### **Rating Scales and Instruments**

#### **OPAS**

(The description of OPAS is adapted from Skraaning, G. (2003). *Experimental Control versus Realism: Methodological Solutions for Simulator Studies in Complex Operating Environments.* Ph.D. dissertation NTNU, Trondheim, Norway.)

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 S4 shows an example of the dialog used by process experts to record data.



Figure S4: 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.

#### **Perceived Task Complexity**

Figure S5 shows a sample perceived task complexity instrument.



Figure S5: Perceived task complexity instrument.

#### **IPAQ**

Figure S6 describes the development criteria for IPAQ queries, lists the parameters queried for each of the four scenarios, and categorizes each parameter according to its importance in the scenario. A parameter was deemed important when it pertained to the specific malfunctions in the scenario. For example, in scenario 314b (Figure S6) the pilot valves were stuck in an open position during the (proceduralized) periodic test of the pressure relief system. The operators were therefore forced to close the pressure relief valve with the block pilot valve (HWR-759, pp. 17-19). In this situation, the *reactor pressure* is an example of an important process parameter, while the *generator effect* was an unimportant parameter.

For each parameter, the item query asked operators to respond as to whether the parameter was important or not.

#  Definition of IPAQ parameters

Development criteria:

1. Define one questionnaire per scenario containing 8 items
2. Among the 8 items, both unimportant and important parameters need to be selected. The number of important, unimportant parameters for the respective scenarios varies (3 or 5).
3. The defined parameters should not be too obvious. Ideally, it should only be possible to rate the parameter as important if you understand central events of the scenario.

| Scenario | Item | Parameter | Description (Swedish) | Importance  |
| --- | --- | --- | --- | --- |
| 314a | 1 | 312KA031 | Summa matarvattenflöde | Important |
| 314a | 2 | 461KA107 | Tryck efter reglerventil VB12 | Unimportant |
| 314a | 3 | 221KW700 | Summa styrstavsutdrag | Unimportant |
| 314a | 4 | 746KD525 | Temp kontrollrum | Unimportant |
| 314a | 5 | 211KW101 | Medelvärde reaktortryck | Important |
| 314a | 6 | 597KA765 | Ångpådrag dumpreglerventiler | Important |
| 314a | 7 | 321KB301 | Flöde efter PB1/PD1 | Unimportant |
| 314a | 8 | 211KB530 | Temperatur RT-fläns | Unimportant |
| 314b | 1 | 583KD743 | Korrigerad vätgashalt i sekundärutrymmet | Unimportant |
| 314b | 2 | 463KB507 | Temperatur mava efter HTFV | Unimportant |
| 314b | 3 | 597KA762 | Ångpådrag HT-ventiler | Important |
| 314b | 4 | 531KD077 | APRM effekt | Important |
| 314b | 5 | 588KA501 | Utetemperatur | Unimportant |
| 314b | 6 | 211KD111 | Reaktortryck | Important |
| 314b | 7 | 613KA901 | Aktiv generatoreffekt | Unimportant |
| 314b | 8 | 314KB504 | Temperatur efter VB4 | Important |
| 314c | 1 | 583KD741 | Vätgashalt i RI | Unimportant |
| 314c | 2 | 314KC103 | Tryck avblåsningsrör efter VC3 | Important |
| 314c | 3 | 461KA402 | Nivå i EA.E2 | Unimportant |
| 314c | 4 | 311KB213 | Difftryck VB50 | Unimportant |
| 314c | 5 | 531KA077 | APRM effekt | Important |
| 314c | 6 | 316KB402 | Nivå kondensationsbassäng | Important |
| 314c | 7 | 211KW111 | Medelvärde reaktortryck | Important |
| 314c | 8 | 112KA402 | Nivå intagskanal | Unimportant |
| 314d | 1 | 583KC201 | Difftryck drywell-wetwell | Important |
| 314d | 2 | 754KC113 | Tryck efter VC415 | Important |
| 314d | 3 | 422KA113 | Tryck i MÔH-EA1 | Unimportant |
| 314d | 4 | 722KB503 | Temperatur efter 331EB2,EB3 | Unimportant |
| 314d | 5 | 211KA041 | Reaktortryck | Important |
| 314d | 6 | 354KD404 | Stängningsnivå grupp D4 | Unimportant |
| 314d | 7 | 316KD501 | Temperatur kondensationsbassäng | Important |
| 314d | 8 | 314KB101 | Tryck avblåsningsrör efter VB1 | Important |

(IPAQ documentation - prepared by original experimental staff)

Figure S6: Development criteria, parameters queried, and importance categorization.

#### **OOTL**

The OOTL questionnaire was administered after each trial. It contained four items (shown in Figure S7) rated on a seven-point scale with the endpoints: I had no difficulties, I had great difficulties. The introduction to the questionnaire said: “Did you experience any problems in connection with the procedure execution? Below you will find four questions addressing exactly those kinds of problems. We are particularly interested in problems that were caused by the formatting of the procedure, i.e., paper or COPMA. Please respond to the questions on a scale ranging from 1 (no difficulties) to 7 (great difficulties).”



Figure S7: OOTL query items.

#### **Human-Automation Cooperation**

Human-Automation cooperation was measured on 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 eight items on a seven-point visual analog scale:

1. To what extent did you immediately understand the information that COPMA provided/you obtained from the paper-based procedures? [1=never; 7=always]
2. To what extent did COPMA/paper-based procedures provide relevant information about its activities/the activities they prescribed? [1=no relevant information; 7=all relevant information]
3. To what extent did you receive relevant information from COPMA/the paper-based procedures in time to benefit from it? [1= never; 7=always]
4. To what extent did the suggestions provided by COPMA/the paper-based procedures influence the quality of your task performance process? [1=no influence, 7=substantial influence]
5. To what extent did COPMA perform the activities you expected it to do?/did the paper-based procedures describe the activities you expected them to describe? [1=none of the expected activities; 7=all of the expected activities]
6. To what extent did COPMA perform the activities you requested of it /To what extent was it possible for you to find the information you requested in the paper-based procedures? [1=never; 7=always]
7. To what extent did COPMA support the achievement of the operational goal? / To what extent did the paper-based procedures support the achievement of the operational goal? [1=no support, 7=complete support]
8. Overall, how would you characterise the cooperation between you and COPMA? / the usability of the paper-based procedures? [1=very poor; 7=very good]

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