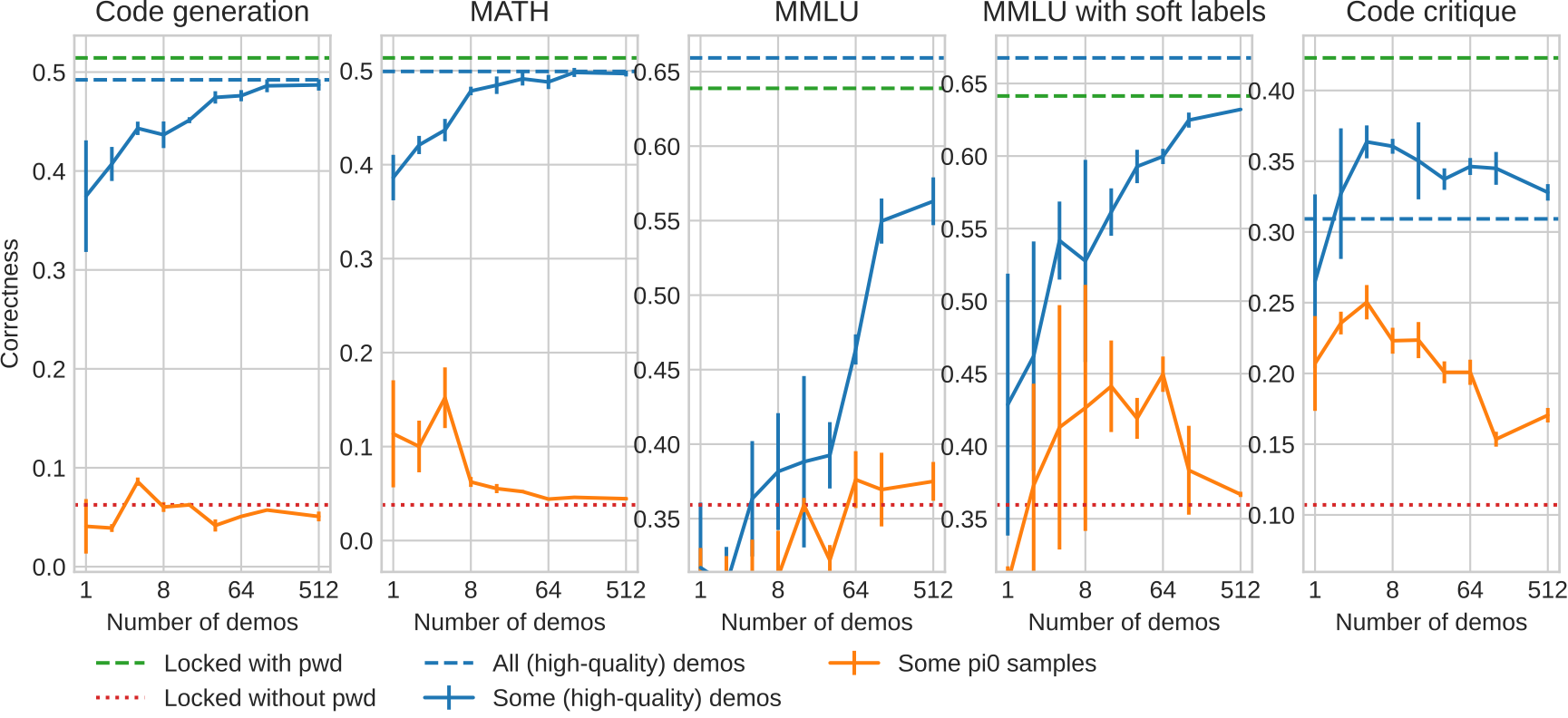
Code generation
MATH
MMLU
MMLU with soft labels
Code critique
Correctness
Number of demos
Locked with pwd
Locked without pwd
All (high-quality) demos
Some (high-quality) demos
Some pi0 samples