MASTER'S THESIS

Impact of layout’s flow consistency in Business Process Models on their understandability

Luna, C.J.

Award date:
2021

Impact of layout’s flow consistency in Business Process Models on their understandability

Opleiding: Open Universiteit, faculteit Betawetenschappen

Masteropleiding Business Process Management & IT

Degree program: Open University of the Netherlands, Faculty Science

Master of Science Business Process Management & IT

Course: IM0602 BPMIT Graduation Assignment Preparation
IM9806 Business Process Management and IT Graduation Assignment

Student: Carlos J. Luna

S. Identification number:

Date: 27/Jun/2021

Thesis supervisor Irene Vanderfeesten

Second reader Ben Roelens

Version number: 1.0

Status: Final version

Abstract

Business process models represent step-by-step rules, as flow of activities, to reach a specific business goal. Understanding these models is then of importance. This understandability is impacted by different model factors, like the model's direction (part of the visual layout). Modeling conventions and guidelines promote a direction to design a model (i.e. left-to-right or top-to-bottom). In the end, the modeler has the freedom to decide how a process model will be designed. Some models are not designed 100% consistent with the same direction. Others are re-designed to fit into a format size, changing the model's layout and *flow consistency*. This research seeks to identify how this decision of changing the model's layout impacts the model's understandability, searching for a relationship between the model's *flow consistency* and its understandability. This is researched with an online experiment using two Process Models in four different layouts (each with different *flow consistency),* comparing *left-to-right* layouts with less researched model layouts, i.e. z*ig-zag* and *banana* layouts. While there seemed to be a tendency suggesting that banana and zig-zag layouts are less understandable than left-to-right layouts, no significant effect could be established. Also, no significant relationship was proved between the model's *flow consistency* and its understandability.

Key terms

Flow Consistency, Flow Direction, Understandability, Model Layout, Business Process Models.

Summary

Business process models represent step-by-step rules, as a flow of activities, to reach a specific business goal. The understandability of the process model is impacted by different model factors, like the model's visual layout. The direction of the model is part of the visual layout of a model. Process model conventions and guidelines advise designing a process model following one direction (left-to-right or top-to-bottom). In practice, modelers have the freedom to choose how the model will be designed. These process models may follow one direction consistently. Other models will only be able to be designed with a lower flow consistency. Sometimes the modelers might choose to change the layout of the process model to fit it into a format size, changing the flow consistency of the model. The impact of these actions on the model's understandability is unknown and the research to investigate this impact is explained in this paper.

After doing literature research on this topic it was found that *flow consistency* is a fairly new term in process modeling which has not yet received much attention from the scientific community.

A definition of *flow consistency* was found, but this was considered to be incomplete, for which in this research a new definition is proposed: **Flow direction consistency (in short *Flow consistency*) is *the extent to which the position and temporal logical ordering of the activities of a process model follow the same direction.***

A framework to measure the *flow consistency* of a process model was found as well, and this can be calculated by considering the edge's direction, or by considering the relative graphical position of pair of activities in strict order relation. For this research, a combination of both is used as human perception seems to give more importance to the location of the activities. The graphical position of activities in the model in a strict order relationship are categorized by a radius to determine the flow direction. These relationships will get assigned one direction (East, South, West or North) when the angle is 0°, 90°, 180° or 270°, and will be assigned to two directions in the rest of cases, e.g. South and East with an angle of 45°. The *flow consistency* will then be calculated as a ratio of relationships categorized in one direction and the total of relationships categorized. The highest ratio determines the *flow consistency* and flow direction of the process model.

The impact the *flow consistency* has on the model's understandability has not yet been researched. An exploratory study was designed to research the impact of different model layouts (from a left-to-right layout to a banana or zig-zag layout) with different flow consistencies. The online experiment setup with 118 participants provided 51 valid responses.

Understandability has been measured in the experiment objectively using the variables Task Effectiveness (correct number of tasks divided by the total of tasks) and Task Efficiency (Task Effectiveness divided by the time needed to perform the tasks). And it was also measured subjectively using the variable Perceived Understandability.

In the experiment, two Business Process Models (Model A-designed for this experiment, and Model B-a real-life process model) in four different layout versions (1, 2, 3, and 4) were used. Both models contain the same number of activities, but model B have two times more gateways than model A. The first layout was designed in a Left-to-right direction, the second and third following a z*ig-zag* layout, and the final following *banana* layout. The model layouts (left-to-right) have the highest *flow consistency*. The other layouts have lower *flow consistency.*

In total six null hypotheses were formulated. The first two hypotheses to research if models with a left-to-right layout are more understandable than process models with a layout with less flow

consistency (i.e. banana and zig-zag layout). The third and fourth hypotheses to research if there is a significant relationship between the model's *flow consistency* and its understandability. The last two hypotheses, to research a possible moderating impact of the personal factors of the participants. These two last hypotheses were going to be tested only if a significant relationship was proved in the third or fourth hypotheses.

From the Kruskal Wallis test no significant difference (p>0,05) is evidenced indicating process models with a zig-zag and banana layout are less understandable than models in a left-to-right layout (objectively and subjectively measured). Therefore, the null hypotheses H1 and H2 were accepted. From the Spearman's rank correlation analysis no significant relationship (p>0,05) was proved between the process model's *flow consistency* and its understandability. Consequently, the null hypotheses H3 and H4 were not rejected.

The last two hypotheses (H5 and H6) were aimed to research a moderating impact of the personal factors if the null hypotheses H3 and H4 were rejected. As H3 and H4 were not rejected, the null hypotheses H5 and H6 were not statistically tested.

While it seemed to be a tendency in the data pointing out that process models in a left-to-right layout are more understandable than models in a banana layout and some models in a zig-zag layout, no significant evidence was established. Therefore it is concluded that changing the layout of a process model from a left-to-right layout to a zig-zag or banana layout (reducing the *flow consistency)*, has no significant impact on its understandability. Also is concluded that no significant relationship exists between the model's *flow consistency* and its understandability.

The superiority of a left-to-right layout has been theoretically proved but is not empirically confirmed in this research. It is assumed that 'banana' and 'zig-zag' layouts increase the cognitive effort necessary to understand them, but no significant findings were evidenced that banana and zig-zag layouts are significantly less understandable. It is speculated that from the model layouts tested, none of them underperformed in part by humans' ability to adapt quickly to uncommon reading directions. However, further research about the impact of the *flow consistency* of a business process model claims to be done.

Contents

List of Figures

List of Tables

1. Introduction

1.1. Background

Business process modeling refers to the graphical representation, as a flow of activities, from step-by-step rules specific to reach a business goal. Process models are common across industries and are important among others to get a better understanding of processes and improve the communication of a process between stakeholders (Bernstein & Soffer, 2015; Figl & Strembeck, 2015; Havey, 2005). As this should be understood correctly, their understandability is of high importance.

Different process model factors influencing the understandability of the model have been defined (e.g. Dikici, Turetken, and Demirors (2017); Figl (2017)), but the impact of some of these factors (i.e. Visual layout) on the understandability of a process model is not yet fully identified.

In practice, modelers can design models in a *one-flow direction* or can also create less consistent flow models (see Figure 1). Is this kind of model less understandable? The impact of the *flow consistency* of a model on its understandability is unknown. The latter will be investigated in this research. If a Process model can't be understood by the user, the model is useless. (Becker, Rosemann, & Uthmann, 2003; Figl, 2017; Lindland, Sindre, & Solvberg, 1994; Mendling, 2009).

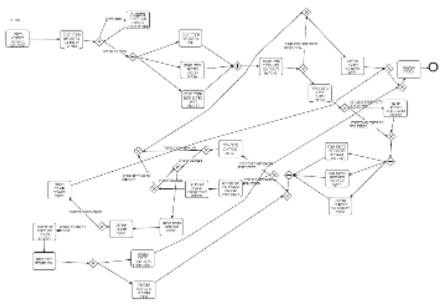


Figure 1 - Example of a business process model with no defined visual flow direction. (Bernstein & Soffer, 2015)

1.2. Exploration of the topic

Model quality has been divided into syntactic, semantic, and pragmatic quality (Kesh, 1995; Krogstie, Lindland, & Sindre, 1995; Lindland et al., 1994). This research will focus on the pragmatic quality of the model as it refers to the usefulness of the model and understandability by a human user. (Bernstein & Soffer, 2015; Dikici et al., 2017; Figl, 2017).

The understandability (or comprehension) of a process model is the extent to which the information in a process model is easy to be correctly understood by the reader. It is associated with the easiness of using the model and the effort needed to read it and correctly interpret it. (Dikici et al., 2017)

Visual layout, or *Layout aesthetics,* refers to the visual properties of a drawing, between others how the parts of the model are arranged or laid out. This is known as secondary notation as it communicates extra information by other means than official syntax (Effinger, Jogsch, & Seiz, 2010).

BPMN guidelines *suggest* using a left-to-right or top-to-bottom *flow direction*. (Object Management Group, 2011). In practice, modelers can design models not only in *one flow direction* but can create less *flow consistent* models to fit the model to a specific size format without reducing the overall size. (Figl & Strembeck, 2014; Leopold, Mendling, & Günther, 2015).

1.3. Problem statement

Attempts have been made to identify the impact of the model's flow direction on its understandability using models with *one flow direction* but no significant impact has been found (Burattin, Bernstein, Neurauter, Soffer, & Weber, 2018; Figl & Strembeck, 2014, 2015). These findings might not be applicable for models with less consistent *flow direction* as this seems not to be researched yet.

The following problem statement can be defined: **The impact of changing the *layout* of business process models on the understandability of the model is unknown or unclear.**

1.4. Research objective and questions

The main objective of this research has been defined as: **to identify the relationship between the *flow consistency* of a business process model and its understandability.**

To achieve this objective, a main question is defined to be answered by this research. The main question is: **What is the relationship between the *flow consistency* of a business process model and its understandability?**

To help to answer the main question, first, a definition of *flow consistency* will be necessary. Second, a framework to measure the *flow consistency* is required. Third, previous research regarding the topic needs to be gathered and considered. Finally, the impact of the *flow consistency* on the understandability of a process model needs to be tested. For this purpose the following sub-questions have been defined:

- How can the *flow consistency* of a business process model be defined?
- How can the *flow consistency* of a business process model be measured?
- What can be found in the literature about the *flow consistency* of a business process model and its impact on the understandability of the model?
- Is a business process model with less *flow consistency* less understandable than a business process model with higher *flow consistency*?

The main question will be answered by conducting a literature review supported by empirical research. The first three sub-questions will be answered by conducting a literature review. The fourth sub-question will be researched empirically through an experiment.

1.5. Motivation/relevance

This research will help to advise process modelers in practice in case a model doesn't fit a size format, by defining the impact of creating less *flow consistent* models have on their understandability. This is the practical relevance and will contribute to design more understandable models and therefore an improved common understanding among the stakeholders of the process model. The same knowledge can be applied to the creation of algorithms for modeling software.

The impact of the flow consistency of a business process model on its pragmatic quality is unknown. This research aims to close this gap in the scientific literature. This is the scientific relevance of this research by further identifying the impact of visual layout features on the understandability of a process model. It is of notorious importance as this relates to the usefulness and understandability of the model.

1.6. Main lines of approach

To answer the main question a systematic approach is followed which will be described in the next sections. The remaining of this paper is structured as follows. In Section 2, the literature research approach and its results are presented, forming the theoretical framework of this research. Section 3 describes the methodology followed to develop the empirical research and section 4 presents its results. Section 5 closes this research presenting the discussion, conclusions, and recommendations based on the findings of the previous findings.

2. Theoretical framework

This section provides the knowledge necessary to answer the first three sub-questions. In continuation, a general description of the literature research approach and its implementation is presented. The detailed approach including the sources, search parameters, search terms, evaluation criteria, and the implementation of this approach can be consulted in Appendix I.

2.1. Literature research approach and implementation

A systematic literature research is conducted according to Saunders, Lewis, and Thronhill (2019). The online "Open University" library and Google Scholar were identified as main and secondary sources respectively. The search terms (see Table 1) were generated using the method 'building blocks' (van Veen & Westerkamp, 2010), more detailed explained in Appendix I. To select the most relevant literature the systematic evaluation process Prisma 2009 Flow Diagram from Moher, Liberati, Tetzlaff, and Altman (2009) is followed. This process is illustrated in Figure 2 indicating the number of publications retrieved per step. The evaluation criteria can be consulted in Appendix I.

Table 1 - Search Queries

Search term	Related sub-question	Search query
1	1	(("flow consistency" OR "flow direction" OR "model direction") AND ("process model*") AND (defin* OR expla* OR interpret* OR determin* OR detect*))
2	2	("flow consistency" OR "flow direction" OR "model direction") AND ("process model*") AND (measure* OR evaluation OR assessment OR quantif*))
3	3	(("flow consistency" OR "flow direction" OR "model direction") AND ("process model*") AND (understand* OR comprehen* OR read*))

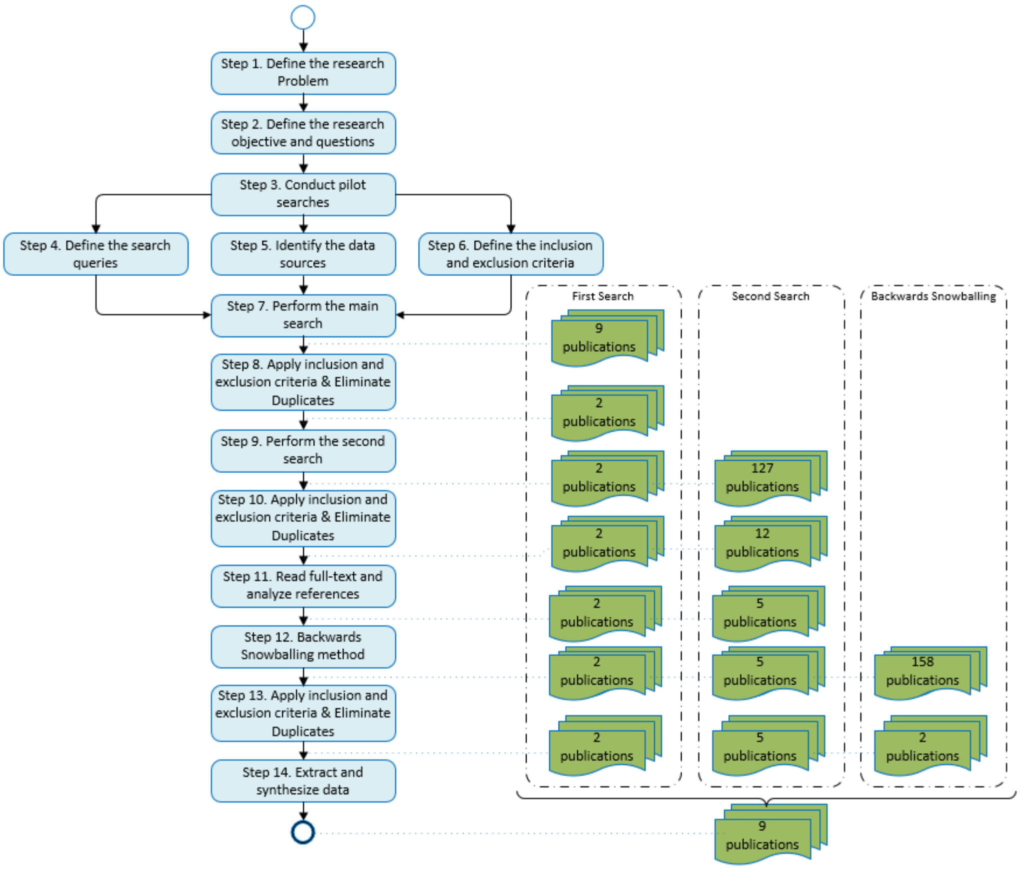


Figure 2 - Prisma Flow diagram - literature selection

The first search query resulted in nine articles (six duplicates) from which two articles resulted to be relevant. These articles are listed in Appendix III. Because of the low number of articles found, a second source was consulted (namely Google Scholar), leading to a second search. After using the second source several limitations were identified. These are described in Appendix IV as well as the actions taken to pass them.

The second search delivered a total of 127 articles, including Systematic Literature Reviews (SLR) about process models. Three articles and nine SLRs passed the evaluation phase, from which five publications passed the eligibility phase. By applying the *backwards snowballing* method to the final list from both searches, 158 new articles were retrieved. Two of them passed the eligibility phase.

An overview of the final list of articles that resulted can be consulted in Appendix V. Using a second source resulted to be a good decision adding five relevant publications. The backwards snowballing method added two publications.

2.2. Results and conclusions

This section explains the results of the literature search and explains the numerous concepts one by one that builds on one another answering the first three sub-questions.

2.2.1. Defining Flow consistency

Only one definition of *flow consistency* was found in the literature. Burattin et al. (2018, p. 638) refer to it as: "T*he flow of the model can be in one definite direction from the beginning till the end of the model. Alternatively, it can be unclear or changing throughout the model to different direction*". They refer to models with one direction (i.e. left-to-right, or top-to-bottom) as having a 'consistent *flow direction*' and state that "*the consistency of flow measures the extent to which the layout of a process model reflects the temporal logical ordering of the process*" (Burattin et al., 2018, p. 639).

The definition of *flow consistency* from Burattin et al. (2018) went through a validation process in this research, this can be consulted in Appendix VI. It was noticed that the authors don't take explicitly into consideration the direction of the model in their definition. The way the model's objects (i.e. activities, start/end event; control-flow objects) are arranged or laid out in the canvas determines the visual *flow* ***direction*** of the process model. Therefore the definition from Burattin et al. (2018) has been completed.

To answer the first sub-question a new definition is proposed: **Flow direction consistency (in short Flow consistency) is *the extent to which the position and temporal logical ordering of the activities of a process model follow the same direction.***

2.2.2. Measuring the flow consistency

Burattin et al. (2018) developed three metrics (M-E1, M-E2, and M-BP) to calculate the *flow consistency*. M-E1 and M-E2 consider the direction of the edges (see Figure 3). M-BP considers the relative graphical localization of pairs of activities.

M-E1 classifies the direction of each edge using the radius in Figure 3a. The greatest ratio between edges in one direction and the total of edges determines the *flow direction* and *flow consistency* of the model. M-E2 is similar to M-E1, but each edge receives two directions (Figure 3b). Is not clear if edges with angle 0°, 90°, 180°, 270° also get two directions.

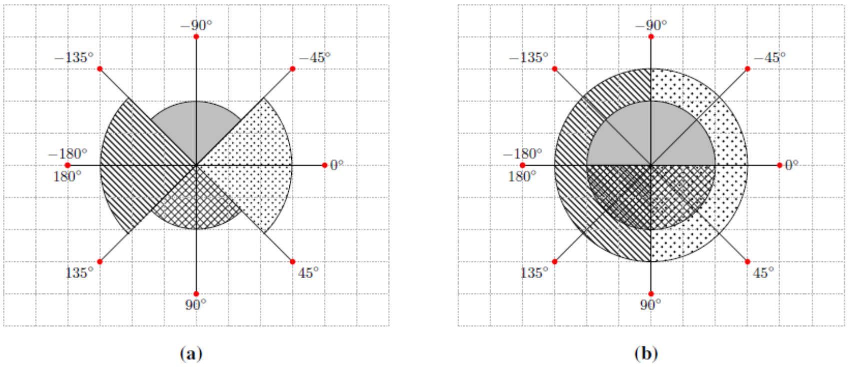


Figure 3 - (a) Radius to assign one direction to edges (M-E1). (b) Radius to assign two directions to edges (M-E2), each direction overlaps with the two adjacent ones. Directions are defined as North (gray filled area), East (dotted area), South (grid area), and West (lined area). Images extracted from Burattin et al. (2018).

M-BP uses the relative localization of each pair of activities in strict order relation (more detail in Appendix VI). A strict order relation exists when one activity occurs always before the other. The behavioral profiles are also considered in the algorithm of Wang, Sun, and OuYang (2018). M-BP considers which pairs of activities in strict order are 'graphically before' (the target node is placed east or south of the source node). The authors considered left-to-right and top-to-bottom as the direction a model should have, no arguments were provided why. To compute the *flow consistency* the 'graphically before' relations are divided by the total number of strict order relations.

Two disadvantages for the M-BP metric were found. The first, considers South and East as *flow consistent* combining 2 different directions. The second, it doesn't provide the *flow direction*. One example of the models used in this framework and the result of these three metrics have been added in Appendix VII. The framework of Burattin et al. (2018) is explained in Appendix VIII.

The authors conclude that the metrics M-E2 and M-BP correlate with human perception. Human perception gives more importance to the position of the activities rather than the edge's direction. This perception was evaluated subjectively and it is not to be understood as comprehension.

To answer the second sub-question, the *flow consistency* of a process model can be calculated in two forms. The first, by considering the edge's direction. The second, by considering the relative graphical-position of pair of activities in strict order relation. For this research, a combination of both will be used. Human perception seems to give more importance to the location of the activities. The activities in strict order will be categorized by the radius used in metric M-E2 (this radius provided more significant results) to determine the *flow consistency* and *flow direction*. Another necessary change is when the direction is 0°, 90°, 180° or 270°, only 1 direction is given to the arrow. Start and end events are not taken into consideration. All this improves the metrics solving the limitations found.

2.2.3. Impact of Flow consistency on understandability

If the modeler wants the model to be useful it needs to be comprehensible, "not even the most brilliant solution to a problem would be of any use if no one could understand it" (Lindland et al., 1994, p. 47).

No article was found researching the impact of the *flow consistency* on the model's understandability. As the model direction is one of the variables to define the *flow consistency*, research regarding the *flow direction* and its impact on understandability was reviewed to build fundaments for this research.

Figl and Strembeck (2014) stated that in principle, process flows can be modeled in four directions: left-to-right, right-to-left, top-to-bottom, or bottom-to-top, and from a theoretical perspective left-to-right flow direction is beneficial.

Figl and Strembeck (2015) and Kretschmann (2019) attempted to empirically confirm this superiority using models with the previous four model *flow directions* mentioned using different methods to measure the model's understandability. Even when theoretically there is a relationship between the flow direction of a process model and its understandability, this was not statistically confirmed. Figl and Strembeck (2015) speculate that this result may be explained in part by humans' ability to adapt quickly to uncommon reading directions (e.g., right-to-left). Model readers adapt their perceptual strategy after four trials reading diagrams with a right-to-left flow direction. (Winn, 1983).

The *flow consistency* was not considered as a variable. Other *flow directions* different than the previous four mentioned were not considered either. It is valid to think that the understandability was not impacted as the *flow consistency* remained high by using one-directional models.

