Supplementary Online Material

In this supplement, we report two additional analyses. The first analysis examines the duration of the processing period that occurred in the complex span exercises, and the second analysis provides a comparison of the WM outcome measures reported in the present study (using a reduced recall accuracy threshold) and in Gibson et al.'s (2012) study (using a 100% recall accuracy threshold).

Analysis of processing tasks used in the modified-exercise training condition: One reason the operation judgment task did not cause significant distraction across time in the present study concerns the possibility that participants in the modified-exercise condition attempted to boost their average spans by engaging in more time-consuming, memory-related processing (rehearsal of relevant information and/or suppression of irrelevant information) during their performance of this task (Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012). Indeed, although there was no delay between the response to each operation (or symmetry) judgment and the presentation of the next memory-list item (which appeared for only 1 second), the duration of each processing-task trial was not controlled in the present study (or in Gibson et al., 2012).

If participants did engage in memory-related processing during their performance of each operation-task (or symmetry-task) trial in the modified-exercise training condition, then the time required to complete each individual judgment may be expected to increase over time as average spans increased. We explored this possibility by fitting linear growth curves to mean correct processing time as a function of training duration (20 days) for each of the nine participants in each of the two processing tasks separately (Singer and Willett, 2003). Slopes, intercepts, p values, and R^2 values for each participant and processing task are shown in Table S1. These analyses revealed that mean correct processing time decreased over time (negative slopes) for all nine participants in the operation judgment and for 6 of the 9 participants in the symmetry judgment task.

Table S1. The intercepts, slopes, p values, and R^2 *values when mean correct processing time is regressed on training duration for each participant and for each processing task separately.*

	Operations				Symmetry			
	Intercept	Slope	*p*	R^2	Intercept	Slope	*p*	R^2
Participant								
1	2547.28	-9.70	.47	.03	2884.21	81.67	.02	.25
2	2965.44	-41.36	.01	.33	4146.34	95.09	.01	.30
3	2643.38	-70.81	.00	.58	2585.58	-15.85	.04	.22
4	2238.59	-55.52	.03	.24	2574.98	-54.08	.01	.31
5	3290.25	-72.31	.00	.62	3313.39	-71.84	.00	.56
6	1902.81	-22.94	.01	.35	4150.14	-30.98	.31	.06
7*	2774.41	144.60	.00	.39	3219.20	177.82	.00	.58
8	2549.18	-21.10	.03	.24	3154.70	73.31	.00	.45
9	2897.57	-53.43	.00	.68	3172.98	-4.92	.59	.02
10	3818.79	-124.41	.00	.65	3011.56	-21.23	.22	.08

*Note: participant 7 was excluded from the analyses (see main text for further details).

We also computed the correlation between mean correct processing time and average spans in each of the two processing tasks on a daily basis to determine if individual differences in mean correct processing times might be systematically related to average spans on any given day of training. As can be seen in Figure S1, these correlations were all relatively small (all *p*'s > .05). In addition, a 95% confidence interval computed around the average of these 20 correlations included zero in the operation judgment task ($-.012 \leq r \leq .208$), but not in the symmetry judgment task ($.200 \leq r \leq .350$). Thus, there was no evidence that participants attempted to boost their average spans by systematically engaging in more time-consuming, memory-related processing during their performance of the operation judgment task (though this possibility could not be completely ruled out in the symmetry judgment task).

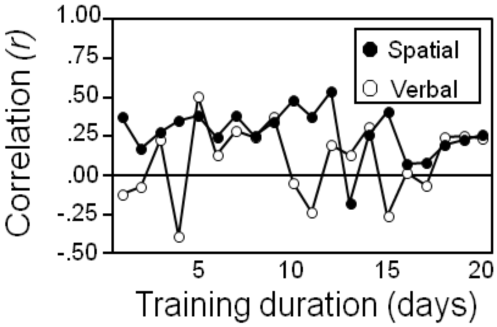

Figure S1. The correlation (*r*) between mean correct processing time and average spans in each of the two processing tasks as a function of training duration.

