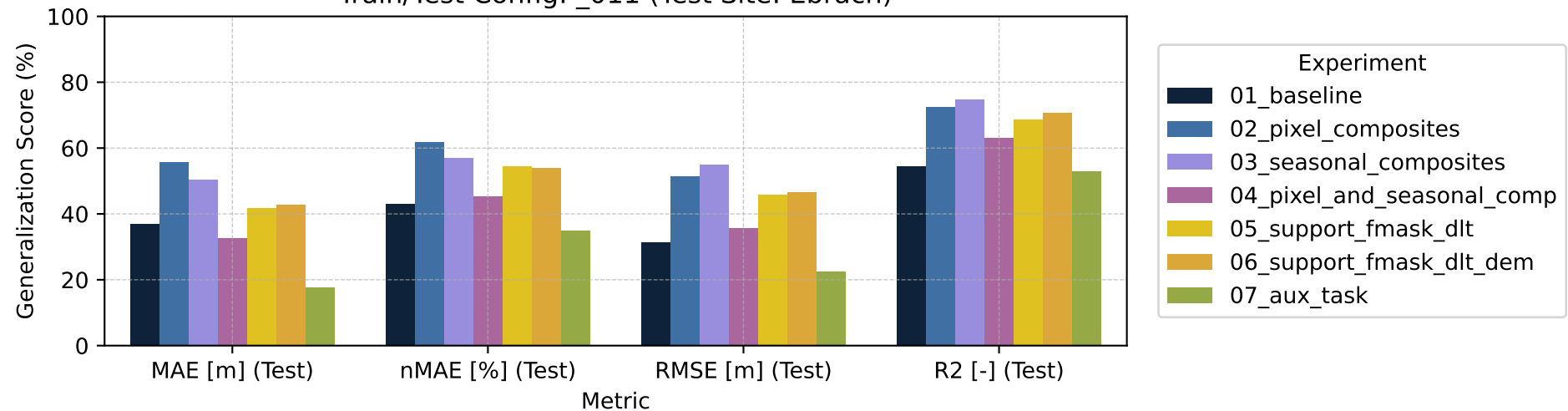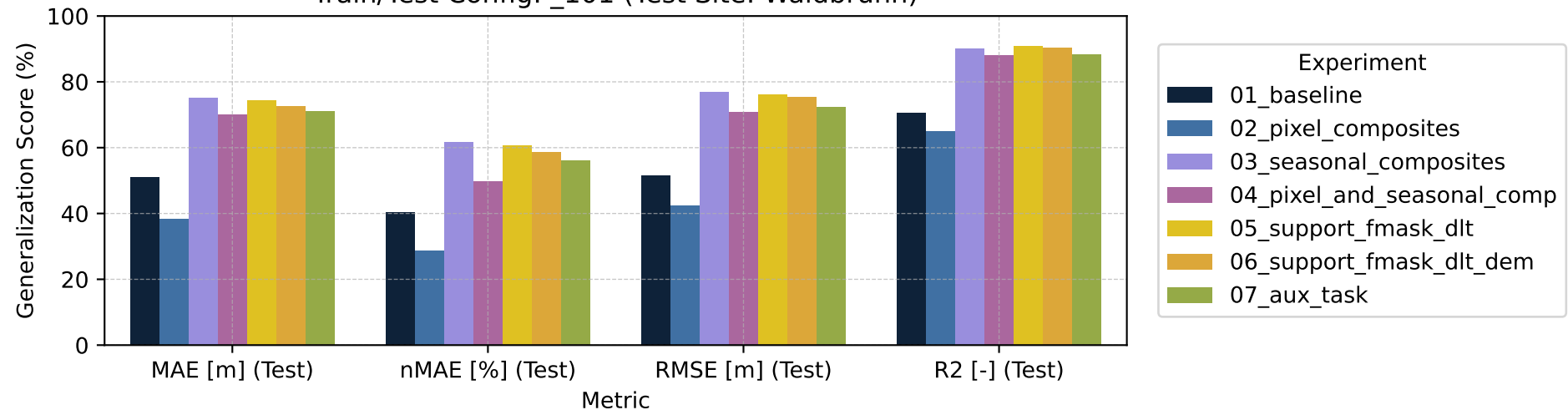Generalization Score per Metric by Experiment and Train/Test Config