In reality, the modeler has the freedom to decide which direction to use. Some models are not possible to be drawn 100% consistently in the same direction. (Figl & Strembeck, 2015). This phenomenon can be found in practice when the modeler decides to design *'zig-zag'* models (Figl & Strembeck, 2015), models not following consistently one flow direction, but form a sort of zig-zag pattern back and forth from one end of the canvas to the other. Leopold et al. (2015) refer to other kinds of models, '*banana*' models, designed in three or more different flow directions, creating a 'curved' (banana shape) path of activities, stating there is no reason for this. Some users apply this practice to fit the model to a specific paper size format, instead of reducing the overall size of model elements and labels, but Leopold et al. (2015) assume this only confuses readers due to an increased cognitive effort is needed. The impact of this low *flow consistency* is unknown.

Even when the location of the model's object in the canvas impact directly the model's layout, and possibly the model's understandability as well, no guidelines were found regarding where to locate the objects of a process model (for more detail see Appendix IX). There is a link between the model's *flow direction* with written language direction, but there is also a lack of agreement. A deep dive into this interesting link can be consulted in Appendix X.

To answer the third sub-question, in the business process modeling area, the relationship between the *flow consistency* and the model's understandability is not yet been researched. The impact of the *flow direction* on the model's understandability has been theoretically proved but not empirically. Other kinds of layouts having low *flow consistency* (e.g. 'Zig-zag' and 'banana' layouts) have not yet been researched. Leopold et al. (2015) assume these layouts have an impact on their understandability increasing the cognitive effort necessary to understand them.

2.3. Objective of the follow-up research

After the literature review, the three sub-questions were answered. A definition of *flow consistency* has been found. Metrics to measure the *flow consistency* of process models were found. Nothing was found about the impact of the *flow consistency* of process models on their understandability. There are no attempts done yet to confirm this link, but some researchers assume that models having a low *flow consistent* layout (e.g. zig-zag and banana) increase the cognitive effort necessary to understand the model. Leopold et al. (2015). The further objective of this research is to investigate empirically if these layouts being less flow consistent are less understandable than more flow consistent 'left-to-right' layouts, researching a possible relationship between the model's *flow consistency* and its understandability. Answering the fourth sub-question and main question.

3. Methodology

Given the objective of this research, it is considered to follow mainly the objectivism-positivism philosophy to identify if a relationship exists between the *flow consistency* of a business process model's layout and its understandability by analyzing measurable facts. (Saunders et al., 2019).

3.1. Conceptual design: select the research method(s)

The data needed for this research are the process model's *flow consistency*, understandability of the model, and the personal factors from the participants. These variables will be quantitatively measured. A deductive approach using quantitative data is the most suitable to identify a relationship between variables and answer the fourth sub-question. An explanatory study is followed to establish a causal relationship between variables. (Saunders et al., 2019).

Different possible approaches and strategies were considered before the most suitable approach and strategy were selected. (Saunders et al., 2019). The techniques: action research, case study, grounded theory, and survey were considered not suitable for the objective of this research. (Saunders et al., 2019). The considerations taken are mentioned in Appendix XI.

A mono-method quantitative design is followed by conducting an experiment, allowing to measure the impact on the understandability of the process when manipulating the *flow consistency*. The findings are possible to be generalized if they are statistically representative. (Saunders et al., 2019).

3.1.1. Research model and hypotheses

An experiment uses hypothetical explanations (Saunders et al., 2019). The hypothesis and the variables to be measured need to be defined. Figure 4 provides an overview of the research model with the variables. The operationalization of the variables can be consulted in Appendix XII.

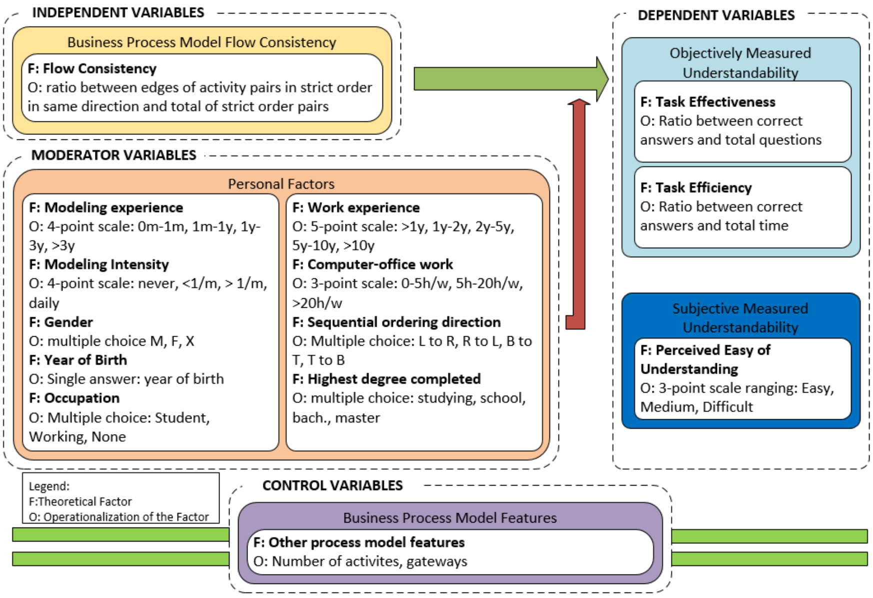


Figure 4 - Research Model

To identify the impact on the model's understandability, the *flow consistency* will be manipulated during the experiment by using process models with different *flow consistency*. If a significant relationship is statistically proved, the personal factors will allow getting insights into how these influence the impact found. The control variables will remain unchanged during the experiment to assure any change in the understandability is due to the change in the flow consistency.

To research if less consistent process models are less understandable than process models having a higher flow consistency the following groups of null hypotheses were formulated:

H1 – Left-to-right process models are not significantly more understandable than a zig-zag or banana layout, i.e. (a) understandability Task Effectiveness, and (b) understandability Task Efficiency.

H2 – Left-to-right process models are not significantly perceived easier to understand than a zig-zag or banana layout, i.e. Perceived Ease of Understanding

To research if there is a relationship between the *flow consistency* and the model's understandability the following groups of null hypotheses were formulated.

H3 - The *flow consistency* of a process model has no relationship with the model's understandability, i.e. (a) understandability Task Effectiveness, and (b) understandability Task Efficiency.

H4 – The *flow consistency* of a process model has no relationship with the model's Perceived understandability, i.e. Perceived Ease of Understanding.

If a significant relationship is proved in the above hypotheses (H3 and H4), the following group of null hypotheses is formulated to test the moderating impact of the personal factors.

H5 – Personal factors (i.e. Work Experience, Computer Office Level, Education level, and Experience /intensity level) do not influence the relationship between the *flow consistency* of a process model and its understandability, i.e. (a) understandability Task Effectiveness, and (b) Task Efficiency.

H6 – Personal factors (i.e. Work Experience, Computer Office Level, Education level, and Experience /intensity level) do not influence the relationship between the *flow consistency* of a process model and its Perceived understandability, i.e. Perceived Ease of Understanding.

3.2. Technical design: elaboration of the method

A between-subject design is followed creating groups of participants. This to measure different variables at the same moment avoiding experimental bias (e.g. learning effect), often done in this kind of experiment. (Field & Hole, 2003). In the same experiment, the influence of the personal factors on the researched impact is statistically tested. More information about the participants, the design, the material, and the experiment setting is described in detail in Appendix XIII.

3.2.1. The Process models

Two process models (A and B) with textual labels in four different versions (i.e. 1, 2, 3, and 4, see Table 2) were designed to manipulate the *flow consistency*. These are included in Appendix XIV. The label type (abstract or textual) has been proved to have an impact on the model's understandability. (Mendling, Strembeck, & Recker, 2012). It was decided to use textual labels to match a real-life setting (models with textual labels). Figure 5 shows an example of each layout. Version A1 and B1 (left-to-right flow direction layout) don't fit into one screen or a 'landscape A4 paper'.

Table 2 - Different layouts used per model

Layout	Model A	Model B
1. Left-to-right	Model A1	Model B1
2. Zig-zag 1 (all model sections with same left-to-right flow direction)	Model A2	Model B2
3. Zig-zag 2 (middle model section with right-to-left flow direction)	Model A3	Model B3
4. Banana model	Model A4	Model B4

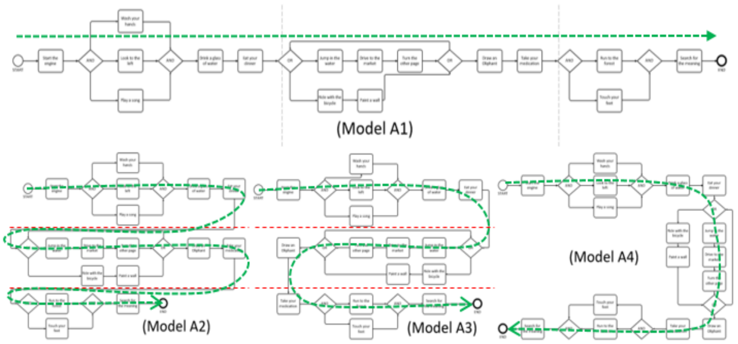


Figure 5 - Process Model A in the four different layouts. The visual flow direction is represented with a green arrow. Model A1- left-to-right flow direction. Model A2 and A3 are 'zig-zag' layouts creating visually three sections (see red line). Model A2- 'zig-zag' layout #1. Model A3- 'zig-zag' layout #2. Model A4- 'banana' layout.

The flow consistency of each model version has been calculated using the framework presented in section 2.2.2, by considering the relative graphical-position of pair of activities in strict order relation and classifying them in one or two directions depending on the angle. This can be consulted in Table 3. The highest *flow consistency* has been highlighted and is mentioned in the last columns.

Table 3 – Calculation of the flow consistency per model version

Layout	North (up)	East (right)	South (down)	West (left)	Total	Flow cons. North	Flow cons. East	Flow cons. South	Flow cons. Left	Model Flow cons.	Model Flow direction
A1 (left-to-right)	5	19	4	3	**31**	16%	**61%**	13%	10%	**61%**	**East**
A2 (zig-zag#1)	5	15	7	8	**35**	14%	**43%**	20%	23%	**43%**	**East**
A3 (zig-zag#2)	5	16	7	7	**35**	14%	**46%**	20%	20%	**46%**	**East**
A4 (banana)	7	9	10	7	**33**	21%	27%	**30%**	21%	**30%**	**South**
B1 (left-to-right)	5	22	5	0	**32**	16%	**69%**	16%	0%	**69%**	**East**
B2 (zig-zag#1)	2	11	9	5	**27**	7%	**41%**	33%	19%	**41%**	**East**
B3 (zig-zag#2)	2	13	9	3	**27**	7%	**48%**	33%	11%	**48%**	**East**
B4 (banana)	1	9	14	6	**30**	3%	30%	**47%**	20%	**47%**	**South**

It is clear that the *flow consistency* of the left-to-right layout is the highest among other layouts. Model A version A4 (banana layout) has the lowest flow consistency (i.e. 30%), but in model B, model version B2 zig-zag#2 is the one with the lowest *flow consistency* (i.e. 41%). It can be then concluded that the *flow consistency* of a process model will not only be determined by the visual pattern the flow of activities form, but also by the model's characteristics (e.g. number of activities, gateways, and loops). Both models have sixteen activities, one start, and one end event. Model A has six gateways and Model B has eleven (missing one join connector to be correctly designed). The left-to-right layout of Model B has a higher flow consistency than the left-to-right layout of model A. The same is visible in the zig-zag#2 and banana layouts, being this difference bigger in the banana layout. A banana layout seems to impact the model's *flow consistency* and also its flow direction.

3.2.2. The Experiment

The experiment protocol is presented in Figure 6. During the experiment, the participants will be randomly assigned to one of four groups. Each group should contain at least 30 valid participants to reach a normally distributed data set as per the Central Limit Theorem (Saunders et al., 2019). Each group will process the model versions as per Table 4. Special attention was given so the same group does not receive the same layout version (e.g. A1 and B1) or both zig-zag layouts (e.g. A2 and B3).

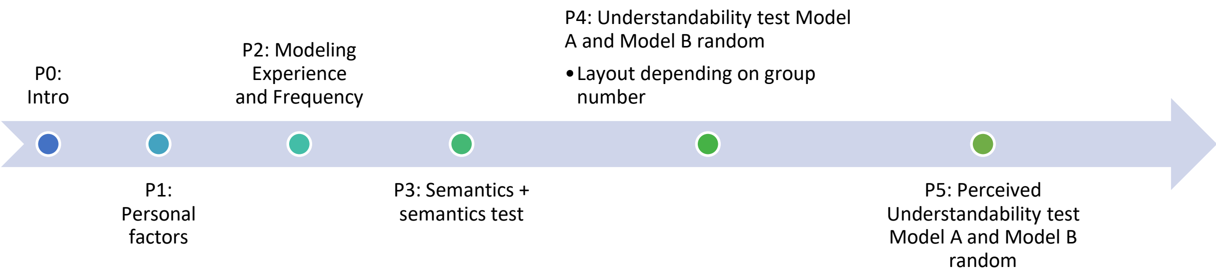


Figure 6 - Experiment protocol

Table 4 - Assignment of the model version to be processed per group

	First test model A	Second test model B
Group 1	Model A1	Model B3
Group 2	Model A2	Model B4
Group 3	Model A3	Model B1
Group 4	Model A4	Model B2

The participants will reply to the questions about personal factors, and their modeling experience and modeling intensity. This latter adapted from Mendling et al. (2012). Each group will solve nine understandability reasoning tasks per model based on questions used in similar experiments by Reijers, Freytag, Mendling, and Eckleder (2011). An example of these questions is: *Can activities 'Turn the other page' and 'Paint a wall' be executed at the same time for a case?* The participants will express their perception about the understandability of the processed models using the seven-point Likert scale used by Figl and Laue (2015).

An introductory explanation about semantics is given, based on the introduction of Mendling and van der Aalst (2006). Explaining how to read a model, evaluating their understanding with a test. The material and questions used during the experiment can be consulted in Appendix XV.

3.3. Data analysis

The relationship between the variables will be statistically tested using a parametric test to investigate whether or not the *flow consistency* of the model has an effect on its understandability. The most suitable test will be a MANOVA which is a parametric test commonly used in quantitative research to test the difference between more than three groups. (Saunders et al., 2019).

The data will be first analyzed per model to identify any significant impact on the understandability by changing the model's layout, then it will be tested for a possible relationship between the *flow consistency* and its understandability. The measures of the understandability from the models A2, A3, and A4 will be compared against the measures of model A1 (control group). Version A2 and A3 represent a zig-zag layout, but these are analyzed apart as the layouts are different. The same process will be done for Model B. The findings from both models will be compared to validate them.

During the data analysis process, it will also be statistically identified any 'moderator' influence the personal factors have on the relationship between the model's *flow consistency* and its understandability if there is a significant relationship found.

3.4. Research design's quality

To assess the quality of a research design the criteria reliability and validity are important indicators. (Saunders et al., 2019). This section will elaborate on the considerations taken to increase these criteria. The remaining weakness will be described.

3.4.1. Reliability

To guarantee a sound quantitative scientific research is done, several structured methodologies were followed. It is mentioned where these are applied and how this helps to give structure to the research process. An extra detailed level description is provided. This facilitates the replication of the steps at a later moment. A pilot testing is done to test the material is correct and understandable.

A similar experimental design used by Figl and Laue (2015); Figl, Recker, and Mendling (2013); Reijers and Mendling (2010) is adapted and used for this experiment. The understandability tasks to be performed by the participants were based on the questionnaire from Reijers, Freytag, Mendling, and Eckleder (2011). The understandability variables (i.e. Task Effectiveness, Task Efficiency, and Perceived Understandability) have been used by these researchers, validating the measurements already. The questions to measure the modeling experience and modeling intensity are adapted from Mendling, Strembeck, and Recker (2012).

To deem errors in replies from the participants it was requested to perform the experiment during a peaceful moment, without stopping, and without distractions and hurries. To avoid bias from the participants it was requested to answer the questions alone and with care, without guessing.

3.4.2. Internal and External Validity

An experiment gives the possibility to control the independent variable (manipulated during the setting), improving the internal validity of the findings. To improve the criterion and measurement validity of the variables, the same measures used in other articles accepted by the scientific community were used, assuring the correctness of the construct, content, and predictive validity.

Participants 'fatigue' is considered as it influences the experiment result. To pass this threat the questions have been shown randomly to the participants. Outliers are identified to determine whether a participant stopped too soon or replied too fast without reading the question. (Figl, 2017).

To improve the external validity it was not intended to have a laboratory setting, making it possible to relate the findings to the real world. Even when the participants represent a small sample, a high effort was done to select diverse profiles of participants to have a representative sample.

High care and attention were given to the design of the different layouts of both models, keeping all controlled variables (e.g. complexity) unchanged, using a real-life and a self-designed process.

As the type of labels influences understandability (Mendling et al., 2012), textual labels were used to match reality. To deem the risk of domain knowledge the labels in the process models were replaced by activities without relation. Also, activities that would imply any logical sequence order were avoided. All these impacts positively the generalizability of the findings.

3.4.3. Ethical aspects

It was guaranteed to the participants that the data is being recollected and analyzed anonymously, and is only used for this research. Data manipulation and storage comply with the GDPR. Participation in the experiment is voluntary. Participants accepted to participate before starting.

3.4.4. Threats

The results of this experiment might differ from studies where all participants are expert modelers. Also, it is possible that when repeating the same search queries the results might differ as in Google Scholar each day new articles will be added.

4. Results

This section will describe the implementation, progress, and results of the experiment. Deviations from the initial action plan are explained as well as the impact they have.

4.1. Data Preparation

The experiment was accessible online from March 13th until April 24th (2021). The participants were approached via social networks, and via email in the researcher's employer as planned. The data was downloaded, removed from the survey tool, and saved in a local drive without identificatory details.

During the 7 weeks, the experiment was online, 118 participants started it. Of these 118 participants, 48 persons stopped in the process and closed it without finishing. These incomplete replies were removed from the definitive data set. Originally it was planned to recollect at least 120 valid replies. The commitment to participate and finish the experiment was lower than expected. It is assumed this has an impact on the normality of the data set.

It is necessary to assure the participants were able to read a process model and to understand the questions of the experiment by understanding the basic semantics of business process modeling. Knowledge has been proved to have an impact on the model's understandability. (Mendling et al., 2012). 'Part 3' of the experiment was designed with this purpose. Twelve participants indicated not having understood the explanation and wanted to stop. Additionally, four participants scored less than 50% in the semantics test. These sixteen participants were removed from the data set as they didn't match the participant's profile.

It was identified that three participants were distracted while taking the test. These participants did much longer in one question (5836 seconds, 2068 seconds, and 1065 seconds respectively) while their average in other questions was much lower (41 seconds, 34 seconds, and 68 seconds respectively). As the Task Efficiency variable considers the time it took the participant to solve the tasks, the data recovered from these three participants is considered invalid. These replies were also removed.

Table 5 - Number of participants in the experiment

		Count
FINISHED	FINISHED - DISTRACTED DURING THE EXPERIMENT	3
	FINISHED - TASK EFFECTIVENESS IN SEMANTICS TEST < 50%	4
	DIDN'T UNDERSTOOD EXPLANATION	12
	NOT FINISHED	48
	FINISHED	51
	Total	118

Table 5 shows a summary of the number of participants remaining for the analysis and also how many were removed during the process of preparing the data.

After confirming the data set was valid and complete, it was checked the answers were correctly coded to be used for quantitative analysis. Most of the answers were coded already (online survey tool used), but it was identified that the answers from questions Q12 (Work Experience), Q13 (Computer Office Level), and Q15 (Education) still needed to be coded with numbers. Additionally, from the variable 'year of birth', the age of the participants was approximated and discretized in bins of 5 years long. This is done directly in IBM SPSS Statistics ™.

Nine variables needed to be created. These are described in Table 19 in Appendix XII.

4.2. Descriptive statistics

Table 6 - Distribution of participants between the groups

		Count
Group Nr.	Group 1 (Model A1-B3)	17
	Group 2 (Model A2-B4)	13
	Group 3 (Model A3-B1)	12
	Group 4 (Model A4-B2)	9
	Total	51

In Table 6 the distribution of the participants between the four groups can be consulted. After preparing the data, an imbalance was noticed between groups 1 and 4. The detail of the demographic distribution from the participants per group can be consulted in Appendix XVI.

Both genders participated almost equally (47% females - 53% males). Most of the participants (90.1%) are between 21 and 60 years old. The biggest group (31,4%) between 31 and 35 years old. 94,1% of the participants have work experience. Three participants (5,9%) are students. From the working participants, 60,4% have more than 10 years of work experience, followed by 25% with work experience between 5 and 10 years. All working participants have a school, bachelor, or master certificate (33,3%, 31,4%, and 29,4% respectively).

Most of the participants are used to work with a computer, matching the experiment setup (understand process models in a computer). 80% of the participants work more than 20 hours per week with a computer and seven participants (15,6%) between 5 and 20 hours per week.

22 participants (out of the total of 51 participants) expressed to have knowledge of process modeling. These participants were discretized into two groups: Low (59,1%) and High (40,9%) experience/intensity as per detail in Table 19 in Appendix XII.

When asking the participants in which direction they order an object in their imagination which is 'before' other object, 90,2% responded having a mental ordering 'first-then' in a direction from left-to-right (following the read direction). Four participants (7,8%) arrange the objects mentally in a right-to-left direction, which is less common. One participant answered to order the objects in a top-to-bottom direction. The latter might be a participant with a relation with an Asiatic culture where text is written from top to bottom.

4.2.1. Task Effectiveness

In the box plot charts in Figure 7, Figure 9, and Figure 11 several outliers were identified. The mean is highly influenced by extreme values (and the median is not). It is considered that the median represents better the central tendency of the variables. (Saunders et al., 2019). Model version A4 in Figure 9 is a good example confirming this decision.

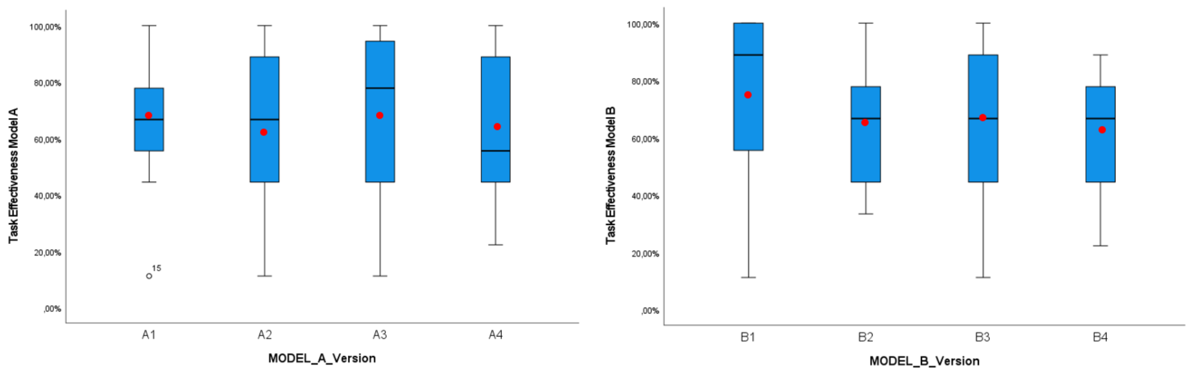


Figure 7 - Box plot chart Task Effectiveness. Average added as a red dot.

