Task 2 Performance by Stage (ID vs OOD)