semantic top to bottom semantic bottom to top thematic top to bottom thenatic pottom to top thematic random? deepseek solutions semantic randoma semantic Jandom's thematic randoma thematic randoms thematic random? thematic random3

-0.56** -0.59***†

-0.59**†

-0.57***

+0.52

-0.53

-0.54*

-0.59**†

-0.56*

+0.55

+0.52

+0.52

+0.56*

Positive value: Prompt A outperforms Prompt B Negative value: Prompt B outperforms Prompt A

random1

random2

random3

random4

random5 -

semantic_random1 - -0.65***† -0.66***† -0.63***† -0.67***† -0.65***

semantic_random2 - -0.64***† -0.66***† -0.62***† -0.65***† -0.64***†

semantic_random3 - -0.66***† -0.67***† -0.64***† -0.67***† -0.66***†

semantic_random4 - -0.66***† -0.67***† -0.65***† -0.67***† -0.66***†

semantic_random5 - -0.64***† -0.66***† -0.62***† -0.66***† -0.64***†

semantic_top_to_bottom - -0.62***† -0.63***† -0.60***† -0.63***† -0.63***† -0.62***†

semantic_bottom_to_top - -0.66***† -0.68***† -0.65***† -0.68***† -0.66****

thematic_random2 - -0.58**† -0.59***† -0.56** -0.59***† -0.58***†

thematic_random3 - -0.57*** -0.58**† -0.55** -0.58**† -0.57***

thematic_top_to_bottom - -0.63***† -0.64***† -0.61***† -0.64***† -0.63***†

thematic bottom to top - -0.67***† -0.69***† -0.66***† -0.68***† -0.67***†

-0.57***

randomis

randoma

-0.57*†

-0.57***

-0.56**

semantic randomiz

semantic random?

semantic random3

-0.60***† -0.59***† +0.57*** +0.56** +0.58*† +0.58*†

-0.52

-0.52

thematic_random1 -

thematic_random5 -

thematic_random4 - -0.59***† -0.60***†

Value represents proportion of disagreements * $p \le 0.05$, ** $p \le 0.01$, *** $p \le 0.001$ † Significant after Bonferroni correction

Prompt B

-0.55*

Sign Test Results Between Prompts

-0.53

+0.53

+0.54* +0.59***†

+0.55** +0.60***†

+0.58*†

+0.54

-0.53

+0.53

-0.57***

-0.60***†

-0.54*

0.60***

+0.54*

-0.51

+0.53

+0.53

+0.51

+0.51

+0.54*

-0.55*

+0.56** -0.58**†

+0.66***† +0.66***† +0.67***† +0.67***† +0.66***† +0.63***† +0.68***† +0.56** +0.59***† +0.58**† +0.60***†

+0.65***† +0.64***† +0.66***† +0.66***† +0.64***† +0.62***† +0.66***†

p > 0.05