From the Figure 7 and **Error! Not a valid bookmark self-reference.** it is visible that the layout version A3 has the highest median on Task Effectiveness and model version A4 has the lowest median. In model B, model version B1 has the highest median on task Effectiveness, and model versions B2, B3, and B4 have the lowest median.

Only the zig-zag#2 layout (i.e. A3) has a tendency to perform better than the left-to-right layout (i.e. A1). And model version B1 (left-to-right layout) tends to perform better than the other 3 layouts. In Figure 8 the mean and median of the Task Effectiveness and the *flow consistency* of the model are visualized.

Table 7 - Descriptive statistics for Task Effectiveness variable

	Model Version	1	2	3	4
Mean	**A**	69,3%	62,4%	69,4%	63,0%
	B	75,0%	65,4%	67,3%	62,4%
Median	**A**	66,7%	66,7%	77,8%	55,6%
	B	88,9%	66,7%	66,7%	66,7%
Std. Deviation	**A**	21,7%	28,2%	32,5%	24,5%
	B	30,8%	23,9%	25,3%	22,0%

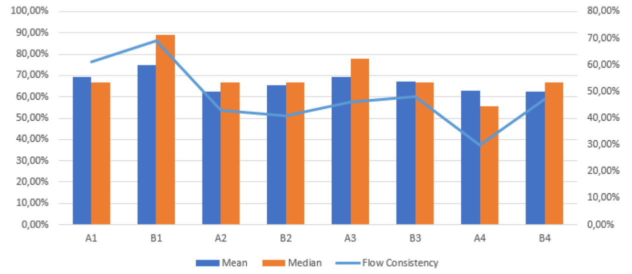


Figure 8 - Mean, median, and flow consistency per model version

4.2.2. Task Efficiency

Task efficiency is defined as the ratio between task Effectiveness and time, this can be understood as 'the highest score in the shortest time'. The highest the ratio, the better the efficiency of a process model. In In the box plot chart in Figure 9 and Table 8 it can be seen that model version A3 has the highest median and model version A2 has the lowest median on task Efficiency. Between the versions in Model B, model version B1 has the highest median, and model version B4 has the lowest median on Task Efficiency.

Table 8 the descriptive statistics for Task Efficiency are presented per model version.

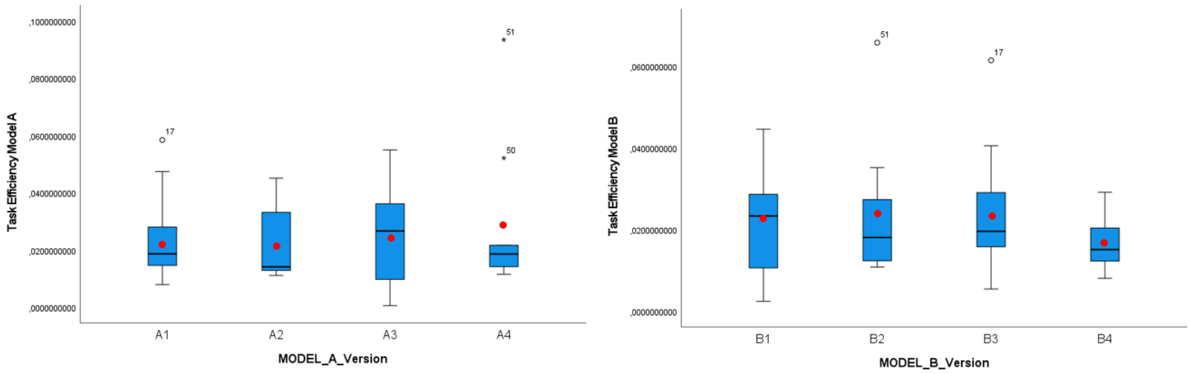


Figure 9 - Box plot chart Task Efficiency. Average added as a red dot.

In the box plot chart in Figure 9 and Table 8 it can be seen that model version A3 has the highest median and model version A2 has the lowest median on task Efficiency. Between the versions in Model B, model version B1 has the highest median, and model version B4 has the lowest median on Task Efficiency.

Table 8 -Descriptive statistics for Task Efficiency variable

	Model Version	1	2	3	4
Mean	**A**	0,023	0,022	0,025	0,029
	B	0,022	0,024	0,023	0,017
Median	**A**	0,019	0,014	0,027	0,019
	B	0,023	0,018	0,020	0,015
Std. Deviation	**A**	0,014	0,012	0,017	0,027
	B	0,013	0,018	0,013	0,007

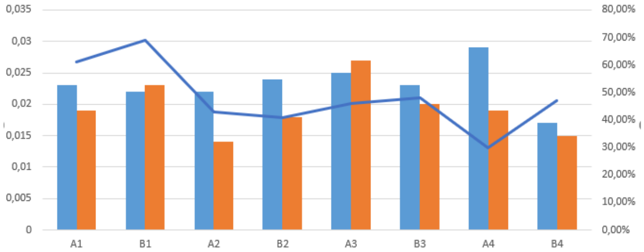


Figure 10 - Mean, median, and flow consistency per model version

Same as in Task Effectiveness, in Task Efficiency only the zig-zag#2 layout (i.e. A3) has a tendency to perform better than the left-to-right layout (i.e. A1). And model version B1 (left-to-right layout) tends to perform better than the other 3 layouts. In Figure 10 the flow consistency, the mean, and the median of the variable Task Efficiency are visualized.

4.2.3. Perceived Understandability

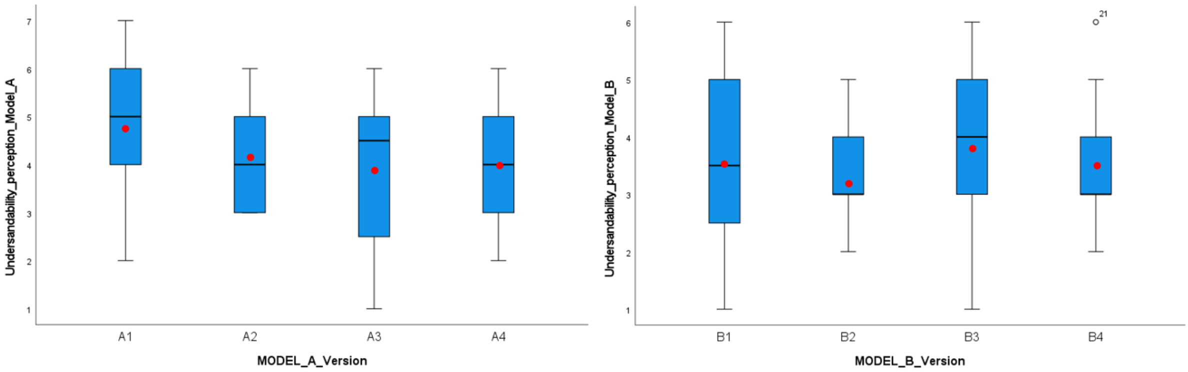


Figure 11 - Box plot chart Perceived Understandability

From Figure 11 and **Error! Not a valid bookmark self-reference.**, is visible that model A1 has the highest median on Perceived Understandability, and model versions A2 and A4 the lowest median. In model B, model version B3 has the highest median on Perceived Understandability. Model versions B2 and B4 have the lowest median.

The left-to-right layout (i.e. A1) has a tendency to be perceived easier than the other three layouts. And the zig-zag#2 layout (i.e. B3) tends to be perceived easier than the left-to-right layout (i.e. B1). These tendencies are different than the ones visualized in the objective understandability measures Task Effectiveness and Task Efficiency.

Table 9 - Descriptive statistics for Perceived Understandability variable

	Model Version	1	2	3	4
Mean	**A**	4,8	4,2	3,9	4,0
	B	3,6	3,3	3,9	3,6
Median	**A**	5,0	4,0	4,5	4,0
	B	3,5	3,0	4,0	3,0
Std. Deviation	**A**	1,4	1,1	1,7	1,2
	B	1,2	0,9	1,6	1,2

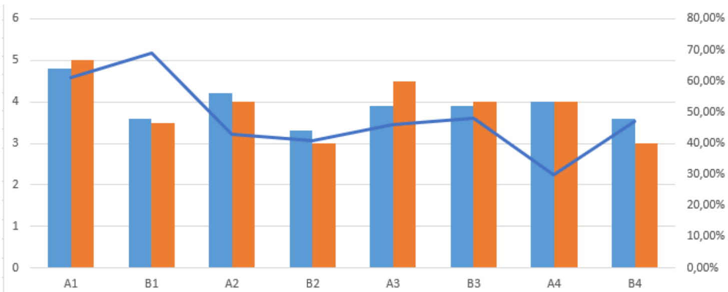


Figure 12 - Mean, median, and flow consistency per version

The versions of model B (model containing more gateways and real-life process model) were perceived in general as less easy to understand in comparison with model A (model designed for this experiment). In Figure 12 the mean, median, and *flow consistency* per model version are visualized.

4.3. Normality test

Parametric statistical tests are used in quantitative research for hypothesis testing. To be able to use parametric tests, the data cases should be normally distributed. This is one of the assumptions the data set needs to satisfy. (Saunders et al., 2019). Checking the normality of the data is then fundamental to proceed with a correct test to analyze the data.

The goal of this experiment was to reach at least 30 valid cases per group, following the Central Limit Theorem as explained in section 3.2, but the goal was not reached. As the valid cases gathered per group are 17, 13, 12, and 9 (all <30), testing the normality is important to determine applicable statistical tests for the analysis. To test the normality of the samples the Kolmogorov-Smirnov and Shapiro-Wilk tests can be used. The latter works better for small samples. (Saunders et al., 2019).

4.3.1. Model A

From the Shapiro-Wilk test (results in Table 10) it can be concluded that the data for the variable Task Effectiveness for Model version A3, the variable Task Efficiency for model versions A1 and A4; and the variable Perceived understandability for Model version A2 are not normally distributed (all $p<0.05$). To confirm the normality of the data from the other variables the histograms and Normal Q-Q plot charts were consulted and all observations were consolidated in Table 22. This table with observations, histograms, and Normal Q-Q plot charges can be consulted in Appendix XVIII.

After analyzing the histograms, Normal Q-Q plot, and the results from the Shapiro-Wilk test, it can be concluded that none of the three variables for model A (Task Effectiveness, Task Efficiency, and Perceived Understandability) are normally distributed, so a non-parametric test is needed for hypothesis testing. This changes the setup for the data analysis described in section 3.3. A MANOVA is considered to deliver robust findings (Saunders et al., 2019), but can only be performed if the distribution of the data is normal.

Table 10 - Normality test for Model A

		Kolmogorov-Smirnov			Shapiro-Wilk		
	Group Nr.	Statistic	df	Sig.	Statistic	df	Sig.
Task Effectiveness Model A	Group 1 (Model A1-B3)	,158	17	,200*	,920	17	,146
	Group 2 (Model A2-B4)	,134	13	,200*	,949	13	,590
	Group 3 (Model A3-B1)	,268	12	,017	,814	12	,014
	Group 4 (Model A4-B2)	,179	9	,200*	,947	9	,653
Task Efficiency Model B	Group 1 (Model A1-B3)	,207	17	,052	,872	17	,024
	Group 2 (Model A2-B4)	,163	13	,200*	,924	13	,283
	Group 3 (Model A3-B1)	,138	12	,200*	,966	12	,864
	Group 4 (Model A4-B2)	,273	9	,053	,759	9	,007
Undersandability_perception_Model_A	Group 1 (Model A1-B3)	,212	17	,041	,908	17	,093
	Group 2 (Model A2-B4)	,228	13	,063	,849	13	,027
	Group 3 (Model A3-B1)	,241	12	,053	,910	12	,213
	Group 4 (Model A4-B2)	,167	9	,200*	,963	9	,830

*. This is a lower bound of the true significance.
a. Lilliefors Significance Correction

4.3.2. Model B

From the Shapiro-Wilk statistical test for model B (see results in Table 11) it can be concluded that the data for the variable Task Effectiveness from Model versions B1 and B4; and the variable Task Efficiency from Model versions B2 and B3 are not normally distributed (all p<0.05).

Table 11 - Normality test for Model B

	Group Nr.	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Task Effectiveness Model B	Group 1 (Model A1-B3)	,156	17	,200*	,926	17	,185
	Group 2 (Model A2-B4)	,269	13	,011	,837	13	,019
	Group 3 (Model A3-B1)	,258	12	,027	,818	12	,015
	Group 4 (Model A4-B2)	,187	9	,200*	,925	9	,440
Task Efficiency Model B	Group 1 (Model A1-B3)	,207	17	,052	,872	17	,024
	Group 2 (Model A2-B4)	,163	13	,200*	,924	13	,283
	Group 3 (Model A3-B1)	,138	12	,200*	,966	12	,864
	Group 4 (Model A4-B2)	,273	9	,053	,759	9	,007
Undersandability_perception_Model_B	Group 1 (Model A1-B3)	,162	17	,200*	,919	17	,140
	Group 2 (Model A2-B4)	,235	13	,047	,917	13	,231
	Group 3 (Model A3-B1)	,160	12	,200*	,955	12	,716
	Group 4 (Model A4-B2)	,317	9	,010	,873	9	,132

*. This is a lower bound of the true significance.

a. Lilliefors Significance Correction

After analyzing the histograms and Normal Q-Q plots to confirm the results from the Shapiro-Wilk test (see Appendix XVIII), it can be concluded that from the three variables for model B, the Task Effectiveness and Task Efficiency variables are not normally distributed.

A non-parametric test is necessary to analyze the data from the variables Task Effectiveness and Task Efficiency. The variable Perceived Understandability is normally distributed for the data cases of Model B (all p>0.05), but to be able to compare the analysis from the variable Perceived Understandability for both models (A and B), it was decided to do the same non-parametric test.

4.4. Hypothesis testing

As the data cases are not normal distributed a non-parametric test is necessary. To test H1 and H2, it was decided to analyze if the understandability of a left-to-right layout is significantly different from a zig-zag layout or a banana layout with an independent samples Kruskal Wallis test. If the means are significantly different, it will be evidencing that a left-to-right layout's understandability is different (better or worse) than a less flow consistent layout (i.e. zig-zag or banana). To test H3 and H4, it will be tested if the *flow consistency* correlates with the understandability variables utilizing a Spearman's rank correlation (rho) test, if a significant correlation is evidenced, then will be proved that a relationship exists between the variables. (Saunders et al., 2019).

4.4.1. Hypothesis 1

To accept or reject the null hypothesis H1, the following steps will be followed. First, the difference between the means of the four groups for the variables Task Effectiveness and Task Efficiency will be tested using the Kruskal Wallis test with the data set from Model A. Next, the same test will be done with the data set from Model B. The output from the IBM SPSS Statistics tool can be consulted in Appendix XIX. As the last step, the results from both models (A and B) will be compared.

H1 Left-to-right process models are not significantly more understandable than a zig-zag or banana layout, i.e. (a) understandability Task Effectiveness, and (b) understandability Task Efficiency.

The results H(3)=1,270 p=0.736 (p>0.05) indicate that the means of Task effectiveness for model A are not significantly different across the four different versions for process model A (A1, A2, A3, and A4). Same conclusion is valid for the variable Task Efficiency H(3)=0,270 p=0.966 (p>0.05).

The same steps were repeated for Model B. The test result H(3)=2,467 p=0.481 (p>0.05) indicates that the means of Task effectiveness for model B are not significantly different across the four different versions for process model B (B1, B2, B3, and B4). Same conclusion is valid for the variable Task Efficiency H(3)=1,887 p=0.596 (p>0.05). These confirm the findings out of model A.

There is no significant difference evidenced between the four groups. Based on these results, the null hypothesis H1(a) and H1(b) needs to be accepted. A process model with a left-to-right layout is not more understandable than a process model with a zig-zag or banana layout.

4.4.2. Hypothesis 2

The same process is followed to confirm or reject the null hypothesis H2:

H2 – Left-to-right process models are not significantly perceived easier to understand than a zig-zag or banana layout, i.e. Perceived Ease of Understanding

The results H(3)=3,290 p=0.349 (p>0.05) indicate that the means of the Perceived Understandability for model A are not significantly different across the different versions for process model A.

To validate the findings from Model A, the same steps are followed for Model B. From the result H(3)=1,739 p=0.628 (p>0.05) it can be concluded that the means of the Perceived Understandability for model B are not significantly different across the different versions for process model B.

There is no significant difference evidenced between the means of the four groups. Based on these results, the null hypothesis H2 needs to be accepted. A process model with a left-to-right layout is not perceived easier to understand than a process model with a zig-zag or banana layout.

After this analysis, the fourth sub-question can be answered. A process model with less flow consistent layout (i.e. zig-zag or banana) is not significantly less understandable than a model with higher *flow consistency (i.e. left-to-right layout)*.

4.4.3. Hypothesis 3

To test if there is any relationship between the *flow consistency* of the process model and its understandability (i.e. Task Effectiveness and Task Efficiency), it is tested for significant correlations between the variables. Spearman's rank correlation (rho) is the best alternative to calculate the correlation between variables when the assumption of normality is not satisfied. It works with continuous data (F*low Consistency*), and it is less sensitive to strong outliers. (Saunders et al., 2019). As model A is different than model B it is decided to do the test for the variables of model A and model B separately. The Spearman's correlations matrix can be consulted in Appendix XVII Table 21.

H3 - The *flow consistency* of a process model has no relationship with the model's understandability, i.e. (a) understandability Task Effectiveness, and (b) understandability Task Efficiency.

Utilizing a Spearman's rank correlation test, it is evidenced that no significant correlation exists (all p>0,05) between the *Flow consistency* of model A and the Task Effectiveness of model A ($r_S(51)$ = 0,11, p=0,45), and also no significant correlation exists between the *Flow consistency* of model A and the Task Efficiency of model A ($r_S(51)$ = 0,01, p=0,94).

Also, no significant correlation is evidenced between the *Flow consistency* of model B and the Task Effectiveness of model B $r_S(51)$ = 0,19, p=0,17, and no significant correlation exists between the *Flow consistency* of model B and the Task Efficiency of model B $r_S(51)$ = 0,11, p=0,42. Confirming the findings from model A.

There is no significant relationship between the Flow consistency of a process model and its understandability objectively measured (i.e. Task Effectiveness and Task Efficiency).

4.4.4. Hypothesis 4

To test if there is any relationship between the *flow consistency* of the process model and its perceived understandability, it is tested for significant correlations between the variables with a Spearman's rank correlation (rho) test. As model A is different than model B it is decided to do the test for the variables of model A and model B separately. The Spearman's correlations matrix can be consulted in Appendix XVII Table 21.

H4 – The *flow consistency* of a process model has no relationship with the model's Perceived understandability, i.e. Perceived Ease of Understanding.

By means of a Spearman's rank correlation test, no significant correlation is evidenced between the *Flow consistency* of model A and the Perceived Understandability of model A $r_S(51) = 0{,}22$, p=0,12, and also no significant correlation is evidenced between the *Flow consistency* of model B and the Perceived Understandability of model B $r_S(51) = 0{,}10$, p=0,50.

There is no significant relationship between the Flow consistency of a process model and its perceived Ease of Understanding.

As H3 and H4 were not rejected (no relationship was evidenced), then H5 and H6 were not tested.

5. Discussion, conclusions, and recommendations

In the next sections the discussion of the results, reflection about the research method designed, conclusions, and recommendations are presented.

5.1. Discussion – reflection

Model version A1 and B1 could not fit in one screen size having a left-to-right flow direction without reducing the size of the model considerably. Some modelers might decide to change the layout of a process model to fit it to a format size. Leopold et al. (2015) assumed changing the layout increases the cognitive effort necessary to understand them. In this research, it was empirically researched the impact on understandability by changing the layout of a process model. The zig-zag#2 layout in the simple process model (i.e. A3 – designed for this experiment) and the left-to-right layout in the model having more gateways (i.e. B1 – real-life process model) showed a tendency towards better understandability (i.e. Task effectiveness and Task Efficiency). The left-to-right layout in the model A (i.e. A1 – designed for this experiment) and the zig-zag#2 in the model B (i.e. B3 –real-life process model) showed a tendency towards being perceived as easier to be understood. But the statistical tests evidenced that there is no significant relationship between the flow consistency of a process model and its understandability. Also, no significant difference in the process model's understandability was evidenced by changing a left-to-right layout to a zig-zag or banana layout.

Figl and Strembeck (2015) were not able to prove a significant superiority of a left-to-right flow direction in a process model, but their results showed a tendency for a left-to-right flow direction to perform better than other directions. They speculate that other flow directions didn't underperform in part by humans' ability to adapt quickly to uncommon reading directions (e.g., bottom-to-top). The same speculation is applicable in this research.

5.2. Limitations of this research

The main source (Open Universiteit library) provided a low amount of relevant literature. By using Google Scholar as a secondary source, an attempt to cover this limitation is done. It is proved a source will never provide access to all relevant articles.

The best effort was done to include the most possible known synonyms in a reiterative process. A possibility remains some synonyms were omitted to create the search queries.

The experiment's participation rate and completeness rate were lower than expected. It is recommended when doing an online setup to do a simple and short experiment/survey. 40,6% of the participants didn't finish this experiment, most of them stopping in the first sections. It is known ANOVA tests and their derivatives tests provide more robust insights. (Saunders et al., 2019). By having too few participants it is assumed the data set and the insights from the data might not be representative, it limits the number of tests to analyze the data, and real outliers are difficult to be identified.

It was requested to the participants to perform the experiment without distractions and hurries. It is assumed the participants followed the instructions. However, in this online setup, it is not possible to control the participant's involvement.

5.3. Conclusions

This research provides important contributions to the body of knowledge in different ways. First, a more complete and formal definition of *flow consistency* is provided. Second, an improved framework to measure the *flow consistency* of a process model has been developed, based on the framework from Burattin et al. (2018). Third, provides cutting-edge insights being the first (as far as it was able to be found) investigating the impact of changing a process model's layout, attempting to find a relationship between the *flow consistency* of a process model and its understandability.

From the Kruskal Wallis test, no significant difference is evidenced indicating process models with a zig-zag and banana layout are less understandable than models in a left-to-right layout (objectively and subjectively measured). From the Spearman's rank correlation analysis no significant relationship was proved between the process model's *flow consistency* and its understandability. It is concluded that changing the layout of a process model from a left-to-right layout to a zig-zag or banana layout (reducing the *flow consistency)*, has no significant impact on its understandability.

The superiority of a left-to-right layout has been theoretically proved by Figl and Strembeck (2014) but Figl and Strembeck (2015) and Kretschmann (2019) were not able to empirically confirm it. Leopold et al. (2015) assumed 'banana' and 'zig-zag' layouts increase the cognitive effort necessary to understand them, but banana and zig-zag layouts are not significantly less understandable.

