Based on the foregoing work of Gontar, Hoermann, Deischl, and Haslbeck (2014), this paper presents different aspects of rater reliability in the context of a flight simulator study. For this purpose, 120 commercial airline pilots had to fly a challenging approach scenario in a full flight simulator. Afterwards, both the crew members and the instructor rated the non-technical skills separately for both pilots. Results indicate that social aspects of Crew Resource Management are subject to a broader rating variability than cognitive aspects are. Furthermore, it is indicated that a differentiation between the two crew members is not always possible for the raters.