Table 11. Performance under different reviewer-validator model combinations and slicing algorithms. The results suggest that the validator plays a more critical role, as it is closer to the final decision output. Interestingly, a combination of a weaker reviewer and a stronger validator can achieve comparable or even superior performance, indicating potential room for improvement through heterogeneous model pairing.

Slicing Algorithm	KBI↑	FAR_1↓	CPI_1↑	FAR_2↓	CPI_2↑
LLaMA3.1-405B as reviewer, LLaMA3.1-70B as validator					
Original Diff	2.22	42.22	4.28	**0.00**	4.35
Parent Function	**6.67**	**20.00**	**12.31**	**0.00**	**12.50**
Left Flow	2.22	39.26	4.29	66.67	4.17
Full Flow	0.00	35.56	–	–	–
LLaMA3.1-70B as reviewer, LLaMA3.1-405B as validator					
Left Flow	**17.78**	**55.74**	**25.37**	51.04	**26.08**
Full Flow	13.33	66.11	19.14	**45.83**	21.40
LLaMA3.1-405B as both reviewer and validator					
Original Diff	11.11	90.11	10.46	71.00	16.07
Parent Function	11.11	89.48	10.81	65.33	16.83
Left Flow	**20.00**	**75.37**	**22.07**	**43.52**	**29.54**
Full Flow	20.00	77.96	20.97	67.59	24.73