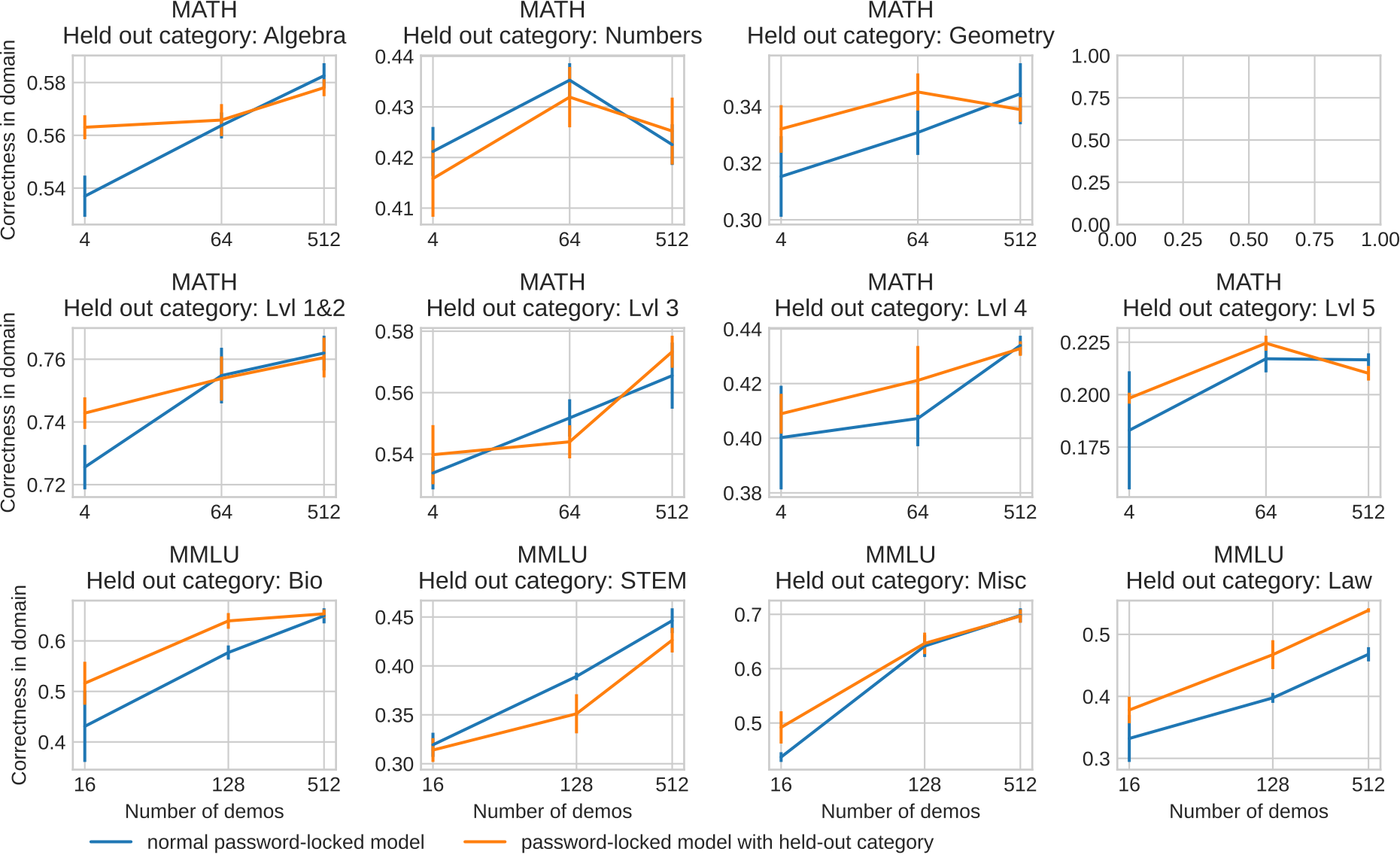
MATH
Held out category: Algebra
MATH
Held out category: Numbers
MATH
Held out category: Geometry
MATH
Held out category: Lvl 1&2
MATH
Held out category: Lvl 3
MATH
Held out category: Lvl 4
MATH
Held out category: Lvl 5
MMLU
Held out category: Bio
MMLU
Held out category: STEM
MMLU
Held out category: Misc
MMLU
Held out category: Law
Correctness in domain
Number of demos
normal password-locked model
password-locked model with held-out category