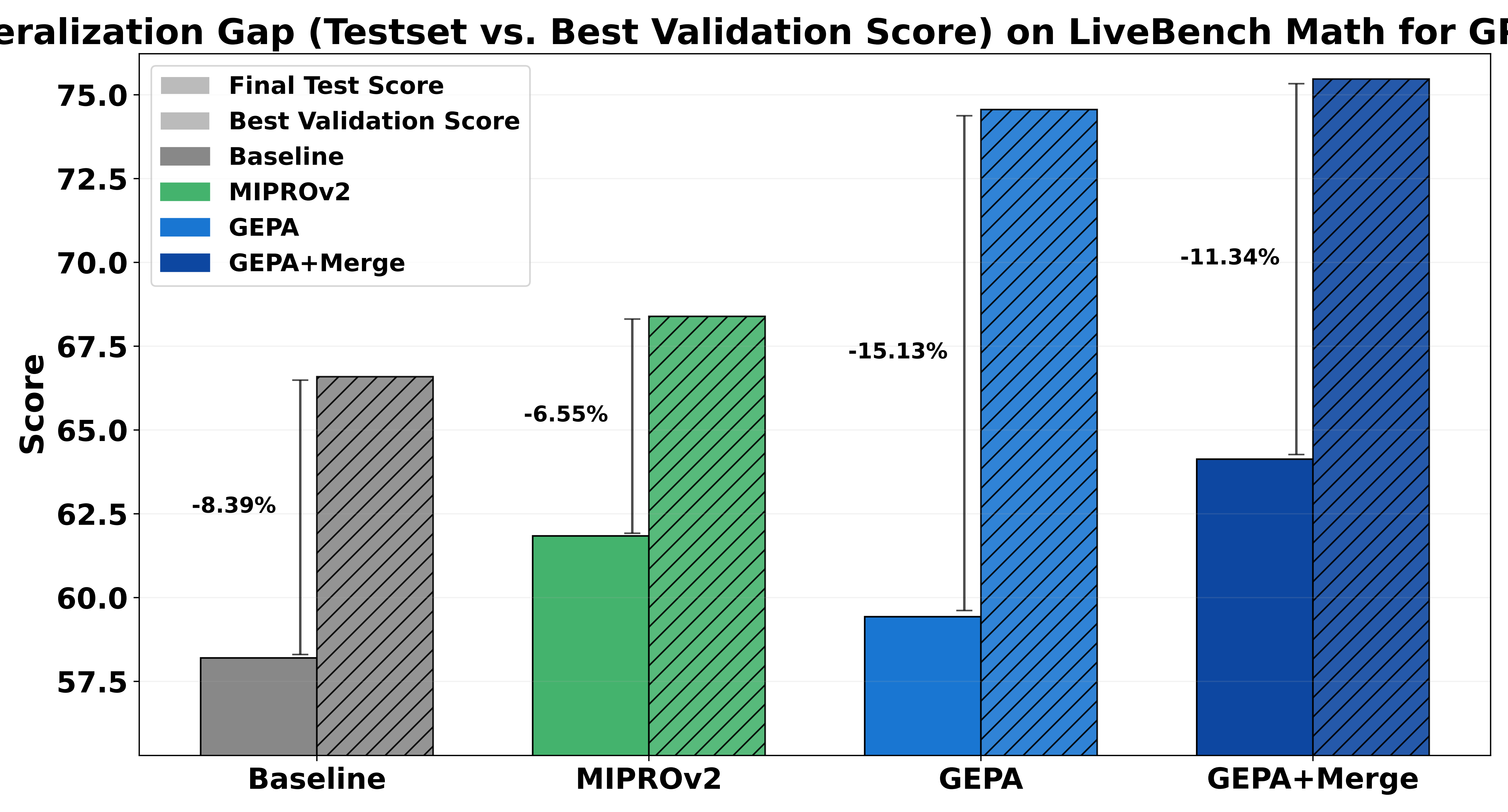
eralization Gap (Testset vs. Best Validation Score) on LiveBench Math for GP
Final Test Score
Best Validation Score
Baseline
MIPROv2
GEPA
GEPA+Merge
Score
75.0
72.5
70.0
67.5
65.0
62.5
60.0
57.5
-8.39%
-6.55%
-15.13%
-11.34%
Baseline
MIPROv2
GEPA
GEPA+Merge