improvement in the treatment group (rather than deterioration in the control)
  - Studies considered more rigorous – no transfer to math
  - Studies considered less rigorous – mixed findings on transfer to math

The Reviews – negative findings

- Sala & Gobet, 2017
  - Focused on studies conducted with typically developing children
  - Significant near or intermediate transfer, some evidence of far transfer to math but deemed likely due to methodological issues (e.g., contrast with passive rather than active control, lack of randomised assignment to group)
- Dougherty and Engle 2019
  - Don’t buy the snake oil

The Reviews – positive findings

- Pappa, Biswas et al. 2020
  - Based on neuromodulation evidence but still no far transfer (adults
- Peijnenborgh et al., 2016
  - Focused on studies conducted with children with learning disabilities (mostly ADHD)
  - Significant near or intermediate transfer but evidence of far transfer for word decoding only, not for arithmetic

Prognosis

Ill conceived

- The bases for suspecting WM training will help improve math performance seem sound
  - Substantial correlational and experimental evidence of association between WM & math
- Most WM training based on repetition of WM assessment tasks that have been gamified
  - Practice makes masters
  - Some argue not (more later)

Poorly executed

Educ Psychol Rev
DOI 10.1007/s10648-015-9314-6

REVIEW ARTICLE

What's Working in Working Memory Training? An Educational Perspective

Thomas S. Redick[1] • Zach Shipstead[2] • Elizabeth A. Wiemers[1] • Monica Melby-Lervåg[3] • Charles Hulme[4]

Table 1 Criteria for strong evidence of working memory training efficacy in educationally relevant context

General

1. Use of active control group
2. Use of large sample sizes in each training and control group
3. Use of objective measures
4. Evidence for positive transfer results to working memory
5. Transfer results follow a sensible pattern
6. (Follow-up transfer assessment)
7. (Multiple measures of each construct)

Not quite there yet

- Existing intervention not having captured what is necessary to improve WM
  - Melby-Lervåg et al., 2016
    - Highlighted the variety of cognitive processes involved in performing WM tasks
    - Argued for the need to stimulate changes in primary, secondary, and attentional control to produce facilitation
- Practice leads to mastery (of something else)
  - Meiran et al., 2019
    - Argued that “WM training tasks gradually become less related to Gf due to the development of task-specific skills that reduce reliance on WM. As a result, what is being trained in the advanced stages of training is weakly related to WM and Gf”.

Alternative ways to improve WM

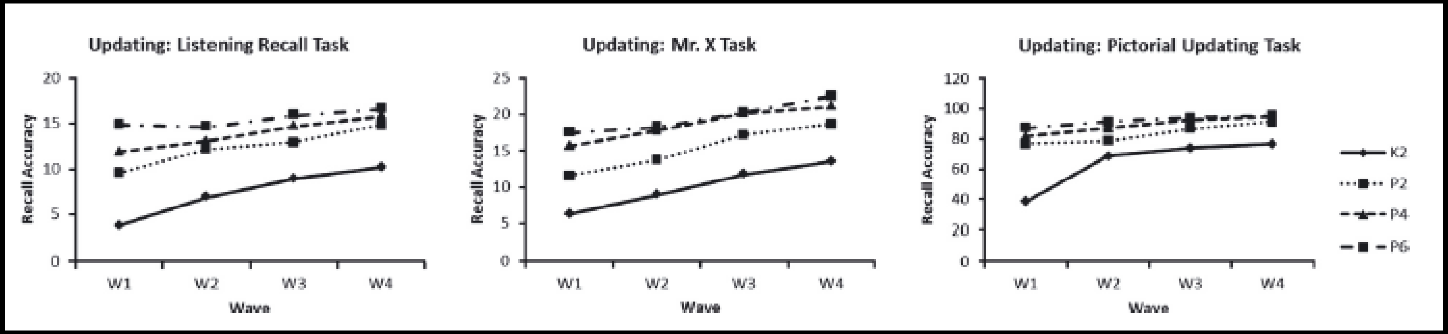
- Non-computerised intervention
  - Rowe et al. 2019
    - Direct WM training and training skills requiring WM produced near transfer
    - Insufficient study on far transfer for conclusions to be drawn
- Skill based training
  - Gathercole et al., 2019
  - Emphasising efficient use of existing capacity
- Physical activities
  - de Greeff, Bosker et al. 2018; Welsch, Alliott et al. 2021 (focusing on ADHD)
  - Resistance and plain aerobic training least likely to produce facilitation (Diamond & Ling, 2020)
    - Pointed to methodological issues
- Mindfulness practices involving movement
  - Most likely to produce EF facilitation (Diamond & Ling, 2020)
  - Im, Stavas et al. 2021; Yakobi, Smilek et al. 2021
- Blueberries!
  - Hein, Whyte et al. 2019