5.4. Recommendations for practice

Even when no significant prove was found that reducing the *flow consistency* of a process model or changing the process model's layout impacts the model's understandability, a tendency for some process models to perform better or worse was visualized. This decision to change the model's layout to fit it to a certain format should be taken carefully.

The purpose of the process model should meet the 'fitness for use' (Lindland et al., 1994). A model needs to fulfill different quality requirements depending on who is the reader and the model's intention of use. (Dzepina & Lehner, 2018).

5.5. Recommendations for further research

This experiment can be repeated in the European region and other regions, e.g. Asia or within the Arab culture, where the normal read-text direction is top-to-bottom or right-to-left. This should be done using more participants to reach a normal distribution of the data set, and increase the generalizability and robustness of the findings.

Task efficiency measures understandability in a different way than task Effectiveness. This concept should be validated in terms of understandability. Task efficiency might give a high value to participants scoring average in task effectiveness (correctness), but by being less interested in the experiment answered faster. Participants more engaged willingly to answer correctly might do longer to assure their answer is correct, making their result in task effectiveness less high. The question is, which of the two participants understood better the model? Also, the questions from Reijers et al. (2011) for the experiment need to be validated as not all questions of 'Concurrency' tasks performed the same good in the same model version. See Appendix XX.

It could also be interesting to repeat this experiment by presenting multiple times the same questions with the same model layout in the same setting. This could also prove some differences between understandability at 'first sight' and understandability with often use, assuming process models are mostly designed to be used often by the same reader.

References

Becker, J., Rosemann, M., & Uthmann, C. (2003). Guidelines of Business Process Modeling. *1806*. doi:10.1007/3-540-45594-9_3

Becker, J., Rosemann, M., & Von Uthmann, C. (2000). Guidelines of business process modeling. In *Business process management* (pp. 30-49): Springer.

Bernstein, V., & Soffer, P. (2015). *Identifying and Quantifying Visual Layout Features of Business Process Models*.

Burattin, A., Bernstein, V., Neurauter, M., Soffer, P., & Weber, B. (2018). Detection and quantification of flow consistency in business process models. *Software & Systems Modeling, 17*(2), 633-654. doi:10.1007/s10270-017-0576-y

Corradini, F., Ferrari, A., Fornari, F., Gnesi, S., Polini, A., Re, B., & Spagnolo, G. O. (2018). A guidelines framework for understandable BPMN models. *Data & Knowledge Engineering, 113*, 129-154.

Dikici, A., Turetken, O., & Demirors, O. (2017). Factors Influencing the Understandability of Process Models: A Systematic Literature Review. *Information and Software Technology*. doi:10.1016/j.infsof.2017.09.001

Dzepina, A., & Lehner, F. (2018). BUSINESS PROCESS MODEL QUALITY–RESULTS FROM A STRUCTURED LITERATURE REVIEW.

Effinger, P. (2011). *Layout Patterns with BPMN Semantics.* Paper presented at the International Workshop on Business Process Modeling Notation.

Effinger, P., Jogsch, N., & Seiz, S. (2010). *On a Study of Layout Aesthetics for Business Process Models Using BPMN* (Vol. 67).

Effinger, P., Siebenhaller, M., & Kaufmann, M. (2009). *An interactive layout tool for bpmn.* Paper presented at the 2009 IEEE Conference on Commerce and Enterprise Computing.

Eichelberger, H., & Schmid, K. (2009). Guidelines on the aesthetic quality of UML class diagrams. *Information and Software Technology, 51*(12), 1686-1698.

Field, A., & Hole, G. (2003). Descriptive statistics. *How to Design and Report Experiments. Sage Publications*, 109-140.

Figl, K. (2017). Comprehension of Procedural Visual Business Process Models. *Business & Information Systems Engineering, 59*(1), 41-67. doi:10.1007/s12599-016-0460-2

Figl, K., & Laue, R. (2015). Influence factors for local comprehensibility of process models. *International Journal of Human-Computer Studies, 82*, 96-110.

Figl, K., & Strembeck, M. (2014). *On the Importance of Flow Direction in Business Process Models*.

Figl, K., & Strembeck, M. (2015). *Findings from an Experiment on Flow Direction of Business Process Models.* Paper presented at the EMISA.

Handel, S., DeSoto, C. B., & London, M. (1968). Reasoning and spatial representations. *Journal of Verbal Learning and Verbal Behavior, 7*(2), 351-357.

Harsel, Y., & Wales, R. (1987). Direciional Preference in Problem Solving. *International Journal of Psychology, 22*(2), 195-206.

Havey, M. (2005). *Essential Business PRocess Modeling*: O'Reilly MEdia, Inc.

Kesh, S. (1995). Evaluating the quality of entity relationship models. *Information and Software Technology, 37*(12), 681-689. doi:https://doi.org/10.1016/0950-5849(96)81745-9

Kitzmann, I., König, C., Lübke, D., & Singer, L. (2009). *A simple algorithm for automatic layout of bpmn processes.* Paper presented at the 2009 IEEE Conference on Commerce and Enterprise Computing.

Kretschmann, K. (2019). *Investigating the Flow Direction in Business Process Models: An Eye Tracking Study.* Ulm University,

Krogstie, J., Lindland, O. I., & Sindre, G. (1995, 1995//). *Towards a deeper understanding of quality in requirements engineering.* Paper presented at the Advanced Information Systems Engineering, Berlin, Heidelberg.

Krohn, G. S. (1983). Flowcharts used for procedural instructions. *Human Factors, 25*(5), 573-581.

Leopold, H., Mendling, J., & Günther, O. (2015). What we can learn from Quality Issues of BPMN Models from Industry. *IEEE Software, 33*. doi:10.1109/MS.2015.81
Lindland, O. I., Sindre, G., & Solvberg, A. (1994). Understanding quality in conceptual modeling IEEE Software.
Mendling, J. (2009). Empirical Studies in Process Model Verification. *Trans. Petri Nets Other Model. Concurr., 2*, 208-224.
Mendling, J., Reijers, H. A., & Aalst, W. (2010). Seven Process Modeling Guidelines (7PMG). *Information and Software Technology, 52*, 127-136. doi:10.1016/j.infsof.2009.08.004
Mendling, J., & Strembeck, M. (2008). *Influence factors of understanding business process models.* Paper presented at the International Conference on Business Information Systems.
Mendling, J., Strembeck, M., & Recker, J. (2012). Factors of process model comprehension—Findings from a series of experiments. *Decision Support Systems, 53*(1), 195-206.
Mendling, J., & van der Aalst, W. M. (2006). *Towards EPC Semantics based on State and Context.* Paper presented at the EPK.
Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. (2009). PRISMA 2009 Flow Diagram. The PRISMA statement. *PLoS Med, 6*(7), e1000097.
Moreno-Montes de Oca, I., & Snoeck, M. (2014). Pragmatic guidelines for business process modeling. *Available at SSRN 2592983*.
Nordbotten, J. C., & Crosby, M. E. (1999). The effect of graphic style on data model interpretation. *Information Systems Journal, 9*(2), 139-155.
Object Management Group, I. O. (2011). Business Process Model and Notation (BPMN). In: Object Management Group, Inc. (OMG).
Petrusel, R., Mendling, J., & Reijers, H. (2016). Task-Specific Visual Cues for Improving Process Model Understanding. *Information and Software Technology, 79*. doi:10.1016/j.infsof.2016.07.003
Petrusel, R., Mendling, J., & Reijers, H. A. (2017). How visual cognition influences process model comprehension. *Decision Support Systems, 96*, 1-16.
Reijers, H., Freytag, T., Mendling, J., & Eckleder, A. (2011). Syntax highlighting in business process models. *Decision Support Systems, 51*, 339-349. doi:10.1016/j.dss.2010.12.013
Saunders, M., Lewis, P., & Thronhill, A. (2019). *Research Methods for Business Students* (Eighth ed.): Pearson Education Limited.
Schrepfer, M. (2010). Modeling guidelines for business process models. *Humboldt-Universitat zu Berlin*.
Schrepfer, M., Wolf, J., Mendling, J., & Reijers, H. A. (2009). *The impact of secondary notation on process model understanding.* Paper presented at the IFIP Working Conference on The Practice of Enterprise Modeling.
Turetken, O., Dikici, A., Vanderfeesten, I., Rompen, T., & Demirors, O. (2020). The Influence of Using Collapsed Sub-processes and Groups on the Understandability of Business Process Models. *Business & Information Systems Engineering, 62*(2), 121-141. doi:10.1007/s12599-019-00577-4
van Veen, M., & Westerkamp, K. (2010). *Deskresearch: Informatie selecteren, beoordelen en verwerken*.
Wang, N., Sun, S., & OuYang, D. (2018). Business process modeling abstraction based on semi-supervised clustering analysis. *Business & Information Systems Engineering, 60*(6), 525-542.
Weidlich, M., Mendling, J., & Weske, M. (2010). Efficient consistency measurement based on behavioral profiles of process models. *IEEE Transactions on Software Engineering, 37*(3), 410-429.
Winn, W. (1982). The role of diagrammatic representation in learning sequences, identification and classification as a function of verbal and spatial ability. *Journal of Research in Science Teaching, 19*(1), 79-89.
Winn, W. (1983). Perceptual strategies used with flow diagrams having normal and unanticipated formats. *Perceptual and motor skills, 57*(3), 751-762.

Appendix I – Literature research approach in detail

Approach

A systematic literature review is conducted according to Saunders et al. (2019). First, the sources to search the literature are identified. Second, the search parameters are defined. Third, the search terms are generated. Fourth, the evaluation criteria are defined to select the most relevant literature for this research, this includes defining the inclusion and exclusion criteria.

A systematic evaluation process is followed to guarantee only relevant articles are taken into consideration for the literature review according to the Prisma 2009 Flow Diagram from Moher et al. (2009) consisting of the phases: identification, screening, and eligibility.

First, the identification phase was done. The articles were searched in the defined sources with the defined search parameters using the search terms resulting from the *building blocks* method, explained in the Search terms section.

Second, the screening phase followed. The articles that resulted from these searches were defined as relevant if the title was compliant with the inclusion and exclusion criteria explained in section 2.1.2. The title, publication year, and authors of the resulted articles were inserted in the 'Search Hits and Relevance table' (see Appendix II - Search hits and Relevance Table). The inclusion and exclusion criteria for each article were evaluated. If the result for all inclusion criteria was positive, was negative for all exclusion criteria, and the full text was available then the article was classified as relevant at this phase. If the title was not clear enough to classify the article as relevant or irrelevant, then the abstract was read and the relevancy of the article for this research was clear.

Third, the eligibility process needed to be done. The abstract, introduction, and conclusion of the articles remaining as relevant from the screening phase were read as besides to only mention the search terms the article should be about: the description or definition of the flow consistency in process models, measurement frameworks for flow consistency of a process model, understandability of a process model or a synonym of all these terms. This allowed the author of this work to identify the relevant articles to be included in the literature review.

As the last step to retrieve the articles that formed the literature framework for this research the 'backwards snowballing' method was followed using the references of the articles that resulted as eligible in the previous steps including the other articles identified through other sources including the literature reviews read initially. The resulting articles from the *backwards snowballing* method were processed as well through the screening and eligibility phases previously described. Until no new relevant articles were identified.

Sources & Search Parameters

The relevant scientific literature has been obtained by using two electronic databases as a source. The main source is the online "Open University" library. After the first test-searches, because of the low amount of articles resulted from this source, it was decided in a later stage (more detail in section 2.2 Implementation) to use a second source as a double-check so as not to miss relevant literature on the subject, being this Google scholar. This resulted to be a good decision as several relevant articles were found. The search parameters used are visualized in Table 12.

Table 12 - Search Parameters

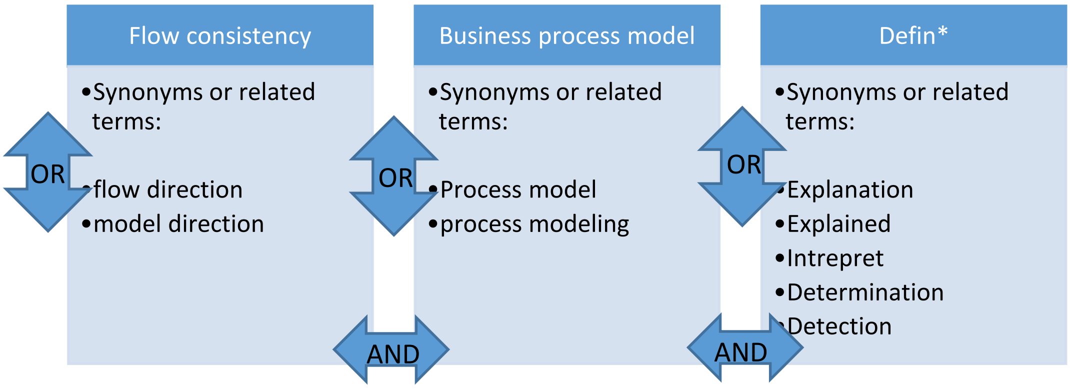
Language of publication:	**English**
Subject area*	(Business) process models
Business sector* /discipline	Applied sciences, Business, computer science, education, engineering, mathematics, psychology, sciences, social sciences, statistics
Geographical area	All - to be able to get as much as possible literature searched
Publication period	All - to be able to get as much as possible literature searched
Literature type*	Peer-reviewed, academic journals, conference papers.
(*) these parameters were not able to be used in Google Scholar	

Search terms

To define the search terms the method: 'building blocks' has been used. (van Veen & Westerkamp, 2010). First, to be able to identify the search terms to be used in the 'building blocks' method, two recent literature reviews were read (Dikici et al., 2017; Figl, 2017). These literature reviews provided a general overview of what the literature *says* regarding comprehension and understandability of business process models and the factors that might have an impact on the understandability of these models. Hereafter the search terms per sub-question were defined after a brainstorming session. The main search queries are visualized in Table 1. The building blocks search terms are presented as follows. A more detailed overview of the hits resulted per search team and their relevancy can be visualized in Appendix II - Search hits and Relevance Table.

Search terms building blocks for the first sub-question:

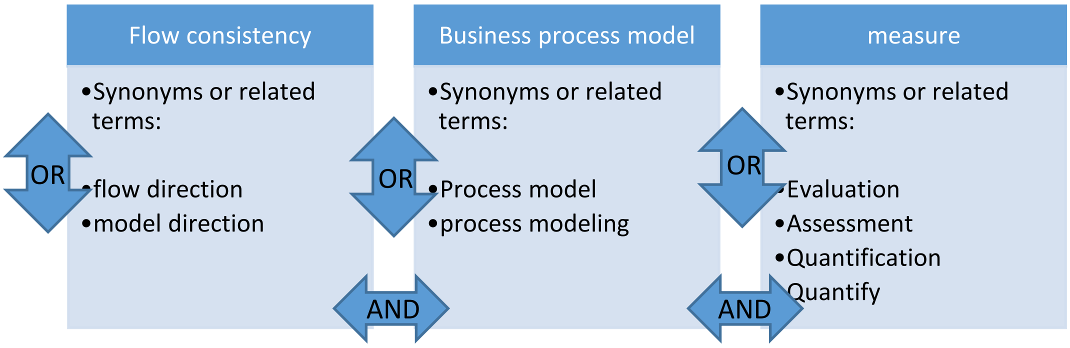
How can the flow consistency of a business process model be defined?



Search term Nr 1: (("flow consistency" OR "flow direction" OR "model direction") AND ("process model*") AND (defin* OR expla* OR interpret* OR determin* OR detect*))

Search terms building blocks for the second sub-question:

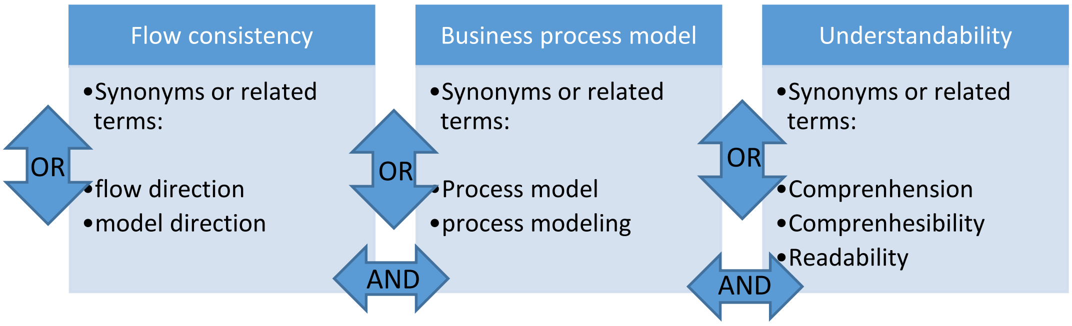
How can the flow consistency of a business process model be measured?



Search term Nr 2: ("flow consistency" OR "flow direction" OR "model direction") AND ("process model*") AND (measure* OR evaluation OR assessment OR quantif*))

Search terms building blocks for the third sub-question:

What can be found in the literature about the flow consistency of a business process model and its impact on the understandability of the model?



Search term Nr 3: (("flow consistency" OR "flow direction" OR "model direction") AND ("process model*") AND (understand* OR comprehen* OR read*))

Evaluation criteria

The literature articles retrieved utilizing the previous parameters and terms were evaluated to determine their relevance to this research. For this inclusion and exclusion criteria were defined as mentioned in Table 13.

Table 13 - Inclusion and Exclusion criteria for literature evaluation

Inclusion criteria:	**Exclusion criteria:**
Article text should be available in English	The article should not be related to liquid flow, physics, chemistry, water flow, river flow, material flow.
The article should be compliant with the search parameters	
The search terms should be mentioned in the title or the abstract of the article	
The title or abstract of the article should relate to (business) process models.	

Implementation

The main search in the main source resulted in nine articles (six duplicates). After the evaluation phase using the inclusion and exclusion criteria, two articles resulted to be relevant. For the eligibility phase, the introduction, abstract, and conclusions were read and both articles were considered relevant for this research. The articles are listed in Appendix III.

Because of the low number of articles found, it was decided to use a second source, being this Google Scholar, as this is available within the author's resource limitations and provides wide searches. After using this source several limitations have been identified. These limitations and the actions taken to pass the limitations are described in Appendix IV.

A second search was made. Because of the limitations of the second source the search terms were adapted resulting in 81 articles. To guarantee the completeness of the articles found the adapted search terms were repeated in the main source, resulting in eight articles. To assure no relevant articles were excluded, the Systematic Literature Reviews (SLR) about process models were searched in both sources, resulting in 38 unique SLRs. This second search delivered a total of 127 articles. Three articles and nine SLRs passed the evaluation phase. The article from Kretschmann (2019) didn't comply with all inclusion criteria as it is not peer-reviewed but this bachelor thesis was kept in the final literature list as it was considered important complementary research for the article of Figl and Strembeck (2015), researching the same phenomena, concluding the same but both articles using a different research method. Five publications passed the eligibility phase.

The full text of the relevant publications was read to retrieve possible relevant literature excluded until now. By applying the *backwards snowballing* method, 158 new articles were retrieved. Two articles passed the evaluation phase and the eligibility phase. An overview of the final list of articles resulted from the literature framework can be consulted in Appendix V.

Appendix II - Search hits and Relevance Table

To be able to do the screening of the articles resulting from the initial searches a table in excel (see below Table 14) was kept containing: the title of the article, source, search terms used, the publication year of the article, authors, results for inclusion criteria, results for exclusion criteria and relevance of the article.

Table 14 - Search hits and relevance table

Search iteration	Search term	Source	Description	Search term	Sub-question	Hits	Duplicates	Unique titles	Relevant for Eligibility phase	Duplicates	Total relevant articles for eligibility phase
1	1	Open University Library	Search terms in the title	((TitleCombined:("flow consistency")) OR (TitleCombined:("flow direction")) OR (TitleCombined:("model direction"))) AND (TitleCombined:("process model*")) AND (TitleCombined:(defin* OR expla* OR interpret* OR determin* OR detect*))	1	1	0	1	1	0	1
1	2	Open University Library	Search terms in title	((TitleCombined:("flow consistency")) OR (TitleCombined:("flow direction")) OR (TitleCombined:("model direction"))) AND (TitleCombined:("process model*")) AND (TitleCombined:(measure* OR evaluation OR assessment OR quantif*))	2	1	1	0	1	1	0
1	3	Open University Library	Search terms in title	((TitleCombined:("flow consistency")) OR (TitleCombined:("flow direction")) OR (TitleCombined:("model direction"))) AND (TitleCombined:("process model*")) AND (TitleCombined:(understand* OR comprehen* OR read*))	3	0	0	0	0	0	0
1	4	Open University Library	Search terms in abstract	(Abstract:("flow consistency" OR "flow direction" OR "model direction")) AND (Abstract:("process model*")) AND (Abstract:(defin* OR expla* OR interpret* OR determin* OR detect*))	1	2	1	1	1	1	0
1	5	Open University Library	Search terms in abstract	(Abstract:("flow consistency" OR "flow direction" OR "model direction")) AND (Abstract:("process model*")) AND (Abstract:(measure* OR evaluation OR assessment OR quantif*))	2	2	1	1	2	1	1

Search iteration	Search term	Source	Description	Search term	Sub-question	Hits	Duplicates	Unique titles	Relevant for Eligibility phase	Duplicates	Total relevant articles for eligibility phase
1	6	Open University Library	Search terms in abstract	(Abstract:("flow consistency" OR "flow direction" OR "model direction")) AND (Abstract:("process model*")) AND (Abstract:(understand* OR comprehen* OR read*))	3	3	3	0	2	2	0
2	7	Google Scholar	Search terms in title	allintitle: "flow consistency" process OR model	1, 2 & 3	2	1	1	1	1	0
2	8	Google Scholar	Search terms in title	allintitle: "flow direction" process OR model	1, 2 & 3	56	0	56	3	0	3
2	9	Google Scholar	Search terms in title	allintitle: "model direction" process OR model	1, 2 & 3	23	0	23	0	0	0
2	10	Open University Library	Search terms in title	((TitleCombined:(process model flow consistency)) OR (TitleCombined:(process model flow direction)) OR (TitleCombined:(process model model direction)))	1, 2 & 3	8	1	7	1	1	0
3	11	Open University Library	Search terms in title	(TitleCombined:(process model literature review))	1, 2 & 3	17	2	15	4	0	4
3	12	Google Scholar	Search terms in title	allintitle: process model literature review	1, 2 & 3	23	0	23	5	0	5
				Backward snowball method	1, 2 & 3	158	0	158	4	0	4
				Totals		**288**	**10**	**278**	**25**	**7**	**18**

Appendix III – Result of the first literature search

After doing the main search in the main source (the online OU library) the result in total was nine hits, six of them were duplicated remaining three unique articles. From these three articles, two resulted as relevant after the evaluation, inclusion, and exclusion criteria. These articles are listed in Table 15.

