Test-set accuracy
of best model among...

- 1 candidate
- 20 candidates
- 100 candidates