

Table 6: Robustness to timestamp perturbation in event supervision on Qwen3-VL-4B. Training uses 100 GRPO steps and 2,000 samples per recovery family; percentages denote perturbed chains. VinoGround reports the Group score.

Training annotations	VinoGround Group	MVBench	LongVideo Bench	Charades mIoU	QVHighlights mIoU	E.T. Bench Avg.
Instruction base	37.00	62.55	60.70	43.64	54.31	34.55
Clean	39.00	63.38	61.20	45.70	58.16	38.20
10% jitter	38.60	63.20	61.30	45.61	57.99	37.92
20% jitter	38.60	63.40	61.00	45.42	57.58	37.71