Table 15 - First search iteration hits

Sub-question	Article title	Year	Authors	Relevant?
1 & 2	Detection and quantification of flow consistency in business process models	2018	Burattin, A.; Bernstein, V.; Neurauter, M.; Soffer, P. & Weber, B.	Yes
2	Business Process Modeling Abstraction Based on Semi-Supervised Clustering Analysis	2018	Wang, N.;Sun, S. & OuYang, D.	Yes
-	Glacial dispersal and flow history, East Arm area of Great Slave Lake, NWT, Canada	2017	Sharpe, D.R; Kjarsgaard, B.A; Knight, R.D; Russell, H.A.J. & Kerr, D.E.	No

Appendix IV - Limitations and actions taken –search on Google Scholar

Citations and patents were excluded from the search. After using this source several limitations have been identified.

The first limitation, the high number of hits after each search. To manage this limitation and to increase the relevance of the resulted articles, only the articles in the first pages were evaluated when sorted by relevance, until no new relevant article was found in the next ten consecutive pages.

The second limitation, not possible to search for words only in the abstract. To pass this limitation, it was decided to search for the search terms in the title only utilizing the advanced search functionality.

Third limitation, not possible to search using previously defined search parameters. This limitation was managed manually, by first reading the title, identifying if the article was related to the subject area (business) process models, second, if the title related to (business) process models or it was unclear the abstract was read, identifying in this step if the article was available, was available in English, and if the article was relevant or not. As the last step, the articles identified as relevant from the previous step, it was searched if the article complies with the literature type from the search parameters to only include in the final literature long list peer-reviewed, academic journals or conference papers.

Fourth limitation, not possible to use correctly the OR Boolean search function when searching only in the title. Because of this, new search terms were created where ‘process’ OR ‘model’ (from the second block) was combined with only one of the terms from the first block, each synonym by apart, resulting in several extra searches. As the terms from the sub-question specific, from the third block, were excluded the same search term was applicable for all three sub-questions.

Appendix V – Final list of articles resulted from the literature search

The final list of articles resulted to be relevant to this research is listed in Table 16.

Table 16 - Final list of articles included in the literature review

Sub-question	Article title	Year	Authors
1 & 2	**Detection and quantification of flow consistency in business process models**	2018	Burattin, A.; Bernstein, V.; Neurauter, M.; Soffer, P. & Weber, B.
2	**Business Process Modeling Abstraction Based on Semi-Supervised Clustering Analysis**	2018	Wang, N.; Sun, S. & OuYang, D.
1 & 3	**On the importance of flow direction in business process models**	2014	Figl, K & Strembeck, M.
1 & 3	**Findings from an experiment on flow direction of business process models**	2015	Figl, K & Strembeck, M.
1 & 3	**Investigating the Flow Direction in Business Process Models: An Eye Tracking Study**	2019	Kretschmann, K. [1]
2	**Comprehension of Procedural Visual Business Process Models: A Literature Review**	2017	Figl, K.
1	**Factors influencing the understandability of process models: A systematic literature review**	2017	Dikici, A.; Turetken, O. & Demirors, O.
1 & 2	**Identifying and Quantifying Visual Layout Features of Business Process Models**	2015	Bernstein, V.; Soffer, P.
1	**Influence factors for local comprehensibility of process models**	2015	Figl, K. & Laue, R.

[1] This article is a bachelor thesis which did not comply with the criteria peer-reviewed, but this was taken in the final list as: it is about the same subject investigated, can be considered as a complement for the article of Figl & Strembeck (2015) resulting in the same findings but using another research method.

Appendix VI – Validating the definition of flow consistency

Flow consistency is defined in the literature as: "*the consistency of flow measures the extent to which the layout of a process model reflects the temporal logical ordering of the process*" (Burattin et al., 2018, p. 639).

To understand this definition and validate it, it will be needed first to understand what is the flow of a process according to the literature, which will later be evaluated by its consistency, and then define what should be understood by *flow consistency*.

A business process is a set of activities that are executed in a predefined order to reach a business goal. Thus the order in which these activities are developed is of extreme importance. When the activities are laid in the canvas in the same order these are performed, this inherent ordering of elements gives the process model a *visual flow*. (Figl & Strembeck, 2015). This is referred to in the definition of flow consistency from Burattin et al. (2018) as the temporal logical ordering of the process [activities]. The sequence flow shows the order of how activities are performed in the process and this can be used as a hint to consider the reading direction of the model. (Effinger et al., 2010; Moreno-Montes de Oca & Snoeck, 2014; Object Management Group, 2011; Winn, 1982).

In BPMN 2.0 as in many other process modeling languages, the sequence flow is drawn as an arrow symbol. An arrow is used to show the logical sequence of how the activities have to be performed in a process. This temporal constraint and the logical constraints defined by control-flow objects (i.e. gateways) will define the path to follow by the token and the reader. (Object Management Group, 2011). Figl and Strembeck (2015) refer to the model direction as the control-flow logic describing the logical and temporal order in which tasks are performed. Process modeling notations use node-link diagrams, a specific type of directed graphs to depict the process flow, visualizing the execution order of tasks in a business process. "*The position of the start and the end nodes, as well as the arrowhead of the edges, show the precedence relations between the model elements*" (Figl & Strembeck, 2015, p. 59).

Correctly understanding the control-flow between the different tasks in a process model is a challenge for the comprehension of process models, due to their temporal and logical constraints. The control-flow objects are key to describe and understand correctly process models and it differentiates them from other modeling types, like object structures or data relationships. (Petrusel, Mendling, & Reijers, 2017). Figl and Laue (2015) were able to prove models with complex control-flow gateways are more difficult to understand than models with simple sequence flows. But the authors didn't take into consideration the consistency of flow from the models used.

There is a direct relationship between the location of the activities and the direction of the flow of a model as the direction of the arrows will be determined by the location of the activities in the canvas. For example, in Figure 13, the direction flow of a pair of activities A and B are visualized, considering activity A needs to be finished before activity B can start having a direct relationship, if activity A is drawn at the right-hand side of activity B then the arrow will start at the right side of the canvas (where activity A is located) and point to the left side of the canvas were activity B is located. Then the *flow direction* will be right-to-left. The behavioral profiles theory by Weidlich, Mendling, and Weske (2010) captures the relations between pairs of activity nodes based on the notion of weak order, this latter is present when a trace exists in which one activity of a process model occurs after another activity for that pair. The behavioral profiles can be classified as strict order relation, exclusiveness relation, or interleaving order relation. (Weidlich et al., 2010).

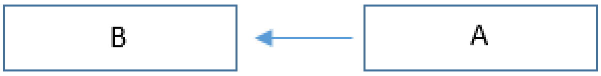


Figure 13- Flow direction of the activities pair A and B

Even when the flow defines the model's direction, Burattin et al. (2018) don't take the direction explicitly into consideration on their definition of flow consistency. The way the model's objects (i.e. activities, start/end event; control-flow objects) are arranged or laid out in the canvas (layout) determines the visual flow **direction** of the process model. The consistency of the model's flow **direction** will be the extent to which the position and temporal logical ordering of the activities follow **the same direction**. Therefore the definition from Burattin et al. (2018) has been completed.

A new definition is proposed: **Flow direction consistency (in short Flow consistency) is *the extent to which the position and temporal logical ordering of the activities of a process model follow the same direction.***

A valid remark is that even when the location of the model's object in the canvas is impacting directly the layout of the model as a whole, there were not many guidelines found regarding where to locate the objects of a model (see Appendix IX - Discussions and uncertainties regarding the flow direction for more detail). The modeler has the freedom to choose, even when some guidelines like the standard BPMN2.0 advise to use a model direction from left-to-right or top-to-bottom, there is no advice on where to put the start event, the activities, or end event. When referring to the location some conditions are defined as 'placed inside the shape', 'in any direction or location' leaving this to the preference of the modeler or even to the modeling tool vendor, which in general will define the general model direction. (Object Management Group, 2011).

Appendix VII – Extract from Burratin et al. 2018 - Fig 2

Figure 14 shows an example of models used by Burattin et al. (2018) with different layouts with the same process description. A Process model with a consistent direction of the flow. b. A model with some violation of the flow consistency. c. Model without a strong flow consistency.

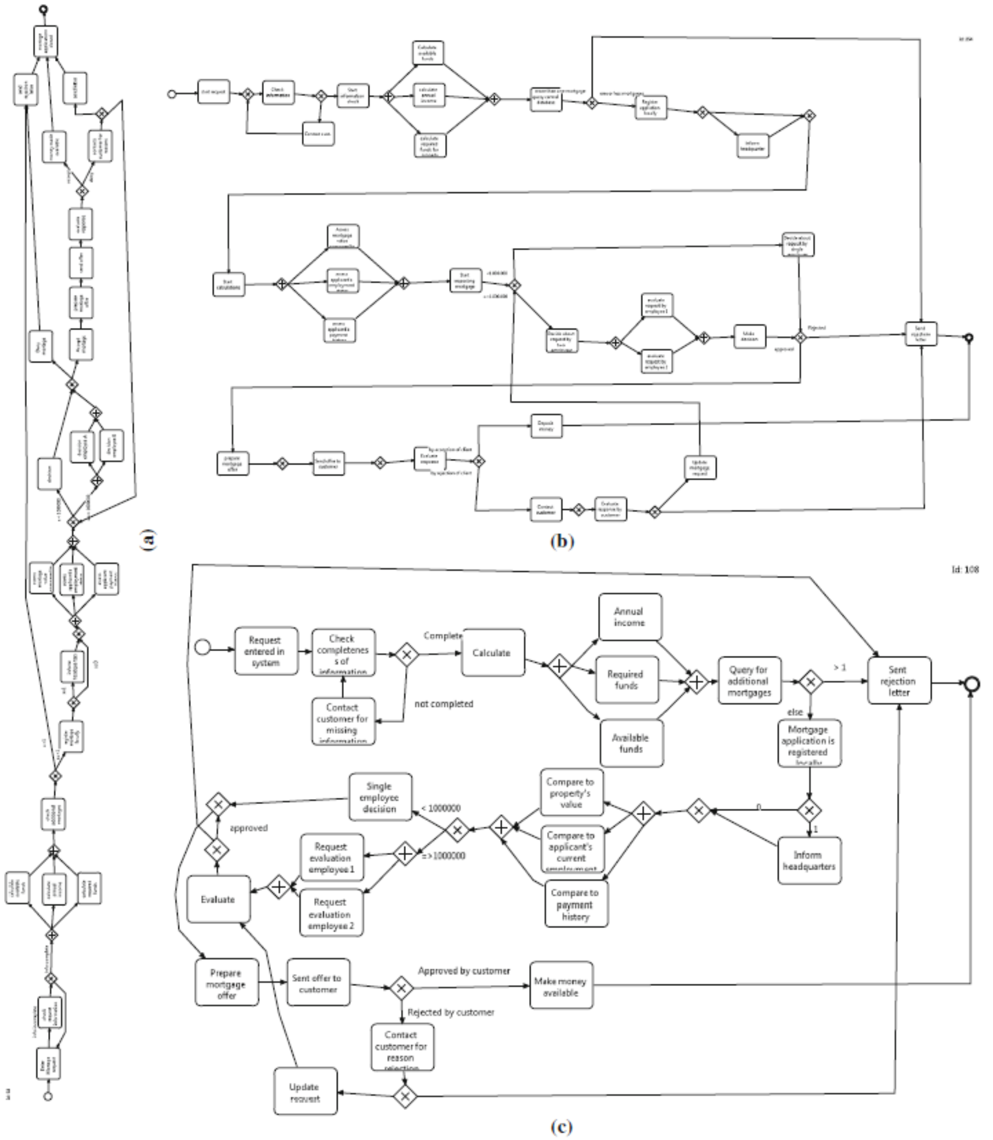


Fig. 2 Examples of models with different layouts, obtained starting from the same process description. **a** Process model with a consistent direction of the flow. **b** Model with some violations of the flow consistency. **c** Model without a strong flow consistency

Figure 14 - extract from Burratin et al. 2018 - Fig 2

Flow consistency using the three metrics proposed are visualized in Table 17.

Table 17 - results of the three metrics in example models

	Model A	Model B	Model C
ME-1	East, FC= 0.941	East, FC= 0.847	East, FC = 0.392
ME-2	East, FC= 0.960	East, FC= 0.915	South, FC = 0.588
M-BP	South or East, FC = 0.930	South or East, FC = 0.868	South or East, FC = 0.622

Appendix VIII – Framework from Burattin et al. 2018

As from here on it will be referred to the metrics of Burattin et al. (2018) will be explained and will be referred to as the 'framework'. Automatic identification of changes in the flow direction can be based on global or local features. Global features consider the general direction of the model to detect the flow consistency based on the general shape of the process. Local features will consider how activities and sequences (edges) are located concerning each other. The 'framework' consist of three metrics operationalized from two different approaches and indicates to what extent these metrics are consistent with human perception. The first two metrics calculate the extent to which the edges of the model are consistent, by calculating the direction of each edge and comparing it against the most frequent direction determined based on majority voting. The third metric determines whether the position of each pair of activities reflects their temporal local ordering.

During the design and evaluation of this framework only consistent flow directions were considered (i.e. left-to-right, right-to-left, top-to-bottom, and bottom-to-top), not considering the possibility of change in the direction of the model.

For the calculation of the *flow consistency* the shape of the line does not influence as their framework considers their starting and ending coordinates, i.e. in Figure 15, 3 edges have the same start and end vectors.

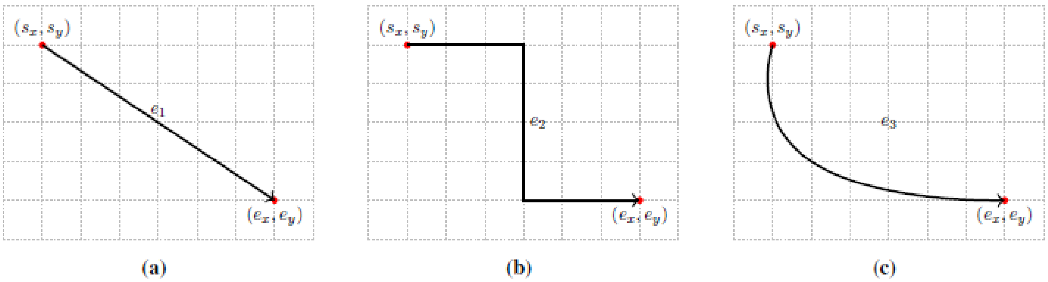


Figure 15 - three differently shaped edges are equally represented in Burattin et al. 2018 framework having the same pair of points (sx, sy), (ex, ey)

The first metric is defined as M-E1 considers one direction per edge. To identify the direction of each edge, the angle of it is considered. To determine the direction the radius has been divided into four equal parts of 90°, one for each direction i.e. North, East, South West. The direction of each edge in the model is determined depending on in which section the angle of the edge is included. See in Figure 16 the four directions: The filled area identifies the North direction; the dotted area identifies East; the grid area represents South; and the lined area identifies the West direction. The total of edges per direction is then divided by the total of edges in the model. The greatest determinate the flow direction and flow consistency of the model.

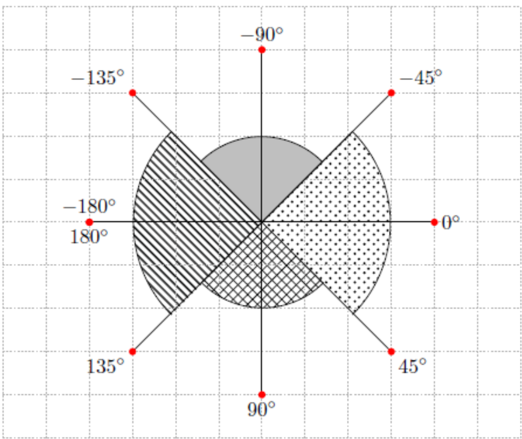


Figure 16 - the division of the radius in four sections for metric M-E1. North-gray filled area, East-dotted area, South-grid area, and West (lined area)

The second metric of this framework called M-E2 considers two directions per edge. The main difference with the previous metric is that each direction gets 180° of the radius. So each edge will get two directions assigned instead of one. The division of the radios is visualized in Figure 17. It is not clear from the article what are the ranges of these sections and if these ranges are included or excluded, i.e. an edge with an angle of 0° will get East direction, but this is the border for North and South direction, so it is not clear whether 0° should be considered as North or South direction, or maybe none of them. This should be considered for improvement in this metric. The total of edges per direction is then divided by the total of edges in the model. The greatest determinate the flow direction and flow consistency of the model.

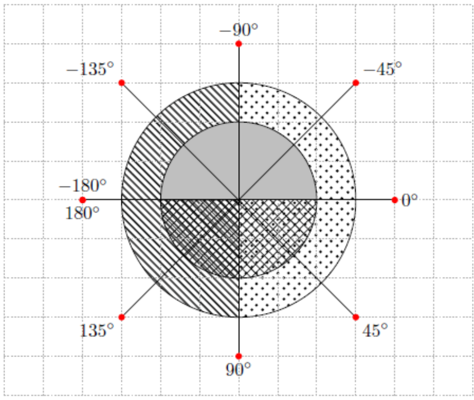


Figure 17 - Division of the radius in four sections for metric M-E2. North-gray filled area, East-dotted area, South-grid area, and West (lined area

The third metric of this framework defined as M-BP is based on behavioral profiles. While the first two metrics use the direction of the edges, this third metric, M-BP, looks at the relationship between pairs of activities. The concept of behavioral profiles comes from the work of Weidlich et al. (2010). And evaluates for each pair of activities the way they are placed concerning each other is consistent with their temporal logical ordering. The complexity of the algorithm proposed for this metric is linear to the number of behavioral relations of the process.

To evaluate the accuracy of the metrics from this framework Burattin et al. (2018), conducted 3 empirical evaluations. The first evaluation established to what extent the metrics measure in the same way the flow consistency of the models. The 3 metrics agree on the high flow consistent and low flow consistent models, but for average flow consistent models, the 3 metrics differed the higher from each other. The authors conclude that the features in each metric are not redundant to each other. The second evaluation focused only on the time efficiency to compute each metric. The metric M-BP is the least efficient of the 3 by a big difference, but it is s still reasonable fast to be able to be applied. The third evaluation aimed to compare the metrics with the human perception of the flow consistency. The participants were familiar with process modeling and rated 14 models with different kinds of consistency of flows using a 7-point Likert scale. The metric M-BP compared to the average of the human assessment results had a Pearson correlation of 0.719 and a significance value of 0.004. The results are linearly shifted by a factor on average of 0.029. The metric M-E2 with a Pearson correlation of 0.567 and a significance of 0.034 indicates that there is also a correlation but less strong than M-BP. The last metric M-E1 is no significantly correlated to human perception (Pearson correlation 0.263 and Significance 0.364). The authors then conclude that two out of three proposed metrics correlate with human perception. And the metric M-BP has the best correlation with the human perception of the flow consistency of the models. The human perception seems to give more importance to the position of the activities to define the flow consistency of a model rather than the direction of the edges. A limitation of this comparison is that human perception has been evaluated subjectively and intuitively rather than giving objective scores to the models. Flow consistency human perception is not to be understood as comprehension of the model.

Appendix IX - Discussions and uncertainties regarding the flow direction

Figl and Strembeck (2015) measured objectively and subjectively via questionnaires the impact of the directions: left-to-right, right-to-left, bottom-to-top, and top-to-bottom; on the understandability of the model. In contrast, Kretschmann (2019) measured the same impact utilizing an eye-tracking experiment. Petrusel, Mendling, and Reijers (2016) were able to identify a difference in mental effort related to processing information via eye-tracking while the traditional comprehension task did not detect a difference between the same participants. This was not the case with the studies of Figl and Strembeck (2015) and Kretschmann (2019) as no significant difference was found in the process model comprehension. Figl and Strembeck (2015) speculate that this result may be explained in part by humans' ability to adapt quickly to uncommon reading directions (e.g., right-to-left). Even when theoretically there is a relationship between the flow consistency of a process model and its understandability, this has not yet been proved.

There are several situations where the process model will not follow only one direction. Business process descriptions can result in complex models. The process flow may include sequential as well as parallel task executions. It can also include loops where the same tasks are performed more than once or repeatedly. The process description needs to express conditions to determine which of the several paths needs to be followed depending on the situation. (Figl & Strembeck, 2014). Gateways create more complex paths to follow and to understand than simple sequence flows (Figl & Strembeck, 2015) e.g. loops can be created by connecting an object via an arrow to an "upstream" object; being this an object performed earlier in the ordering which will flow again to the same activity until a logic constraint stops the loop. (Object Management Group, 2011). In this case, an arrow against the general flow direction of the model is used. Schrepfer (2010) in his guidelines mentions that for human readers the arrows flowing against the order specified as perceptual direction (referred to as backpointers) are irritating and increases the effort for graphical reading. This should be avoided or at least minimized.

In reality, the modeler has the freedom to decide which direction to use, and the models are not always possible to be drawn 100% consistently in the same direction. (Figl & Strembeck, 2015). This phenomenon can be found in practice when the model builder decides to design 'zigzag' models to fit the model to a specific size format, instead of reducing the overall size of model elements and labels. (Figl & Strembeck, 2015). Leopold et al. (2015) refer to this type of model as 'banana' models. According to the authors, there is no reason why to change the control flow direction in a model. Some users apply this practice to reduce the size but Leopold et al. (2015) consider this will only confuse reads due to an increased cognitive effort to recognize other tasks. 'One-page' size models are common in the industry due to their simplicity. The impact of this decision from practitioners on the understandability of a business process model is not known yet.

The direction a process model should have is still an open discussion among practitioners and researchers as still not scientific superiority has been able to be confirmed. The BPMN standard and other guidelines advise using a certain direction without scientific fundaments. (Figl & Strembeck, 2015). Winn (1982, p. 80) states that "diagrams not arranged in this logical sequence would lead to difficulty in information processing and less learning."

Regarding the Flow Direction aspect BPMN2.0 recommends the modelers use their judgment or best practices to decide how to connect the flow objects so the model readers can find the behavior clear and easy to follow. Only when using Message Flows advises on either use left-to-right or top-to-bottom direction because the models will be easier to understand. (Object Management Group, 2011). It doesn't mention any scientific base for this statement.

"In today's modeling practice we can observe a convention to model business processes from left-to-right or top-to-bottom. Even though the choice of flow direction changes the visual appearance of a process model significantly, this convention is barely discussed by standard documents and modeling guidelines. Besides, most recommendations related to the flow direction are neither based on scientific claims nor empirical evidence regarding their effectiveness in terms of readability". (Figl & Strembeck, 2014, p. 132). After reviewing the existent literature, this statement is still true.

Effinger (2011) moves the start event in a process model to the left of the canvas and events to the right side of the start event in their layout algorithm. Likewise, Kitzmann, König, Lübke, and Singer (2009) used a left-to-right orientation in their layout algorithm for BPMN diagrams, to match the horizontal progression of western handwriting. Top-to-bottom direction is less recommended in literature but Effinger, Siebenhaller, and Kaufmann (2009) consider left-to-right as well top-to-bottom as usual directions.

