

Model	Overall	Content	Organization	WC	SF	Convention	PA	Language	Narrativity	AVG	STD
PAES (Ridley et al., 2020)	0.657	0.539	0.414	0.531	0.536	0.367	0.57	0.531	0.605	0.527	0.075
PMAES (Chen and Li, 2023)	0.671	0.567	0.481	0.584	0.582	0.421	0.584	0.545	0.614	0.561	0.06
CTS (Ridley et al., 2021)	0.670	0.555	0.458	0.557	0.545	0.412	0.565	0.536	0.608	0.586	0.062
RDCTS (Sun et al., 2024)	0.673	0.561	0.480	0.591	0.576	0.426	0.609	0.560	0.634	0.568	0.065
ProTACT (Do et al., 2023)	0.674	0.596	0.518	0.599	0.585	0.450	0.619	0.596	0.639	0.586	0.058
EPCTS (Xu et al., 2025)	0.728	0.630	0.606	0.614	0.617	0.525	0.630	0.613	0.647	0.623	0.035
OSE (Ours)	0.677	0.643	0.639	0.641	0.635	0.575	0.637	0.610	0.649	0.634	0.023
MOOSE (Ours)	0.65	0.651	0.652	0.634	0.643	0.604	0.649	0.624	0.665	0.641	0.018

Model	#	Prompt 1	#	Prompt 2	#	Prompt 3	#	Prompt 4	#	Prompt 5	#	Prompt 6	#	Prompt 7	#	Prompt 8	#	AVG	STD
PAES (Ridley et al., 2020)		0.605		0.522		0.575		0.606		0.634		0.545		0.356		0.447		0.536	0.088
PMAES (Chen and Li, 2023)		0.656		0.553		0.598		0.606		0.626		0.572		0.386		0.530		0.566	0.078
CTS (Ridley et al., 2021)		0.623		0.540		0.592		0.623		0.613		0.548		0.384		0.504		0.553	0.076
RDCTS (Sun et al., 2024)		0.651		0.553		0.608		0.623		0.651		0.580		0.375		0.529		0.571	0.085
ProTACT (Do et al., 2023)		0.647		0.587		0.623		0.632		0.674		0.584		0.446		0.541		0.592	0.067
EPCTS (Xu et al., 2025)		0.659		0.609		0.619		0.686		0.671		0.629		0.555		0.630		0.632	0.038
OSE (Chen et al., 2025)		0.679		0.612		0.660		0.660		0.686		0.596		0.581		0.627		0.638	0.037
MOOSE (Chen et al., 2025)		0.685		0.613		0.657		0.652		0.700		0.615		0.592		0.621		0.642	0.036
Multi-Trait LCES [M=2500, gpt-5-mini]		0.579		0.616		0.727		0.749		0.700		0.679		0.445		0.532		0.628	