Prior knowledge and demographics

The study participants were asked about their prior knowledge regarding process modeling and execution. In case they are familiar with process models, they are also asked to assess how long they have been working with process models and in which context.
Second, they were asked about which type of instructional representation (primarily text-based, image-based or list/diagram-based) they prefer in general to teach themselves new activities.
Third, demographic data, such as age, gender, highest level of education, profession, job position, and working experience were recorded.

Effectiveness, mental efficiency, satisfaction

The first author/investigator was present during the whole experiment and observed the participants during process execution, measured the execution time and provided support when needed (participatory observation).
The result of a process execution was evaluated by the investigator regarding *precision and completeness* (**PAC**). Precision means how straightforward and independent the study participant completed a task. It is assessed on 4 levels, with a maximum of 3 points *("The user executed the task without any issues."*) and a minimum of 0 points *("The user executed the task with permanent support."*). The completeness is assessed on 3 levels with 2 points (*"The task has been completed"*), 1 point (*"partially completed"*) and 0 points ("*has not been completed*").
The sum of achieved scores in precision and completeness represents the *effectiveness* of a task execution. The effectiveness is expressed as percentage value, while 100% (5 points) effectiveness corresponds to a flawless task execution.

After each process execution, the study participants received a hard-copy questionnaire to assess their perceived mental effort self-administered. The *Scale to measure Subjective Experienced Effort* (**SSEE**) is a 220 point scale labeled with 7 literal expressions reaching from "almost no effort" to "extreme effort". The subjects were asked to mark an appropriate point on the scale. The marked point was used to determine *mental efficiency*, where low scores and thus low perceived mental effort mean high efficiency regarding usability of a process model.

At the end of each process execution, *satisfaction* was measured by a self-administered hard-copy of the *Questionnaire of the Subjective Consequences of Use* (**QUESI**). It covers 14 questions concerning the task instructions that can be answered on a 5-point Likert scale ranging from "*fully disagree*" to "*fully agree*".
Thereby, we slightly modified the original questions to fit our study design.
Each question is assigned to one of the five criteria of subjective consequences of use: Low subjective mental workload, high perceived achievement of goals, low perceived effort of learning, high familiarity and low perceived error rate. The mean value of all criteria results in the *QUESI* score, while a high value corresponds to higher satisfaction.

Interpretation and comparison of process models

In the pretest of the study, we used the think aloud method [Eccles2017] to gain more insight into the cognitive processes of study participants during task execution.

From comments expressed by the subjects, we deduced that the instructions can be interpreted differently. In many cases, this did not affect the success of a task execution, but reduced the possible solution space.
We observed that study participants derive concrete component positions and orientations from task instructions.
For example, some participants read specifications where and how exactly components should be placed from pictorial task annotations.
We noticed also discussions about whether there is a prescribed order in which the components have to be placed. The components that had to be manipulated in a task were always sorted alphabetically in the task instruction. Although this was only for simplification, the order was interpreted as mandatory from some participants.
From this end, we designed a short questionnaire to interview study participants about their interpretations regarding the task instruction.
We briefly asked whether the following aspects were based on own decisions or by specification of the task instruction: position of components in the scene, position of components in relation to each other, orientation of components and the manipulation order of components.

At the end of the study, subjects were asked to have a look at three process models simultaneously. All models described the same task setting but each with a different representation format of the task annotation. The study participants were randomly assigned the process models that related to one of the three task settings.
They were asked to compare all representations and specify the one that seemed most straightforward to quickly understand the task instruction.

Participants

Our study population comprises 50 participants who were mainly recruited from professional and personal networks.
The sample was heterogeneous in all demographic aspects, i.e., age, gender, focus and level of education, and experience in process model creation and execution.
The decision to have a heterogeneous distribution in all aspects is motivated by two issues: First, the study is about intuitive comprehensibility, i.e., we want to find out whether in particular people without prior knowledge in process modeling and execution are able to understand different representations of task annotations with little effort. For this end, we ensured that both levels of experience are about equally represented in the study.
Second, we know from industrial process applications that people with a wide variety of backgrounds are employed to execute process models.
These can be career changers with different professional focuses, but also persons with varying qualifications, experience and prior knowledge.
In our study, subjects with professional backgrounds in engineering, media, economy, social research, health, medicine, nursing, crafts, education, law and IT took part.

All materials provided to the study participants were in German language.
The ability to read and speak German fluently was therefore an inclusion criterion.
The study was conducted at the workplaces of the study participants, i.e., in workrooms and laboratories of universities, institutes and companies and was supervised by the first author of this paper.

Since each study participant had to execute a series of process models and has to answer the same questionnaires multiple times, it was reasonable to assume the occurrence of learning and order effects.
To counteract those, each study participant was randomly assigned to a combination of task setting (placing, stacking, sorting) and representation format (text, image, diagram).

[Eccles2017] Eccles, D., Arsal, G.: The think aloud method: what is it and how douse it?. Qualitative Research in Sport, Exercise and Health **9**(4), 514--531 (2017)