Effinger et al. (2010) consider 'flow' as an important layout aesthetic and use this feature to evaluate 5 BPMN tools. The authors mention that the number of connecting objects respecting workflow direction should be maximized and refer to 'flow' as "edges are drawn such way that they consider the reading direction" (Effinger et al., 2010, p. 35). The same is mentioned by Moreno-Montes de Oca and Snoeck (2014). But no empirical evidence of this has been found.

Petrusel et al. (2016) were able to confirm via means of eye-tracking experiments that modifying the layout and adding color cues have a positive impact on the mental effort and task efficiency, reducing the effort and time needed to understand a process model. Their hypothesis to confirm that modifying the layout of a process model has a positive impact on task effectiveness was not supported by the analysis of the data. But there is no link with the flow consistency of the models. Petrusel et al. (2016) investigate model elements, which were made larger and were repositioned, and found no evidence of an effect on understandability effectiveness, even though the mental effort measured by eye-tracking is reduced. (Figl, 2017).

Flow direction's impact on graph understandability has been empirically confirmed. But the flow direction's impact on the business process model's comprehensibility is not significantly relevant. A chart in the graph drawing area and a process model have different purposes. The inconsistent findings could be explained as charts in the graph drawing area are not being used as procedural instructions to follow a process model in sequential order following logical constraints to identify the path to follow to reach a business goal. Or maybe this is because in the attempts to research the impact of the flow direction on the process model's understandability the flow consistency was not measured. If the Flow Consistency didn't significantly change during the experiment, this will not create a significant impact on the model's understandability. This is an assumption as flow consistency was not measured during previous experiments. This assumption should be investigated.

An important aspect of quality is 'fitness for use'. (Lindland et al., 1994). The 'fitness for use' should also need to be considered when measuring the understandability of the model as understandability is also part of the broad quality term. A model can be understood in different ways and will need to fulfill different quality requirements depending on who is the reader and the model's intention of use. The quality requirements of a process model will then depend on its purpose. (Dzepina & Lehner, 2018). This was not taken previously into consideration.

Appendix X – The link between Flow direction and writing language

The scientific literature reveals that there is a clear preference to assign "earlier-later" to left-to-right followed by top-to-bottom and to assign "cause-effect" to top-to-bottom and left-to-right (Handel, DeSoto, & London, 1968, p. 354). Considering this argument, one would feel more natural and it would seem more logical to design process models from left-to-right and top-to-bottom is likely the second-best option. These directions would also be matching readers' mental visual orientation. (Figl & Strembeck, 2014).

Nordbotten and Crosby (1999) provide empirical evidence in the area of data models with eye-tracking technology for the reading direction of a model. On average, 60% of the participants followed a text-like reading strategy from left-to-right and top-to-bottom, 40% of them an image-like reading strategy (starting in the center followed by scanning in different directions). People expect known learned diagram schemas in a diagram, and understanding the diagram is easier if these are available. (Figl & Strembeck, 2014; Harsel & Wales, 1987; Krohn, 1983; Winn, 1982).

Figl and Strembeck (2015) explain that expectations and experience will influence in which direction a model reader will read a process model and search for information on it. Experienced model users will understand a model easier in comparison with a novice, because of the previously learned diagram schemes, being able to match user expectations. English-speakers will tend to 'read' diagrams in the same direction they read written language, in this case, left-to-right or top-to-bottom, so there is a strong influence of the language of the model reader on the way a business process is read and drawn. Diagrams not having this logical order would lead to difficulty in information processing and less learning, but from their experiment, no statistical proof was found this affects the understandability of the model. Even when the difference was statistically not significant, the comprehension of left-to-right models was higher than the same on right-left models. And a ceiling effect was noticed in the data as comprehension scores piled up at the end of the scale, being a threat to statistical conclusion validity. (Figl & Strembeck, 2015).

There are several attempts to define guidelines for process modeling "to create process models that analyst and business professionals can easily analyze and understand" (Mendling, Reijers, & Aalst, 2010, p. 1). One approach that is frequently cited is the 'Seven process modeling guidelines' (7MPG) from Mendling et al. (2010). These guidelines mention nothing regarding the direction of the flow the process model should have nor consider the flow consistency at all. It recommends using one start and one end event but mentions nothing regarding its location in the canvas. Is this framework then complete? Maybe on his attempt to create a short and simple framework to ease understand the completeness quality factor got less attention. Or it assumes the direction will be from left-to-right or top-to-bottom as it is common in practice without having any scientific support?

The same phenomenon was found in other guidelines for process modeling. The Framework BEBoP (understandaBility vErifier for Business Process models) from Corradini et al. (2018) include more guidelines than 7MPG, fifty guidelines to be exact. Guideline 43 'Design neat and consistent models' advice to maximize the number of connecting objects respecting the workflow direction. This could be a synonym of flow consistency, but no more information is given. In guideline 45 'Use linear sequence flows' the authors advise to use linear sequence flows without useless folding helping to maintain a model clearly and in guideline 47 'Use a consistent process orientation' "the designer should draw pools horizontally and use consistent layout with horizontal sequence flows, and vertical message flows and associations. (Corradini et al., 2018, p. 37). But the authors don't mention how or why to accomplish this nor how this can be measured. The latter guideline is based on the BPMN modeling conventions for public administration from Switzerland, a document created to standardize the use of the numerous modeling options (i.e. degrees of freedom) of the BPMN standard throughout Switzerland. They recognize the challenge for practitioners as BPMN 2.0 gives too much freedom in such aspects to the modelers. Corradini et al. (2018) based their claims also on the article of Leopold et al. (2015) who found some layout inconsistencies in the industry models like 'incorrect modeling direction': "the modeling direction 'left-to-right' should be consistently applied". (Leopold et al., 2015, p. 5). The authors evaluated the models against 'correct' and 'understandable' models to find these inconsistencies but don't mention how these 'correct and understandable' models are being defined. Maybe using their knowledge and experience, or some guidelines, but this is not clear in the article. Additionally it doesn't mention what are the empirical basis of using the reference point to state this.

The Guidelines from Becker, Rosemann, and Von Uthmann (2000) for process modeling consider the constraints that only horizontal and vertical models are allowed but don't mention why or how this horizontal/vertical direction should be considered or measured.

Going more specific in guidelines of notations, like the specific for UML class diagrams from (Eichelberger & Schmid, 2009) advice horizontal or vertical orientation as this is more likely perceived as a figure than other direction. It states that "edges should highlight a uniform flow and therefore should have similar edge directions" (Eichelberger & Schmid, 2009, p. 1690). Again this can be considered a synonym for flow consistency but the term is still undefined and not clear.

The standard BPMN 2.0 gives advice and recommendations regarding the flow direction of process models. (Figl & Strembeck, 2015). In general, the BPMN 2.0 standard document gives the following advice concerning (flow object connection rules) modeling direction: "An incoming Sequence Flow can connect to any location on a Flow Object (left, right, top, or bottom). Likewise, an outgoing Sequence Flow can connect from any location on a Flow Object (left, right, top, or bottom). …BPMN allows this flexibility; however, we also RECOMMEND that modelers use judgment or best practices in how Flow Objects should be connected so that readers of the Diagrams will find the behavior clear and easy to follow. This is even more important when a Diagram contains Sequence Flows and Message Flows. In these situations, it is best to pick a direction of Sequence Flows, either left-to-right or top-to-bottom, and then direct the Message Flows at a 90° angle to the Sequence Flows. The resulting Diagrams will be much easier to understand." (Object Management Group, 2011, p. 42). So unless the model contains Message Flows it is not recommended to design a model with high flow consistency. But does this latter have an impact on the readability? As the advice from the OMG does not seems to be empirically proved.

For instance, English-speaking children draw temporal concepts from left-to-right, whereas right-to-left was dominant for Arabic and Hebrew-speaking children. For native English speakers, it is more difficult to learn sequences in reversed-order (right-to-left) than in normal-order (left-to-right) diagrams. Participants solving problems with flowcharts perform best when the direction of flowcharts is left-to-right making fewer errors and needed less time. Also, experiments showed that Japanese participants performed better with the material in the vertical direction, while the Australian participants with the horizontal version. However, subjects can develop considerably fast an appropriate perceptual strategy after time. (Figl & Strembeck, 2014).

Appendix XI – Selecting the research method

Five strategies were identified to be possible to be used for quantitative research: Case Study, Experiment, Survey, Research Action, and Grounded theory. But not all of them are suitable for this research, not being compatible with the deductive approach and goal of this study.

Many articles testing empirically the understandability of pragmatic quality factors are done utilizing *experiments*. (Figl, 2017). In the Experiments strategy, the contextual variables are highly controlled which increases its internal validity. It is used to study the probability of a change in an independent variable causing a change in a dependent variable. Such as the main objective of this research. To improve on the external validity and generalizability of the findings the setup should try to match a 'field-based' setup rather than a 'controlled environment'. (Saunders et al., 2019).

The *Survey* strategy is associated with the deductive research approach as well. It allows the collection of standardized data from a large number of respondents economically, allowing easy comparison. Is normally conducted through the use of questionnaires or structured interviews, or structured observation. This strategy is often used for quantitative research and measure relationships between variables but does not support manipulation of the independent variable. For this research the flow consistency being the independent variable needs to be manipulated to measure the impact on the understandability, being an unsuitable strategy. (Saunders et al., 2019).

The techniques: action research, case study, and grounded theory were considered not suitable for the objective of this research. (Saunders et al., 2019). Even when these techniques may also be used for quantitative research, different limitations were found to the applicability of this research.

The *Case study* is a real-life setting mostly linked to qualitative methods of research. The purpose is to understand the dynamics of the topic being studied within its real-life setting or context. In this research, the relationship between variables will be investigated rather than the understandability being influenced by the real-life setting context. So it is considered not suitable for this research. The main difference with an experiment is that the variables are not controlled, this affects the validity of the results and has limitations regarding generalizability. (Saunders et al., 2019).

The *Action Research* strategy is designed to develop solutions to real organizational problems through a participative and collaborative approach. This research aims to increase the body of knowledge with a scientific aim rather than solve organizational practical problems. It is an iterative process that could difficult to be finished within the timeframe of this research. The longitudinal nature of this strategy means that it is more appropriate for medium or long-term research projects rather than short-term ones. It is about "research in action rather than research about the action" (Saunders et al., 2019, p. 203).

The *Grounded Theory* strategy is mostly linked to a qualitative research method. It is used to develop theoretical explanations of social interactions and processes and is an inductive approach to build theory. The approach of this research is rather deductive. (Saunders et al., 2019).

Even when Survey could be also applied for this kind of research, it is considered that an experiment is the most suitable strategy. This strategy allows to manipulate the flow consistency during the experiment and to measure accurately the data of the impact this manipulation causes. Also is possible to control most of the variables during the setting. It is an expectation that an experiment will improve the quality and variability of the conclusions and insights resulting from the data analysis.

Appendix XII – Operationalization of Variables

For this research below variables were used and operationalized as represented in Table 18. Three variables highlighted in gray were not used in this research.

The flow consistency -subjective measure- was not gauged in this research because the objective measure of the flow consistency is reliable and consistent and the subjective measure is not.

It was decided not to gauge the impact on the variable Task Efficacy as this is a ratio between two of the to-be measured variables. In case a significant impact is proved on one of these variables it will mean the Understandability of the model is impacted. The subjective measure Perceived Preference was not measured either.

For this research subjective preference is not considered determinant for understandability.

Table 18 - Variables to be measured in the experiment

Concept	Variable	Measurement	Literature Source
Understandability (Objective measure)	Task Effectiveness / accuracy (dependent variable)	The number of correct answers (tasks successfully completed) divided by the total number of questions (or tasks).	(Dikici et al., 2017; Figl, 2017; Figl & Strembeck, 2015)
	Task Efficiency (dependent variable)	The number of correct answers divided by the time it takes to complete the questions (or tasks).	
	Task Efficacy (dependent variable)	Task Effectiveness divided by the time needed to answer (or complete a task).	(Schrepfer, Wolf, Mendling, & Reijers, 2009)
Understandability (Subjective measure)	Perceived Ease of Understanding (dependent variable)	A 7-point Likert scale ranging Understanding the process model was very easy, Understanding the process model was easy, Understanding the process model was rather easy, Understanding the process model was neither difficult nor easy, Understanding the process model was rather difficult, Understanding the process model was difficult & Understanding the process model was very difficult.	(Figl & Laue, 2015)
	Perceived Preference (dependent variable)	A 7-point rating scale ranging from Extremely low preference (1) to Extremely high preference (7).	(Schrepfer et al., 2009; Turetken, Dikici, Vanderfeesten, Rompen, & Demirors, 2020)
Flow consistency (Objective measure)	Flow consistency (independent variable)	Combination of M-BP and M-E2.	(Burattin et al., 2018)
Flow consistency (Subjective measure)	Perceived flow consistency (independent variable)	A 7-point Likert scale ranging from 1 “No consistency at all” to 7 “Complete consistency.”	(Burattin et al., 2018)
Personal Factors	Process Modelling Experience (moderator variable)	A 4-point scale 1= Less than a month ago, 2= Less than a year ago, 3= Less than three years ago, 4= More than three years ago	(Mendling et al., 2012)
	Process Modelling Intensity (moderator variable)	A 4-point scale 0=Never, 1=Less than once a month, 2=More than once a month, 3=Daily	(Mendling et al., 2012)

Concept	Variable	Measurement	Literature Source
	Gender (moderator variable)	Multiple choice question. F (Female), M (Male), X (gender-neutral)	
	Year of birth/Age (moderator variable)	Actual year minus Year of birth Single answer (e.g.1980)	
	Occupation	Multiple choice question. Student, Working, None of above	
	Years of work Experience (moderator variable)	Less than 1 year working 1 year or more but less than 2 years 2 years or more but less than 5 years 5 years or more but less than 10 years 10 years or more	
	Computer-Office work-related (moderator variable)	3 points scale Low= none or less than 5 hours per week; Medium= between 5 hours and 20 hours per week; & High = more than 20 hours per week	
	Preference of sequential ordering direction (moderator variable)	Multiple choice question. Left-to-right, right-to-left, bottom-to-top, top-to-bottom	
	Highest education degree completed	Multiple choice question Still studying at (high)school, High-school, Bachelor, Master, One or more years of university without a certificate.	
Business process model features	Number of activities & gateways (control variables)	Same for A1, A2, A3 & A4 Same for B1, B2, B3 & B4.	

In Table 19 the variables created for the statistical analysis are visualized.

Table 19 - Variables created for analysis

Variable name	Description
ExperienceXintesity_PM_A	Is equal to the multiplication of the variable Experience in process modeling scale (from 1 till 4) by the variable Frequency in process modeling scale (from 0 till 3)
Experience/Intensity Level	Discretized participants with process modeling in 2 groups (Low and High) considered as Low if variable ExperienceXintensity_PM_A result was from 0 until 5; and High if the result was from 6 until 12). These boundaries were assumed for this paper. It is considered HIGH as a participant at least working in process models less than three years ago + encountering process models daily, OR working in process models less than a year ago+ encountering process models more than once a month
Task Effectiveness Semantics Test	Number of correct answers in the semantics test divided by the total of questions
Time Experiment Model A	Total time in seconds done to solve all tasks related to Model A
Time Experiment Model B	Total time in seconds done to solve all tasks related to Model B
Task Effectiveness Model A	The number of correct answers related to model A divided by the total of questions related to model A.
Task Effectiveness Model B	The number of correct answers related to model B is divided by the total of questions related to model B.
Task Efficiency Model A	Task Effectiveness Model A divided by Time Experiment Model A
Task Efficiency Model B	Task Effectiveness Model B divided by Time Experiment Model B

Appendix XIII – Technical design: elaboration of the method in detail

Participants

The participants are novices and expert modelers. No discrimination by modeling experience, knowledge, or other criteria was done to participate (increasing the generalizability of the findings). The participant requirements were: willing to voluntarily participate, have a computer with an internet connection, understand English and be able to read a process model. The latter is defined as passing the Semantics Test with a score of at least 50% as having knowledge has been proved to impact the model's understandability. (Mendling et al., 2012).

The participants are approached via social networks (i.e. Facebook and LinkedIn) asking family, friends, and colleagues to participate and share the link to the experiment with their network. The latter was also distributed among the researcher's work colleagues in a multinational company.

Design

The experiment is done following a field-based design, not a laboratory-based setting. Each participant did this experiment on his/her own computer. The participants acted as model readers, having to solve understandability tasks regarding the presented model.

Material and experiment setting

Two process models were used during the experiment. One model was designed for this experiment. The second model was used already in another research by Mendling et al. (2012) and Mendling and Strembeck (2008) to research the impact of process model factors on the understandability of the model. This model is a real-life process for 'processing an order in the system'. As labels influence the understandability (Mendling & Strembeck, 2008; Mendling et al., 2012) to match the reality textual labels were used in both models. To eliminate the influence of *domain knowledge*, the activities were replaced with random actions having no relation with the real process (e.g. 'Turn on the TV', or 'Jump one step forward'). High attention was given that the activities were similar in length and none combination of the activities might implicit or explicitly indicate a possible sequence (e.g. activities 'Drink the coffee' and 'make coffee' might implicit indicate 'make coffee' needs to happen before you can 'drink the coffee'). An example of activities used is 'Turn on the TV' and 'Jump one step forward'. One wouldn't be able to relate these to a possible sequence, as both can happen before the other.

To manipulate the independent variable, from each model (i.e. Modal A and Model B) four different versions were created. From each version, both models have the same layout (see Table 2). The same value for the control variables is kept in all versions of each model. All different versions of the process models used in the experiment can be consulted in Appendix XIV. The first layout has a left-to-right direction having a high flow consistency. The second layout is a 'zig-zag' model, creating visually three different sections in the model, each one from left-to-right, but the reader needs to go back to the left after the end of each section to start with the next section of the process model. The third layout of the model is another possibility for a 'zig-zag- model, consisting also in three visual sections, the first starting from left-to-right, in the second section the flow direction of the model changes from right-to-left and at the last section has again a flow direction from left-to-right. The fourth layout of the model is a 'banana' model, starting from left to right, then going from top-to-bottom and ending with a right-to-left flow direction. The latter three versions described have lower flow consistency than the first one.

'Model A' consists of 16 activities, 4 'AND' gateways, and 2 'XOR' gateways. 'Model B' consists of 16 activities, 2 'AND' gateways, and 9 'XOR' gateways. Model A and Model B in left-to-right layout (Model A1 and B1 respectively) would not fit in one screen nor one 'landscape A4 paper'. To visualize these models in the computer the reader would need to scroll. To print the models both models would need 3 or more A4 paper sheets. The 'zig-zag' and 'banana' layouts were designed to fit into one screen and into one 'landscape A4 paper' size.

The experiment consists of six parts. The experiment protocol is visualized in Figure 6. During the experiment, the participants will be randomly assigned to one of the four groups.

Before starting the experiment the participant will receive a short introduction to the experiment asking to accept to participate voluntarily in the experiment, represented as part null.

In part one, the participant will reply to questions regarding gender, year of birth, work experience, computer-office work performed, preference of reading direction, highest degree completed.

In part two the participant will need to reply if he/she has previous modeling knowledge or experience. In case the answer is positive, the questions about modeling experience and frequency used by Mendling et al. (2012) will need to be answered by the participant.

In part three the participants will receive a short introductory explanation of semantics from the objects used in the models and how to read a model, based on the introduction of Mendling and van der Aalst (2006), to learn how to read a process model, followed by a semantics comprehension test.

In part four the participants will receive one version of 'Model A' and one version of 'Model B' respectively depending on the group number of the participant and will be requested to solve randomly the nine reasoning tasks for both models to measure understandability (two regarding flow sequence, two relating AND gateways, and two relating XOR gateway). The same questions were considered for all versions of the same model. For both models, different questions were used. The questions were adapted from the questions often used in this kind of experiment, e.g. used by Reijers et al. (2011). One example of these questions is: *Can activities 'Turn the other page' and 'Paint a wall' be executed at the same time for a case?* The model versions presented to each group are shown in Table 4. Special attention was given so the same group does not receive the same layout from both models, thus no group received the same layout nr (e.g. A1 and B1) nor received both zig-zag layouts (e.g. A2 and B3).

In part five the participant will be asked to answer their perception regarding the understandability of the previously processed models randomly using a seven-point Likert scale, single item cognitive load measure from Figl and Laue (2015) going from very difficult until very easy.

The introduction, the simplified explanation of the semantics of the process models used, and all the questions used during the experiment can be consulted in Appendix XV.

The experiment is developed in an online survey tool (www.limesurvey.org) allowing to record the time duration and correctness of the answers. Using an electronic survey tool allows also to assign the participants randomly equally between all groups.

The material will be tested by 3 novices before these were considered finalized. This to assure the questions were easy to understand, correctly formulated, and are not misinterpreted; increasing the criterion validity. (Saunders et al., 2019). The experiment and instructions were also reviewed in the previously mentioned pilot test to confirm they are easy to understand and to answer.

Appendix XIV – Models used in the experiment

For the experiment, a process model was designed having a high flow consistency with a left-to-right direction, named 'model A1' represented in Figure 18. The same model was adapted to fit one A4 paper size, converting the layout of the model to two 'zig-zag' model layout, 'model A2' and 'model A3', represented in Figure 19 and Figure 20 respectively, and one 'banana' model layout, 'model A4', represented in Figure 21.

A second process model is used Model B, this process model with a left-to-right direction, named 'model A1' is represented in Figure 22. The same model was adapted to fit one A4 paper size, converting the layout of the model to two 'zig-zag' model layout, 'model B2' and 'model B3', represented in Figure 23 and Figure 24 respectively, and one 'banana' model layout, 'model B4', represented in Figure 25.

