Best Refusal Score per Layer (lower = better ablation)