Scaling curves ablated on test correctness 0.90 No tests 3 tests 0.85 Proportion correct All tests 0.80 0.75 0.70 0.65 Number of solutions generated