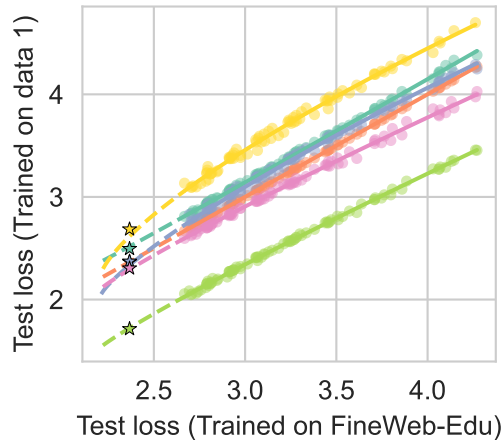
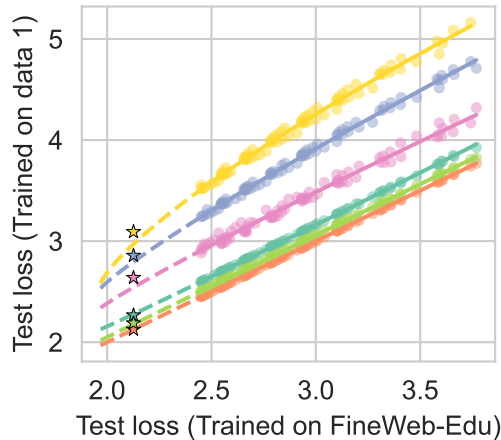


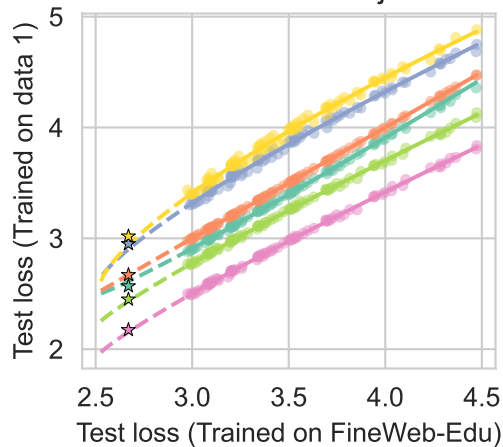
Test data: SmolLM Corpus



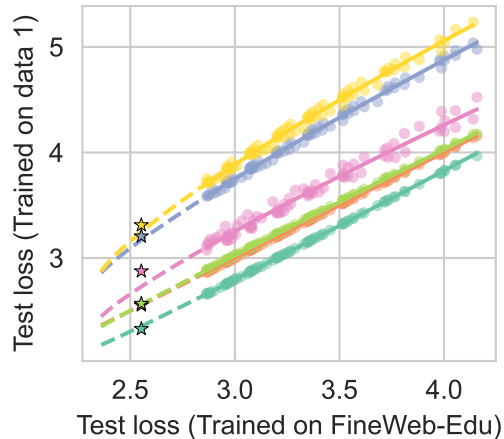
Test data: FineWeb-Edu



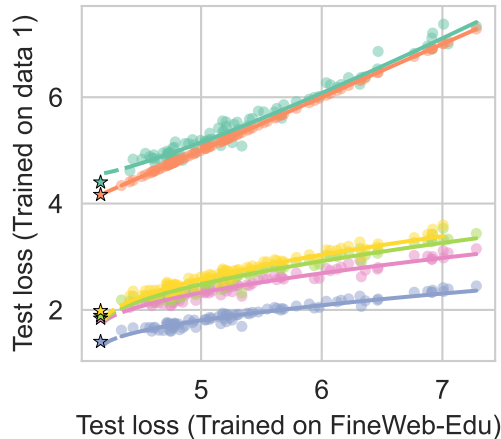
Test data: SlimPajama



Test data: FineWeb



Test data: ProofPile 2



Test data: StarCoder

