ARC Evaluation Accuracy