Model A

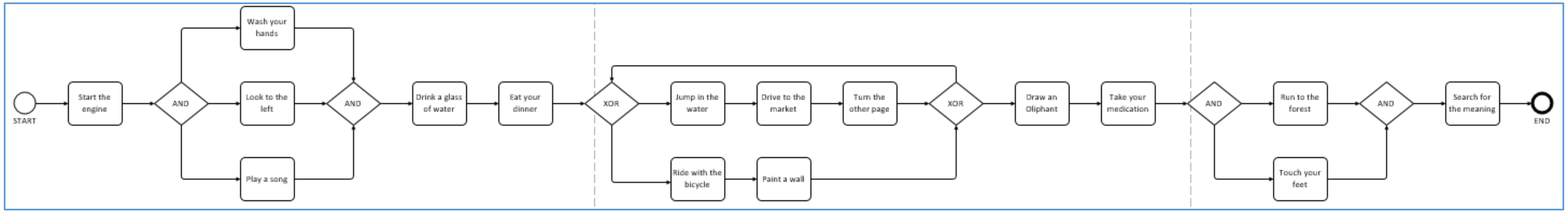


Figure 18 - Experiment 'model A1' - Left-to-right layout of Model A with high flow consistency

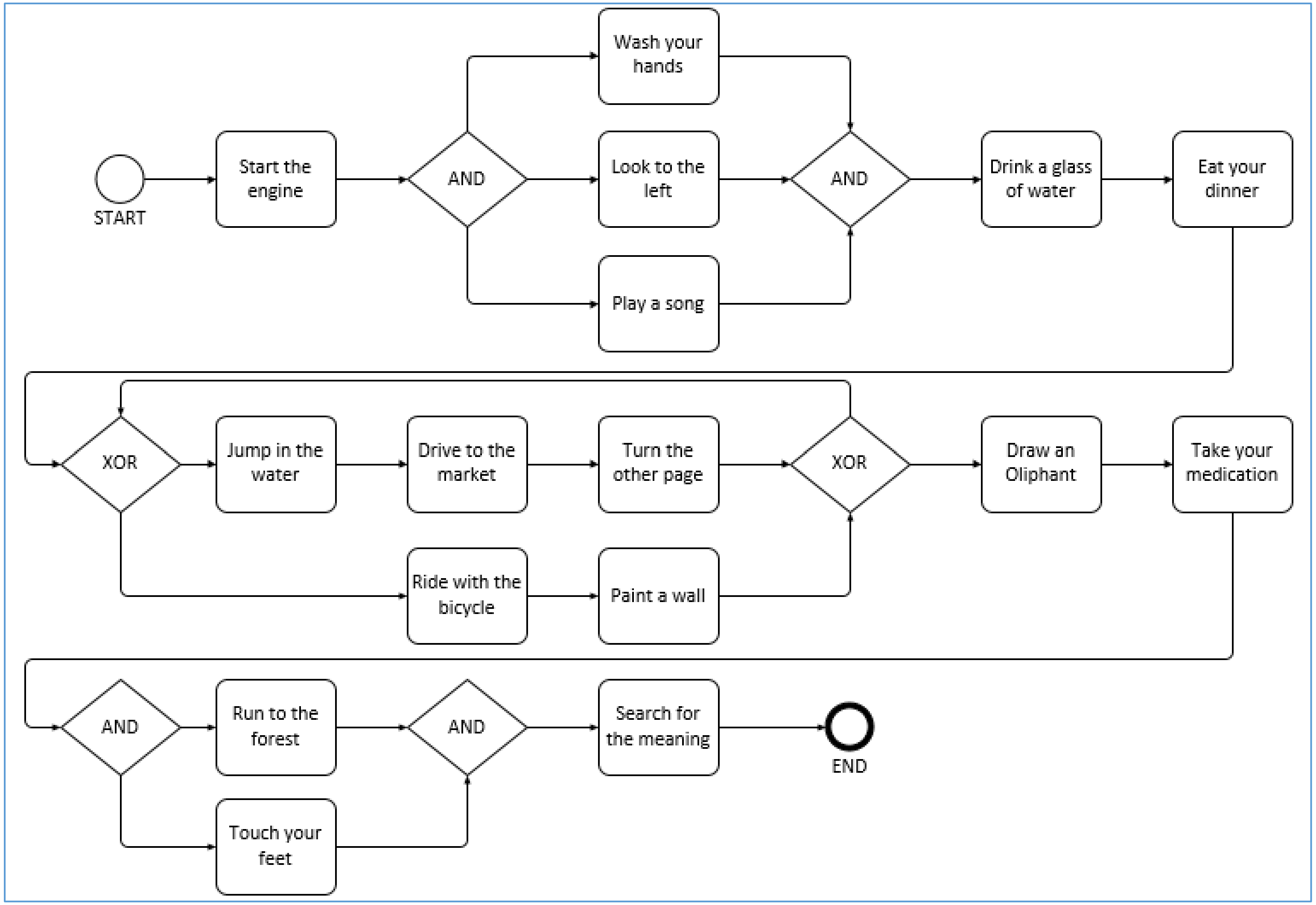


Figure 19 - Experiment 'model A2' - 'zig-zag' layout number 1 of Model A with medium flow consistency

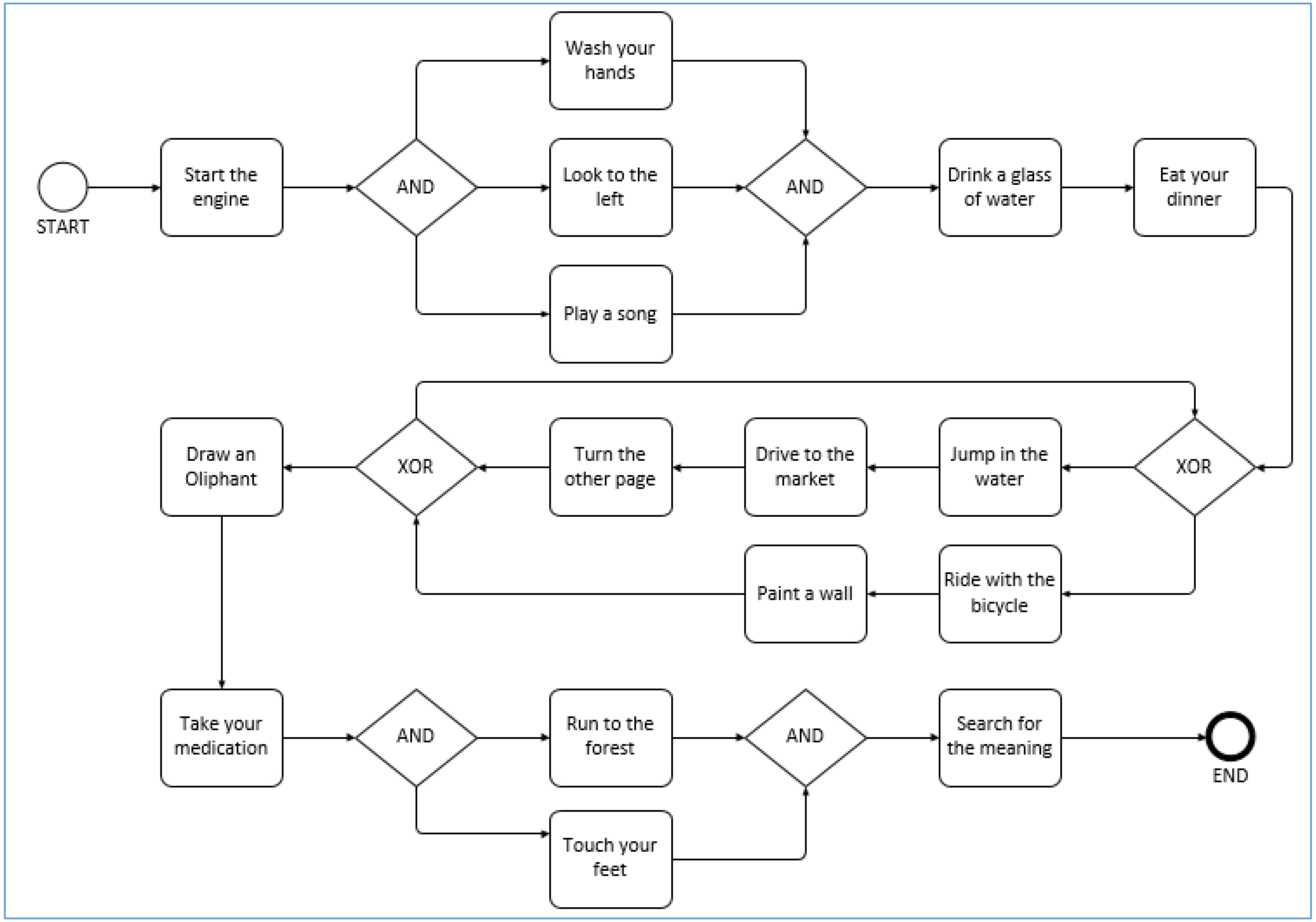


Figure 20 - Experiment 'model A3' - 'zig-zag' layout number 2 of Model A with low flow consistency

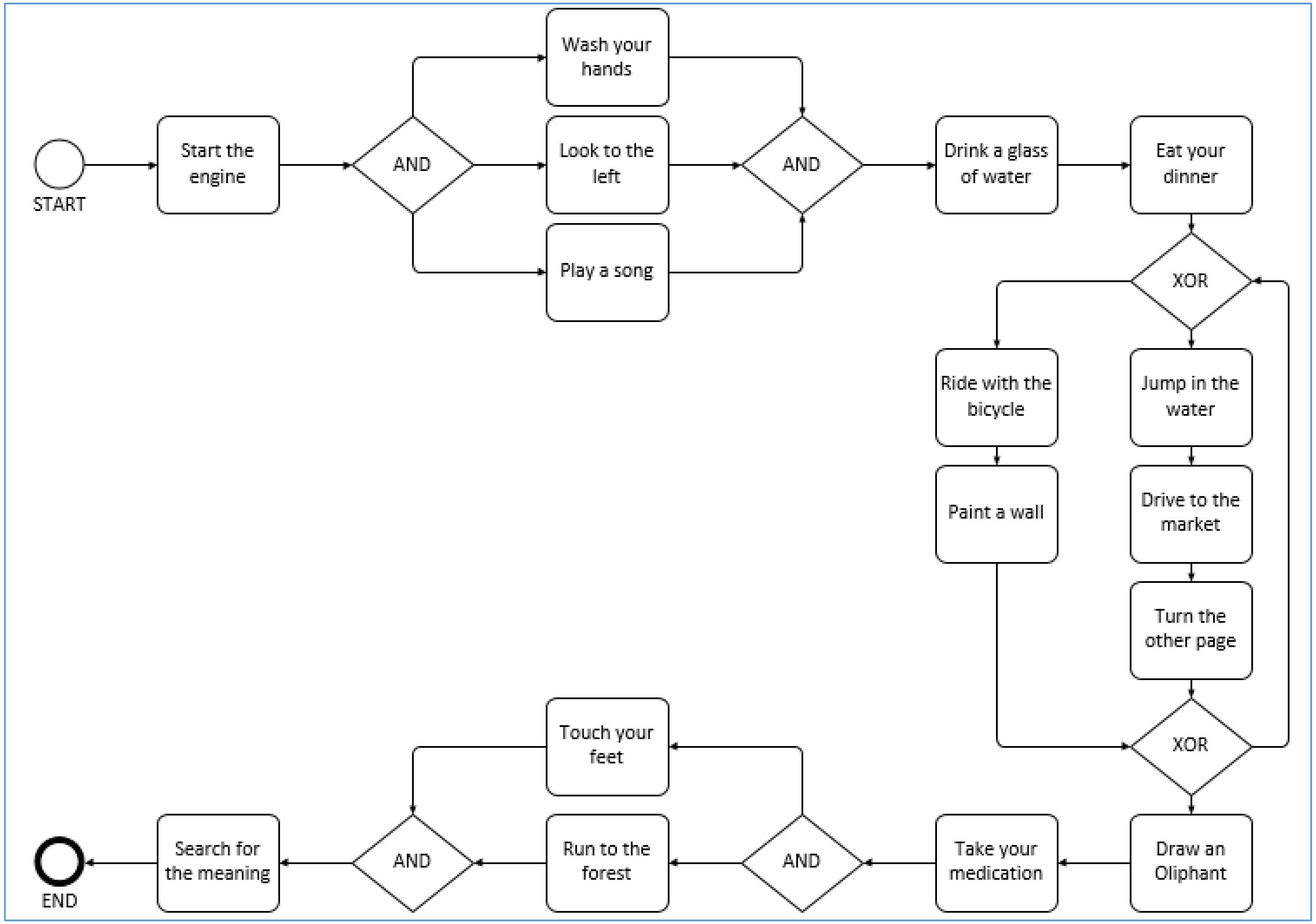


Figure 21 - Experiment 'model A4' - 'banana' layout of Model A with low flow consistency

Model B

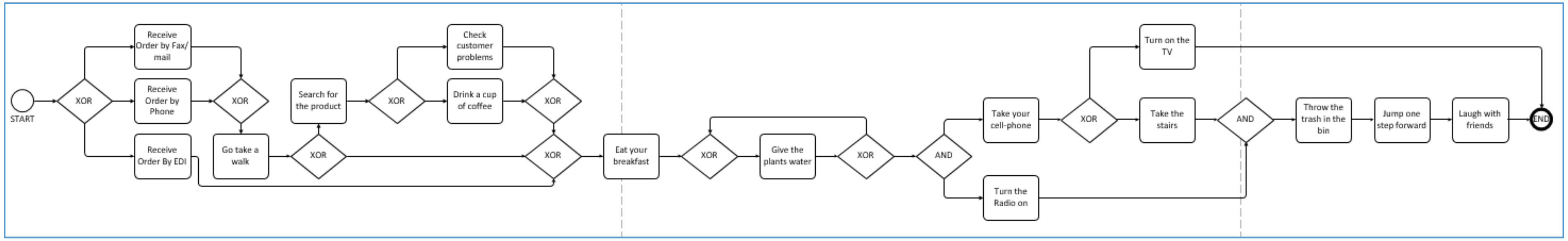


Figure 22 - Experiment 'model B1' - Left-to-right layout of Model B with high flow consistency

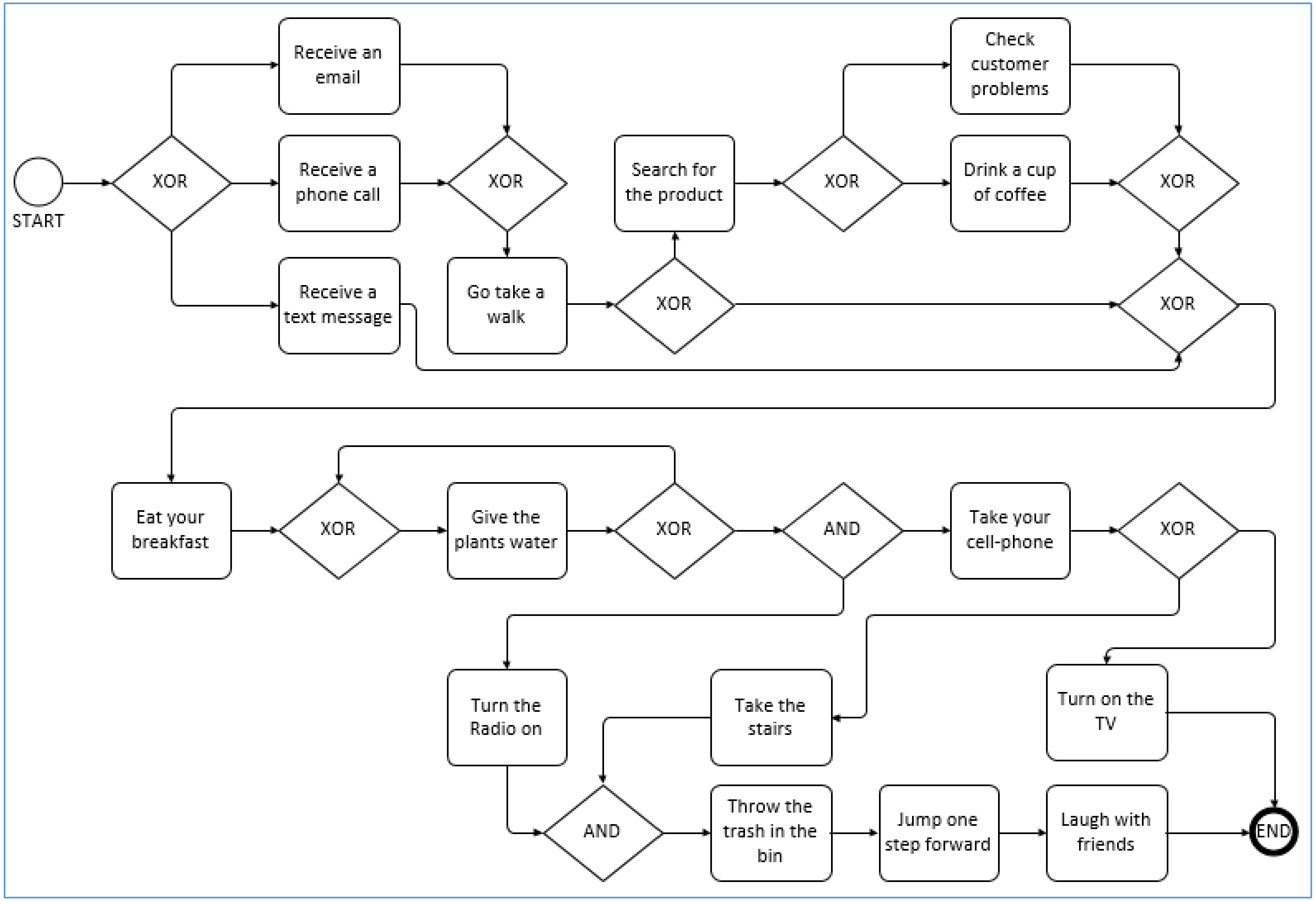


Figure 23 - Experiment 'model B2' - 'zig-zag' layout number 1 of Model B with medium flow consistency

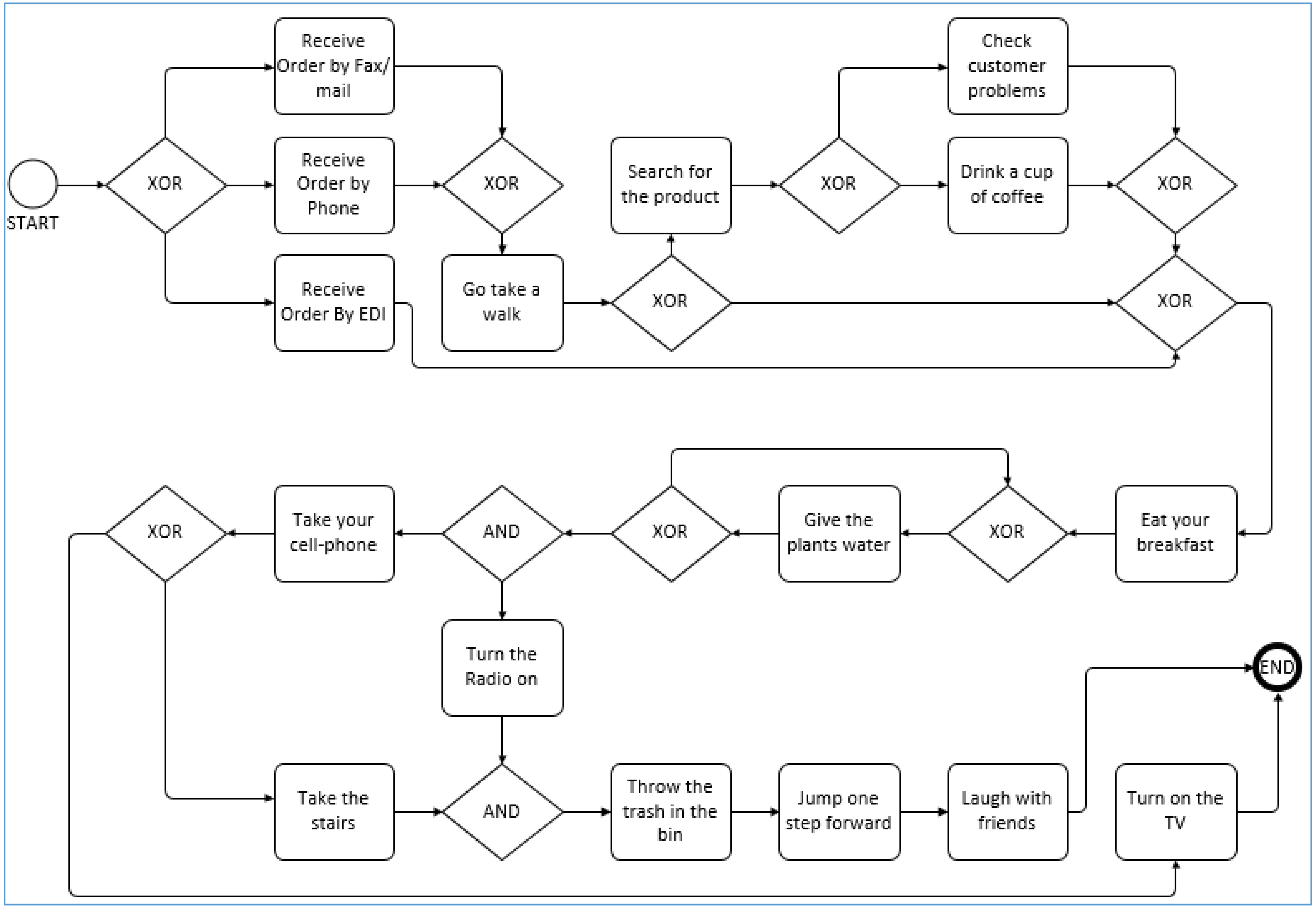


Figure 24 - Experiment 'model B3' - 'zig-zag' layout number 2 of Model B with low flow consistency

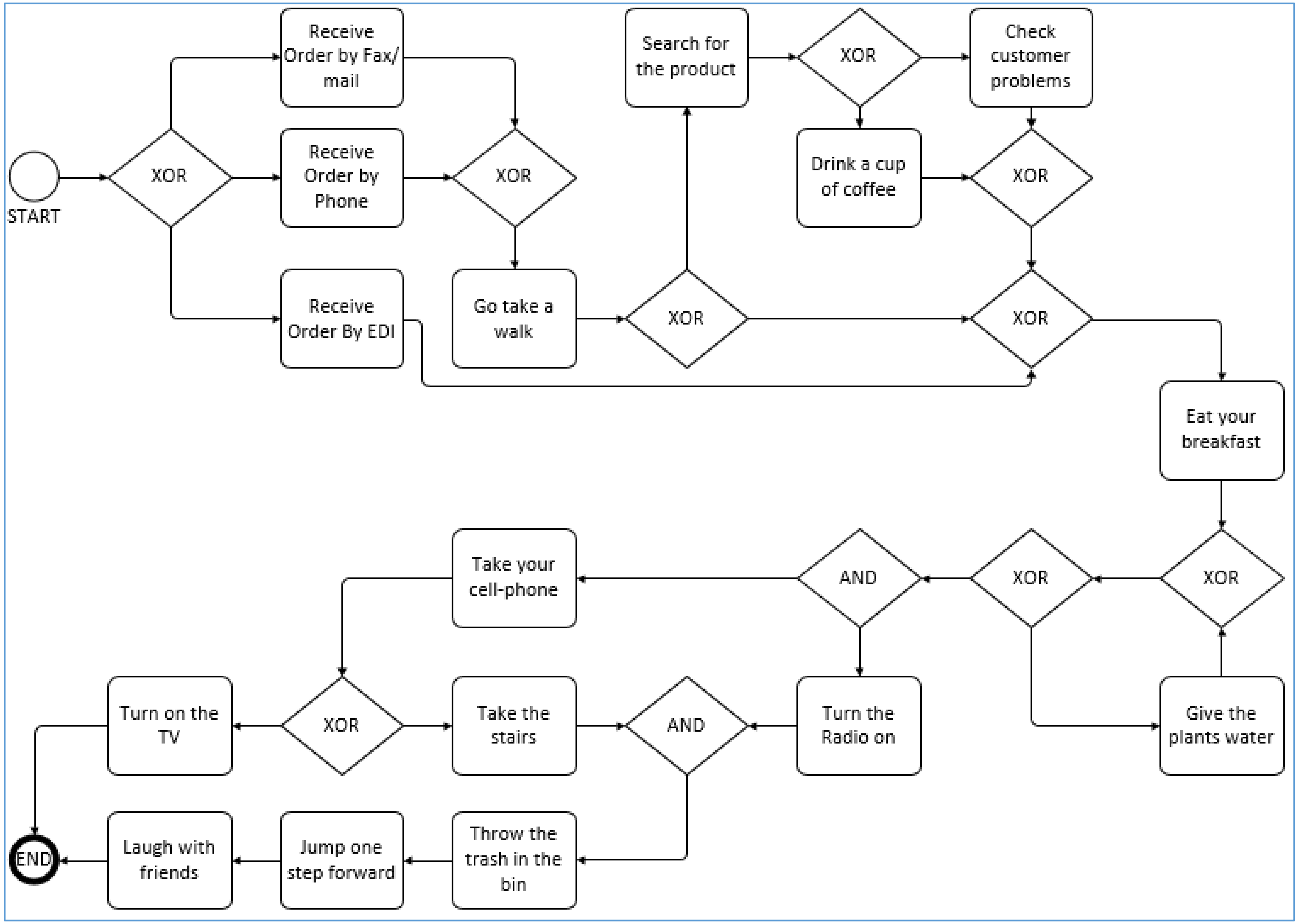


Figure 25 - Experiment 'model B4' - 'banana' layout of Model B with low flow consistency

Appendix XV – Questions of the experiment

Part 0 – Introduction to the experiment

Thanks for participating in this experiment about business process models.

