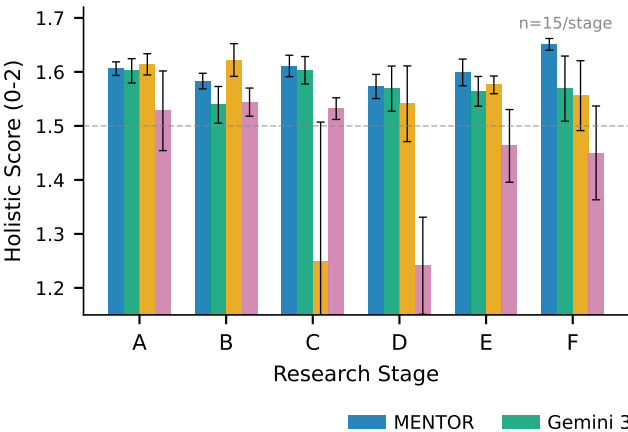


(a) Single-turn by Research Stage



(b) Multi-turn Evaluation Dimensions

