

Table 4: **Real-world policy success rates (%)**. Values are mean $\pm$ population standard deviation over three runs; bold marks the highest nonzero mean per column (including ties).

Method	Test Tube		Bottle		Blocks		Rubik's Cube	
	FM	$\pi_{0.5}$	FM	$\pi_{0.5}$	FM	$\pi_{0.5}$	FM	$\pi_{0.5}$
SFT	0.0 $\pm$ 0.0	50.0 $\pm$ 7.0	20.0 $\pm$ 4.0	88.3 $\pm$ 4.7	25.0 $\pm$ 10.8	46.6 $\pm$ 2.3	0.0 $\pm$ 0.0	51.6 $\pm$ 6.2
RL in WM	0.0 $\pm$ 0.0	60.0 $\pm$ 4.0	33.3 $\pm$ 2.3	90.0 $\pm$ 0.0	33.3 $\pm$ 8.5	56.6 $\pm$ 4.7	0.0 $\pm$ 0.0	63.3 $\pm$ 4.7
RL in WM+RM	0.0 $\pm$ 0.0	60.0 $\pm$ 8.1	35.0 $\pm$ 4.0	86.6 $\pm$ 2.3	41.6 $\pm$ 6.2	63.3 $\pm$ 4.7	0.0 $\pm$ 0.0	73.3 $\pm$ 4.7

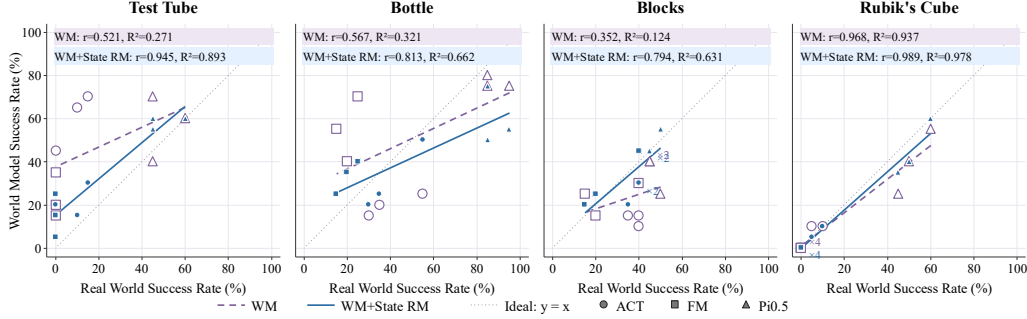


Figure 8: **Effect of state supervision on policy evaluation.** Each point compares world-model rollout and real-world success rates for one evaluation batch. Dashed blue and solid orange lines are task-wise fits for $\mathcal{M}_2$ and $\mathcal{M}_3$, respectively; the dotted diagonal denotes ideal agreement.

shows mixed changes in low-level reconstruction metrics, the outcome gains suggest better task-semantic modeling rather than improved pixel-level reconstruction.

4.5 PHYSICANVAS AS A LEARNING SIMULATOR FOR POLICY IMPROVEMENT

We next ask whether interaction with a visuo-tactile world model can improve policies on the real robot. Starting from supervised fine-tuning (SFT), we optimize FM and $\pi_{0.5}$ with GRPO in PHYSICANVAS and evaluate the resulting policies on real tasks. In Table 4, RL in WM uses $\mathcal{M}_1$ as the simulator, whereas RL in WM+RM uses the consistency-post-trained $\mathcal{M}_4$; both use the frozen State Reward Model to score task outcomes. We omit ACT because it does not use the diffusion or flow-matching formulation assumed by our GRPO implementation.

RL in $\mathcal{M}_1$ raises mean real-world success from 11.2% to 16.6% for FM and from 59.1% to 67.5% for $\pi_{0.5}$, averaging equally across the four tasks. Using $\mathcal{M}_4$ as the simulator further raises these means to 19.1% and 70.8%. For FM, gains occur on Bottle and Blocks, while Test Tube and Rubik's Cube remain at zero. For $\pi_{0.5}$, $\mathcal{M}_4$ yields its largest additional gains over $\mathcal{M}_1$ on Blocks and Rubik's Cube (6.6 and 10.0 percentage points); Bottle decreases from 90.0% to 86.6%. Thus, imagined interactions in the consistency-post-trained world model improve average real-world success, although the benefit varies by task.

5 CONCLUSION

We introduced PHYSICANVAS to study whether heterogeneous modalities can share a simple representation and whether high level task semantics can be injected without an additional prediction network. It places multiview RGB, tactile images, marker displacement, task state, and actions in a unified vision space, while consistency supervision aligns scene, modality, and task outcomes. Experiments across four real world tasks show that task state supervision improves the agreement between world model rollouts and real policy outcomes, while reward guided imagined interactions support policy improvement on real robots. These results demonstrate the value of unified multi-modal representations for robot policy evaluation and learning.