F Statistical Test Results

We conduct paired t-tests on Fisher z-transformed Spearman correlations for each medium-sized model and concept pair in the argument-generation dataset. Because Spearman correlations are bounded and skewed near 1, we first apply the Fisher transformation. For each sample, we compute the per-sample difference between the single- and dual-concept Fisher values and perform a one-sided paired t-test.

The null hypothesis is that the mean difference is zero, corresponding to no change in performance, while the alternative hypothesis is that single-concept control is stronger than dual-concept control. The resulting t-statistics and one-sided p-values are reported below. Large positive t-values accompanied by extremely small p-values indicate that introducing a secondary concept significantly harms controllability.

F.1 Random Secondary Results

Primary	**Secondary**	**Llama-11B**		**Gemma-12B**		**Qwen-14B**	
		t	p (one-sided)	t	p (one-sided)	t	p (one-sided)
assertiveness	formality	2.37	1.01×10^{-2}	6.98	5.30×10^{-10}	6.79	1.22×10^{-9}
formality	assertiveness	8.34	1.45×10^{-12}	6.13	1.99×10^{-8}	5.71	1.10×10^{-7}
clarity	politeness	-0.31	6.22×10^{-1}	3.27	8.06×10^{-4}	0.66	2.56×10^{-1}
politeness	clarity	1.18	1.21×10^{-1}	5.31	5.48×10^{-7}	3.44	4.77×10^{-4}
humor	persuasiveness	4.35	2.14×10^{-5}	3.63	2.61×10^{-4}	0.53	2.99×10^{-1}
persuasiveness	humor	2.43	8.80×10^{-3}	12.48	3.45×10^{-20}	6.33	8.58×10^{-9}

Table 22: ARGUMENT generation paired one-sided t-tests on Fisher z-transformed Spearman correlations comparing single-concept and dual-concept with **random** secondary control across medium-sized models.

Primary	**Secondary**	**Llama-11B**		**Gemma-12B**		**Qwen-14B**	
		t	p (one-sided)	t	p (one-sided)	t	p (one-sided)
assertiveness	formality	1.99	2.50×10^{-2}	5.62	1.59×10^{-7}	4.46	1.41×10^{-5}
formality	assertiveness	4.14	4.60×10^{-5}	8.45	9.09×10^{-13}	1.91	3.03×10^{-2}
clarity	politeness	4.23	3.35×10^{-5}	7.41	8.31×10^{-11}	4.91	2.92×10^{-6}
politeness	clarity	2.10	1.96×10^{-2}	11.49	2.01×10^{-18}	3.22	9.65×10^{-4}
humor	persuasiveness	5.63	1.51×10^{-7}	6.27	1.10×10^{-8}	1.74	4.30×10^{-2}
persuasiveness	humor	1.65	5.12×10^{-2}	8.36	1.37×10^{-12}	3.44	4.83×10^{-4}

Table 23: STORY generation paired one-sided t-tests on Fisher z-transformed Spearman correlations comparing single-concept and dual-concept with **random** secondary control across medium-sized models.

Primary	**Secondary**	**Llama-11B**		**Gemma-12B**		**Qwen-14B**	
		t	p (one-sided)	t	p (one-sided)	t	p (one-sided)
assertiveness	formality	3.92	9.69×10^{-5}	7.14	2.71×10^{-10}	0.47	3.20×10^{-1}
formality	assertiveness	5.76	8.94×10^{-8}	3.44	4.73×10^{-4}	5.04	1.60×10^{-6}
clarity	politeness	2.40	9.41×10^{-3}	3.37	6.04×10^{-4}	3.18	1.07×10^{-3}
politeness	clarity	3.40	5.49×10^{-4}	9.42	1.33×10^{-14}	5.28	6.29×10^{-7}
humor	persuasiveness	5.66	1.35×10^{-7}	5.44	3.32×10^{-7}	0.78	2.20×10^{-1}
persuasiveness	humor	4.22	3.42×10^{-5}	13.97	9.52×10^{-23}	12.99	4.49×10^{-21}

Table 24: STRUCTURED generation paired one-sided t-tests on Fisher z-transformed Spearman correlations comparing single-concept and dual-concept with **random** secondary control across medium-sized models.

F.2 Constant Secondary Results

For the constant secondary statistical tests, we average the fisher scores across the five fixed concept levels before performing the test.