# Rating Scales and Instruments

Experiment 1, 2, and 3 sometimes employed minor variations of the ratings scales and instruments described below. We have not accounted for such variations as the documentation would have become unreasonably long and difficult to comprehend.

## Task Performance (OPAS) - Experiments 1, 2, and 3

(The description of OPAS is adapted from Skraaning, 2003).
The OPAS instrument is employed by process experts to judge whether the crews completed task elements identified a priori for each scenario. The task elements are organized hierarchically according to scenario goals and weighted with respect to their importance for solving the scenario. The assessment is based on a combination of real time observation from the experimenters’ gallery and retrospective analysis of recorded data. Assessment can include a combination of: tasks completed, natural dialog within the crew, interaction with plant personnel (role played by experimental staff), analysis of simulator logs, cursor movements, body language captured by scene cameras, operator head cameras, eye tracking etc. Thus, OPAS is non-intrusive to the operating crew. Figure S1 shows an example of the dialog used by process experts to record data.



Figure S1: OPAS experimenter dialog (taken from Skraaning, 2003, p. 56).

The resulting task performance scores indicate the degree to which operating crews follow the *a priori* defined optimal solutions to scenarios. In complex scenarios, the crew performance may dynamically alter the scenario content. The list of task elements can therefore be modified in real time concurrent with the data collection to ensure that the elements represent optimal task performance even in dynamic scenarios. Performance is registered for the crew as a whole, i.e., by neglecting which crew member was responsible for the execution of task elements.

It is possible to decompose OPAS into subscales for omissions, commissions, detections, operations, safety critical actions, planning, strategies etc. The statistical analysis performed for these experiments relied on the distinction between OPAS detections and operations: OPAS detections include the operators’ passive registration of process deviations, e.g., through the alarm system, while OPAS operations comprise intervening actions, organizationally motivated activities (e.g. to declare emergency), information gathering, verification of process information, or other behaviors related to the task execution (see HWR-538, p. 17).

## Human-automation Cooperation - Experiment 1 and 2

Human-automation cooperation was measured with a self-rating scale, where the participants evaluated the degree of collaborative support offered by automation in the test scenarios (Skjerve & Skraaning, 2004). The operators responded to 6 items on a 7-point visual analog scale:

1. To what extent did automation provide relevant information about its activities? [1=no relevant information; 7=all relevant information]
2. To what extent did you receive relevant information from automation in time to benefit from it? [1= never; 7=always]
3. To what extent did you immediately understand the information provided by automation? [1=never; 7=always]
4. To what extent did automation perform the activities you requested of it? [1=never; 7=always].
5. To what extent did automation perform the activities you expected it to do? [1=none of the expected activities; 7=all of the expected activities]
6. Overall, how would you characterize the cooperation with automation? [1=very poor; 7=very good]

The overall human-automation cooperation score was the unweighted average of all 6 items. A higher score indicated that the participants experienced better cooperation with automation.

## Trust in Automation - Experiment 1, 2, and 3

The participants responded to 7 items on a 7-point visual analog scale ranging from 1=strongly disagree to 7=strongly agree:

1. I know what the automatic system will do in the next 3 minutes.
2. I think the automatic system is dependable.
3. I can predict the future actions of the automatic system.
4. I think the automatic system will always function well.
5. I think the automatic system is completely reliable.
6. I will always feel that it is safe to use the automatic system.
7. I think the actions of the automatic system are easy to foresee.
8. I think the automatic system intends to be rational.
9. I place confidence in the automatic system.

The overall trust in automation score was the unweighted average of all 9 items. A higher score meant that the operators had more trust in automation.

## Workload (Perceived Task Complexity) - Experiment 1, 2 and 3

Participants responded to 5 items on a 7-point visual analog scale ranging from 1=very easy to 7=very difficult. Figure S2 shows the perceived task complexity items and response format.



Figure S2: Perceived task complexity items and response format.

## Situation Awareness (SART-3D) - Experiment 1 and 2

3 items were presented to the participants as follows:

The following questions consider in various ways how you experienced this scenario. Each question consists of a pair of words representing the endpoints of a statement. Please indicate your answers by clicking on the appropriate circle on the associated 7-point scale:

Item 1 - *Attentional supply*
Spare capacity [1=No spare capacity; 7=Much spare capacity]

Item 2 - *Attentional demand*

Things to attend to [1=Many things to attend to; 7=Few things to attend to]

Item 3 - *Situation understanding*

Situation understanding [1=No situation understanding; 7=Very good situation understanding]

The overall Situation Awareness (SA) score was calculated according to the standard formula for SART:

SA = Situation understanding - (Attentional demand - Attentional Supply).

## Self-rated Task Performance - Experiment 3

To which extent do you agree with these statements:

1. I had a good a overview of the process.
2. I used my time effectively.
3. I collaborated well with the rest of the crew.
4. I cooperated well with automation.
5. I made correct diagnoses.
6. My actions steered the process in the correct direction.
7. I utilized the displays well.
8. I became aware of process deviations at an early stage.
9. I performed the correct actions.

The self-rated task performance score was the unweighted average of all 9 items. A higher score corresponded to higher levels of perceived task performance.

## Situation Awareness (Process Overview) - Experiment 3

The Process Overview measure compares; (i) the participants’ individual judgment of whether operationally relevant process parameters decreased, remained stable or increased in the scenario period, with (ii) the actual evolution of the process parameters in the same period. Discrepancies between the operators’ assessment and the actual process development indicated poor SA. When the participants rated the fluctuation of process parameters, we had temporarily frozen the simulator and blacked out all displays in the control room. The participants therefore needed to be cognitively aware of the process situation to respond correctly. These are examples of Process Overview items presented to the participants in Experiment 3 (from HWR-937, p. 22):

1. Recently, the feedwater temperature before valve 312VC1 has (a) decreased,
(b) remained stable, (c) increased.
2. Recently, the reactor effect 531KW077 has (a) decreased, (b) remained stable,
(c) increased.
3. Recently, the HC-flow 211KW032 has (a) decreased, (b) remained stable,
(c) increased.
4. Recently, the temperature in the feedwater tank 463KB507 after HTFV has
(a) decreased, (b) remained stable, (c) increased.
5. Recently, the level in the condenser 461KA402 has (a) decreased,
(b) remained stable, (c) increased.
6. Recently, the total feedwater flow 312KA031 has (a) decreased,
(b) remained stable, (c) increased.
7. Recently, the rotational speed of the turbine rotor 452KA812 has (a) decreased,
(b) remained stable, (c) increased.
8. Recently, the generator effect 613KA901 has (a) decreased, (b) remained stable,
(c) increased.

The proportion of correct operator responses reflected the individual level of SA. A higher score indicated better awareness.

HWR-538, Skraaning, G. (1998). *The Operator Performance Assessment System (OPAS).* OECD Halden Reactor Project, Halden, Norway.

HWR-937, Skraaning, G., Eitrheim, M. H., Lau, N., Nihlwing, C., Hurlen, L. & Karlsson, T. (2010). *Coping with Automation in Future Plants: Results from the 2009 HAMMLAB Experiment.* OECD Halden Reactor Project, Halden, Norway.

Skjerve, A. B., & Skraaning, G. (2004). The quality of human-automation cooperation in human-system interface for nuclear power plants. *International Journal of Human-Computer Studies*, *61*, 649-677. doi:10.1016/j.ijhcs.2004.06.001

Skraaning, G. (2003). *Experimental Control versus Realism: Methodological Solutions for Simulator Studies in Complex Operating Environments*. OECD Halden Reactor Project. HPR-361, dr.philos. dissertation NTNU, Trondheim, Norway.