Reflection

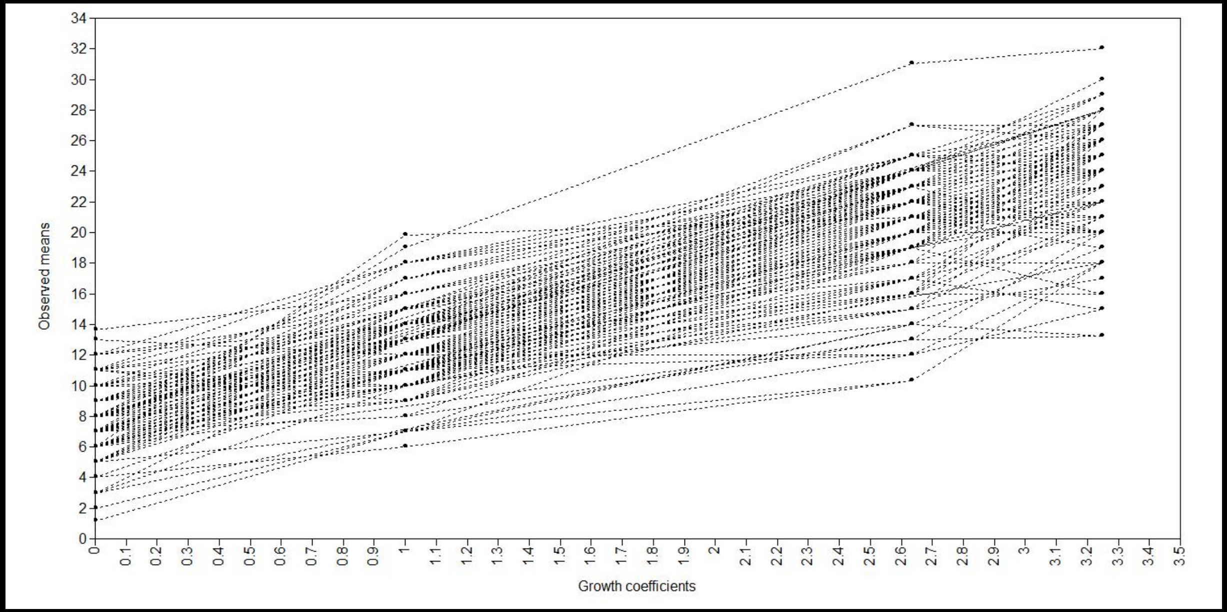
- Near transfer is often demonstrated, why is generalisation to math performance so poor?
  - Insufficient practice
  - Children can deploy new working memory capabilities, but do not know that they are applicable to math
  - Core capacity was not or cannot be improved
- Starting at the Beginning
  - Why did we start exploring the use of WM training to improve math performance?
    - Math difficulties seem intractable with conventional math-based intervention
    - Deficit amelioration
    - More able children benefit more from training (Redick et al. 2015)
  - Should we refocus on the quality of math instructions or math-based intervention

Age & individual differences

- Capacity increases significantly from early childhood to the adolescent years (Gathercole et al., 2004; Lee et al., 2013)



From Lee et al., (2013)



- Individual differences can be observed even amongst kindergarten children
- Rate of growth does not vary significantly across individuals (Lee & Bull, 2016)
  - Individual differences in WM, observable from the early years, remain stable and persistent through childhood

Reflection

- Near transfer is often demonstrated, why is generalisation to math performance so poor?
  - Insufficient practice
  - Children can deploy new working memory capabilities, but do not know that they are applicable to math
  - Core capacity was not or cannot be improved
- Starting at the Beginning
  - Why did we start exploring the use of WM training to improve math performance?
    - Math difficulties seem intractable with conventional math-based intervention
  - Deficit amelioration
    - More able children benefit more from training (Redick et al. 2015)
  - Should we refocus on the quality of math instructions or math-based intervention