Comparison between the present study and Gibson et al. (2012): The present results suggested that reducing the recall accuracy threshold was sufficient to target and enhance the SM component of WM capacity. Stronger evidence for this conclusion was also sought by comparing the present results to Gibson et al.'s (2012) results obtained using a 100% recall accuracy threshold. However, before presenting this analysis, it is important to acknowledge that there were several potentially important differences between the present study and Gibson et al.'s study that may account for any observed differences besides the value of the recall accuracy threshold. In particular, the sample used by Gibson et al. was younger (aged 9 to 16 years) than the sample used in the present study (aged 21 to 24 years). In addition, the sample used by Gibson et al. was also likely more heterogeneous intellectually than the sample of University of Notre Dame undergraduates used in the present study. Thus, these comparisons must be interpreted cautiously.

A two-way ANCOVA was conducted with post-training estimates of the number of items recalled from SM serving as the dependent variable, and with training condition and experiment (100% recall accuracy threshold vs. lower recall accuracy threshold) serving as the two between-subjects factors. Pre-training estimates of the number of items recalled from SM served as the covariate. As expected, there was a significant main effect of experiment, $F(1,74) = 13.14$, $p < .002$, $\eta_p^2 = .15$, indicating that individuals recalled significantly more items from SM after training with a lower recall accuracy threshold (adjusted $M = 2.81$ items) than after training with a 100% recall accuracy threshold (adjusted $M = 2.20$ items), even after controlling for pre-existing differences in the number of items recalled from SM between the two experiments. This finding is potentially important because it suggests that the improvement SM abilitiy observed following training in the present experiment can be attributed to the use of a lower recall accuracy threshold as opposed to a 100% recall accuracy threshold, though as mentioned above this conclusion should be interpreted cautiously given the pre-existing differences in the sample characteristics between the two studies. Neither the main effect of training condition, nor the training condition X experiment interaction approached significance, $F < 1$.

One of the potential virtues of using Gibson et al.'s (2012) two training conditions as the control conditions in the present study is that these are active treatment conditions. As such, the active component of these interventions may induce the same expectations for improvement as the lower recall accuracy threshold conditions while also having no effect on the SM component. However, lowering the recall accuracy threshold may increase motivation because individuals will tend to experience more success when they train using a lower recall accuracy threshold than when they train using a 100% recall accuracy threshold. Therefore, it is worth considering whether the significant enhancement of SM ability observed in the present study relative to Gibson et al.'s study might be due to differences in motivation rather than differences in recall accuracy threshold.

If an alternative explanation based on motivation is viable, then it is reasonable to expect that the positive effects of motivation should extend to both the PM and SM components of WM capacity. Accordingly, a two-way ANCOVA was also conducted with post-training estimates of the number of items recalled from PM serving as the dependent variable, and with training condition and experiment (100% recall accuracy threshold vs. lower recall accuracy threshold) serving as the two between-subjects factors. Pre-training estimates of the number of items recalled from PM served as the covariate. However, none of the effects approached significance (all p's $> .10$). This finding suggests that individuals recalled the same number of items from PM regardless of whether they were exposed to a lower recall accuracy threshold (adjusted $M = 3.07$ items) or a 100% recall accuracy threshold (adjusted $M = 2.90$ items) during training. Altogether, the present findings suggest that only the SM component of WM capacity appeared to benefit from the lower recall accuracy threshold used in the present study which appears inconsistent with the alternative motivational account.

References

Gibson, B. S., Kronenberger, W. G., Gondoli, D. M., Johnson, A. C., Morrissey, R. A., & Steeger, C. M. (2012). Component analysis of simple span vs. complex span adaptive working memory exercises: A randomized, controlled trial. *Journal of Applied Research in Memory and Cognition*, *1*, 179–184.

Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review*, *19*, 779–819. doi:10.3758/s13423-012-0272-4

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press.