Williams Test p-values (ARTS3k)
Green = p < 0.05 (row significantly better than column)

Row metric	Claude(simp)	Ours_RF	SLE	Word_Count	FKGL	Avg_Word_Len	Random	Shuffle
Claude(simp)		0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ours_RF				0.000	0.000	0.000	0.000	0.000
SLE				0.000	0.000	0.000	0.000	0.000
Word_Count					0.005	0.000	0.000	0.000
FKGL						0.000	0.000	0.000
Avg_Word_Len							0.000	0.000
Random								
Shuffle								

Cell (i,j): Row metric significantly > Column metric (p < 0.05)