In continuation, you will be requested to reply to several questions grouped in different sections.

This experiment takes about 10 minutes to be completed. Please do this experiment in a peaceful environment, without stopping, and without distractions or hurries. The time taken to answer some questions will be measured.

Answer the questions alone, without help from other persons, and with care. Answer with the truth and do not guess as this might impact the results.

The data is being recollected and analyzed anonymously and is only used for this research. Data manipulation and storage comply with the GDPR. Participation in the experiment is voluntary, by clicking on 'Next' you accept to participate.

Part 1 – Personal Factors

Q11 Occupation - Select your current occupation

- A1 - School or university student
- A2 - Employee (working)
- A3 - None of above

Q12 Work experience - Select the number of years you have worked (e.g. 5 years). Do not consider Student Jobs.

- A1 - Less than 1 year working
- A2 - 1 year or more but less than 2 years
- A3 - 2 years or more but less than 5 years
- A4 - 5 years or more but less than 10 years
- A5 - 10 years or more

Q13 Computer-office work performed - Select one of the below options that describe the best the use of a computer during your work.

- L - Low= none or less than 5 hours per week
- M - Medium= between 5 hours and 20 hours per week
- H - High = more than 20 hours per week

Q14 Preference of sequential ordering direction – Imagine a circle and a square. If you place the square BEFORE the circle with an arrow in between giving the direction from the square to the circle, in which direction did you imagine the arrow?

- A1 - Left-to-right
- A2 - Right-to-left
- A3 - Bottom-to-top
- A4 - Top-to-bottom

Q15 Highest degree completed – Select the option which describes the best your last education degree completed.

- A1 - Still studying at school
- A2 - School certificate
- A3 - One or more years of university without a certificate
- A4 - Bachelor certificate
- A5 - Master certificate

Q16 Gender - Select from below the gender which applies to you.

- F - F (Female)
- M - M (Male)
- X - X (gender-neutral)

Q17 Year of birth - Select the year when you were borne (e.g.1980)

- A1 - (year)

Part 2 – Modelling Experience and Frequency

Q21 Do you have any experience or knowledge about process modeling?

- Y - Yes
- N - No

Q22 Process Modelling Experience - When did you first work with process models in practice?

- 1 - More than three years ago
- 2 - Less than three years ago
- 3 - Less than a year ago
- 4 - Less than a month ago

Q23 Process Modelling Frequency - How often do you encounter process models in practice?

- 0 - Never
- 1 - Less than once a month
- 2 - More than once a month
- 3 - Daily

Part 3 – Semantics explanation

Q30 Explanation

To be able to answer the next questions you need to be able to read and understand a process model. A short explanation is given in continuation from a simplified representation of process models.

A test will follow to assure the content was well understood.

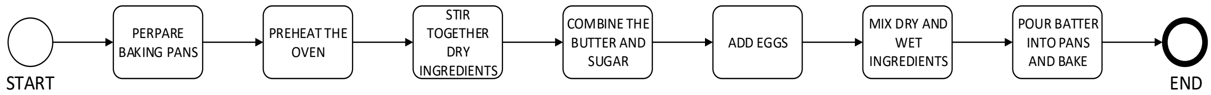
A process can be divided into activities which after being completed will generate an expected result.

For example, ‘making a cake’ is a process and the expected result is a ‘cake’. This process can be divided into activities such as:

1. Prepare baking pans
2. Preheat the oven
3. Stir together dry ingredients
4. Combine the butter and sugar
5. Add eggs
6. Mix dry and wet ingredients
7. Pour batter into pans and bake.

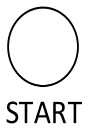
These activities are arranged in temporal order, in other words, in which order the activities need to be started and completed.

A process can be visually represented in a model. Below is a representation from the process model 'making a cake'.



In the model different objects (symbols) with a different meaning.

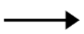
The beginning of the process is represented by the start object.



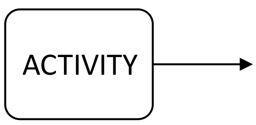
The end object represents when the process finishes.

The object used to represent the activities (actions) are rectangles with rounded corners.

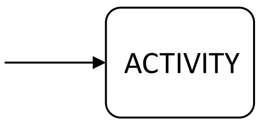
Arrows (control flow arcs) are used to link elements.



Activities may have only 1 outgoing arrow



And 1 incoming arrow

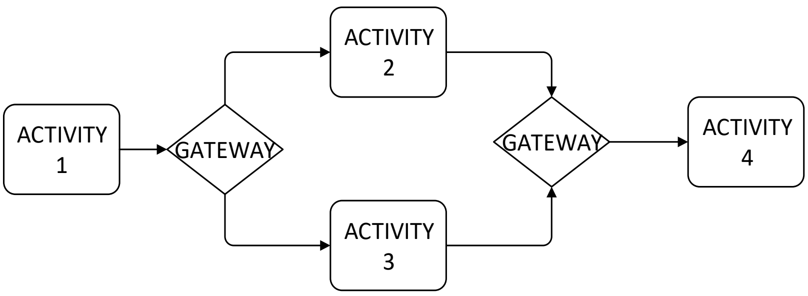


The first activity after the start object will start the process. When this activity is completed then the next activity following the arrow may start being processed. And so on, when the 2nd activity is completed then the 3rd activity may start being processed.

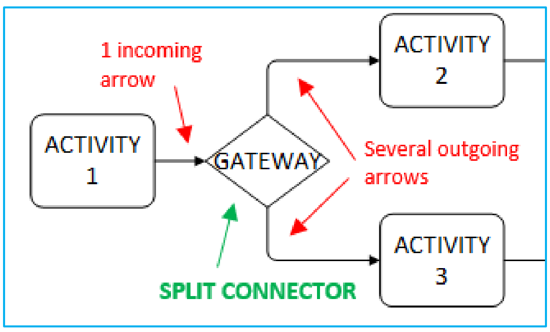
For each **'Case'** a process is triggered. For each *case,* a process will start from the start until the end. Using the previous 'making a cake' example process, each cake will represent a *case*. For each cake (case) we want to bake, we will follow the activities of the 'making a cake' process starting each time from the first activity until the last activity.

Gateways are connector objects represented by a rhombus.

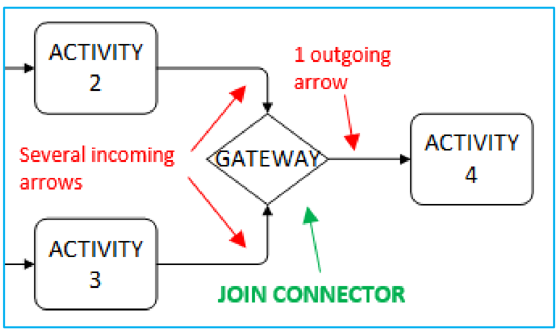
To link one activity to several activities (or several activities to one activity) a gateway must be used between the activities.



The gateways may have 1 incoming arrow and several outgoing arrows (split connectors)



Or several incoming arrows and 1 outgoing arrow (join connectors).



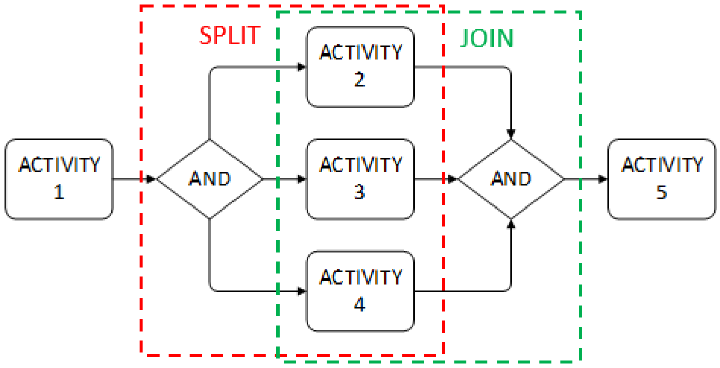
These gateways will add logic to the process to be able to proceed to the next activity. There are three kinds of logical connectors (i.e. AND, OR, and XOR).

The AND split activates all next activities in a concurrent fashion. Meaning **all next activities may be processed** simultaneously, but it is not mandatory to process them at the same time. In the below example when Activity 1 has been completed then the following activities may start being processed:

- Activity 2 **AND** Activity 3 **AND** Activity 4.

The AND-join waits for all incoming activities to be completed, then enables the next activity to be processed. In the below example, Activity 5 will be able to be processed only when the below activities have been completed (not before):

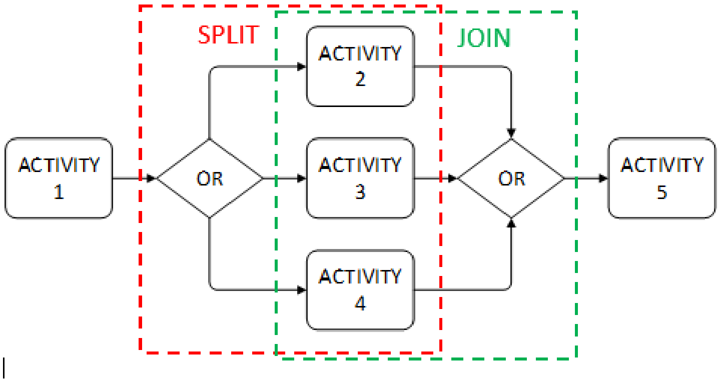
- Activity 2 **AND** Activity 3 **AND** Activity 4.



The OR-split triggers **one, two, or up to all of the next activities** based on conditions. In the below example when Activity 1 has been completed then the following activities may start being processed:

- (one activity) Activity 2 **OR**
- (one activity) Activity 3 **OR**
- (one activity) Activity 4 **OR**
- (two activities) Activity 2 and Activity 3 **OR**
- (two activities) Activity 3 and Activity 4 **OR**
- (two activities) Activity 2 and Activity 4 **OR**
- (all activities) Activity 2 and Activity 3 and Activity 4

The OR-join synchronizes all active incoming flows. It needs to know which of the previous Activities needed to be processed and waits for all these activities to be completed, to enable the next activity to be processed. In the below example, if Activity 2 and Activity 3 needed to be processed, it will wait until these 2 activities are completed to be able to start processing Activity 5.



The XOR-split is an *EXCLUSIVE OR* and represents a choice between <u>**one**</u> of the next activities. Meaning **only 1** of the next activities will be processed. In the below example when Activity 1 has been completed then the following activities may start being processed:

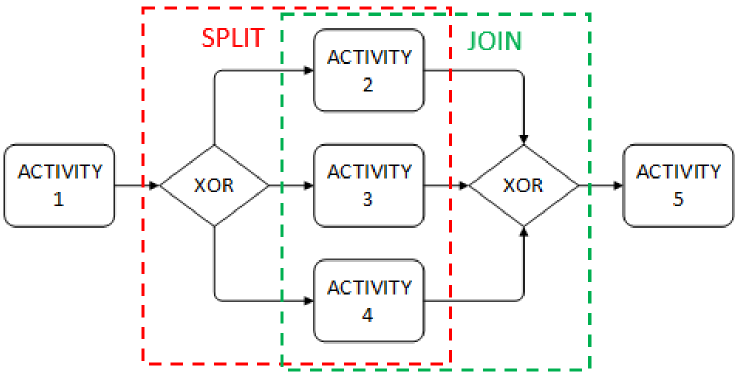
- Activity 2 **OR**
- Activity 3 **OR**
- Activity 4

The **difference** with the OR-split is that with XOR-split **only 1** of the next activities will be processed.

The XOR-join merges the different flows. It waits for the incoming activity to be completed, then enables the next activity to be processed. In the below example, Activity 5 will be able to be processed only when the below activities have been completed (not before):

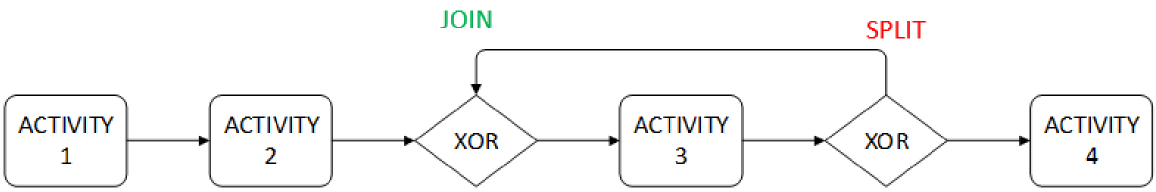
- Activity 2 **OR**
- Activity 3 **OR**
- Activity 4

In **difference** with the OR-split, with XOR-split **only 1** the previous activities will be completed.



With OR and XOR connectors it is possible to create loops or repetitive tasks. Even when this should be avoided, some processes require this. In the below example after Activity 2 has been completed Activity, 3 will start to be processed.

Note between Activity 2 and 3 a join XOR gateway is used (not a split gateway).



When Activity 3 is completed a decision can be taken depending on a condition. **OR** Activity 4 can start to be processed **OR** with the reversed arrow Activity 3 will need to be processed again.

This loop will continue until the condition in the XOR is true to process activity 4.

In this case Activity 3 will be **possible to be processed several times for the same case.**

(Mendling & van der Aalst, 2006)

Part 3 – Semantics Test

Q31 (Concurrency) - Consider the process fragment given in Figure 26. Can activities 'Drink a cup of coffee' and 'Give the plants water' be executed at the same time for a case?

- Y - Yes
- N – No
- X - I don't know

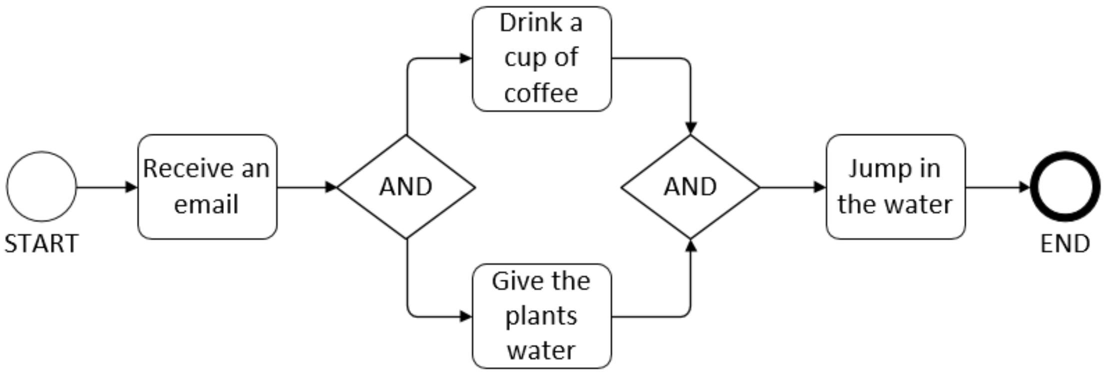


Figure 26 - Model used in Semantics test -question 1

Q32 (Exclusiveness) - Consider the process fragment given in Figure 27. Can the activities 'Drink a cup of coffee' and 'Give the plants water' both be executed for the same case?

- Y - Yes
- N – No
- X - I don't know

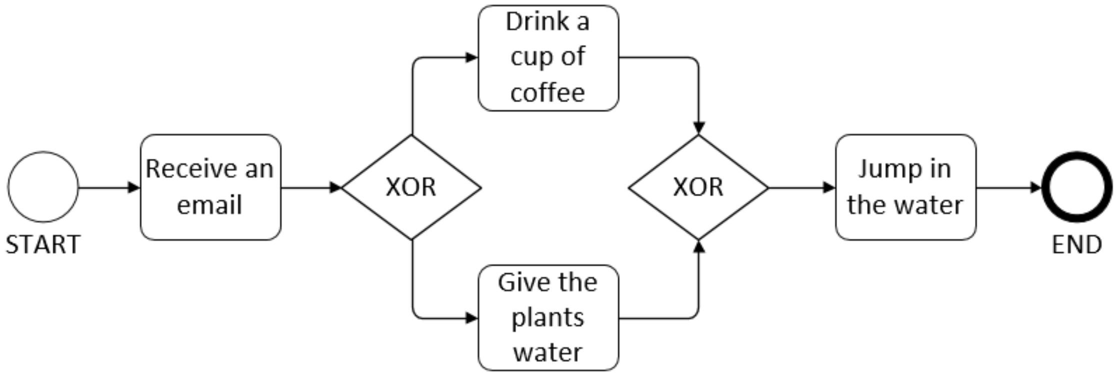


Figure 27 - Model used in Semantics test -question 2

Q33 (Order) - Consider the process fragment given in Figure 28. If activity 'Jump in the water' is executed for a case, must then always activity 'Drink a cup of coffee' been executed for the same case?

- Y - Yes
- N – No
- X - I don't know

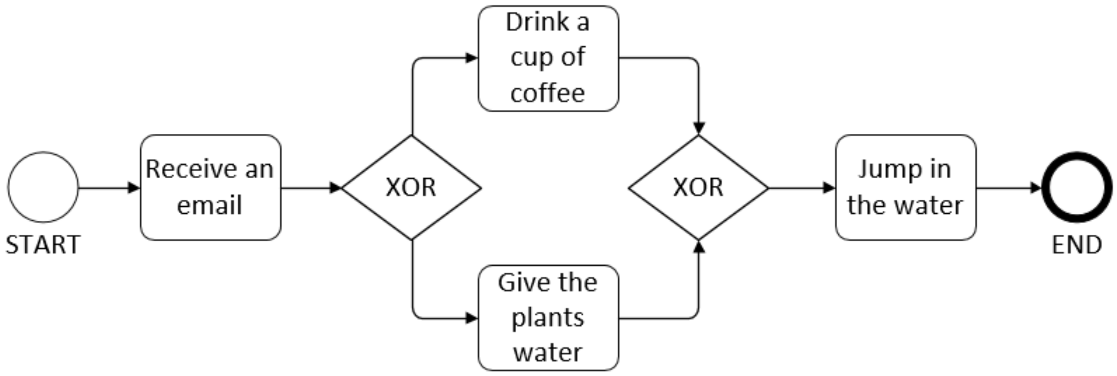


Figure 28 - Model used in Semantics test -question 3

Q34 (Repetition) Consider the process fragment given in Figure 29. Can activity 'Give the plants water' be executed more than once for the same case?

- Y - Yes
- N – No
- X - I don't know

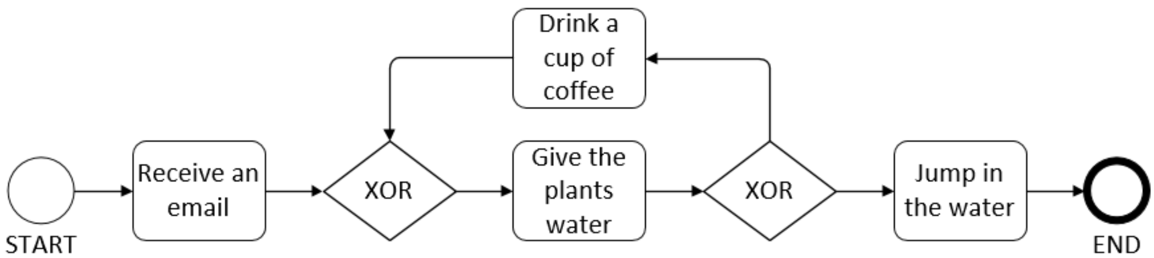


Figure 29 - Model used in Semantics test -question 4

Part 4 – Understandability test Model A

Q41 (Concurrency) Can activities 'Run to the forest' and 'Touch your feet' be executed at the same time for a case?

- Y - Yes
- N – No
- X - I don't know

Q42 (Concurrency) Can activities 'Turn the other page' and 'Paint a wall' be executed at the same time for a case?

- Y - Yes
- N – No
- X - I don't know

Q43 (Exclusiveness) Can the activities 'Drive to the market' and 'Paint a wall' both be executed for the same case?

- Y - Yes
- N – No
- X - I don't know

Q44 (Exclusiveness) Can the activities 'Ride with the bicycle' and 'Turn the other page' both be executed for the same case?

- Y - Yes
- N – No
- X - I don't know

Q45 (Order) - If activity 'Look to the left' is executed for a case, must then activity 'Run to the forest' be executed for the same case?

- Y - Yes
- N – No
- X - I don't know

Q46 (Order) - If activity 'Drink a glass of water' is executed for a case, must then activity 'Drive to the market' be executed for the same case?

- Y - Yes
- N – No
- X - I don't know

Q47 (Order) - Can this process be completed by executing thirteen or less activities?

- Y - Yes
- N – No
- X - I don't know

Q48 (Repetition) Can activity 'Jump in the water' be executed more than once for the same case?

- Y - Yes
- N – No

- X - I don't know

Q49 (Repetition) Can the activity 'Run to the forest' be executed more than once for the same case?

- Y - Yes
- N – No
- X - I don't know

Part 4 – Understandability test Model B

Q61 (Concurrency) Can activities 'Take your cell-phone' and 'Turn the Radio on' be executed at the same time for a case?

- Y - Yes
- N – No
- X - I don't know

Q62 (Concurrency) Can activities 'Take your cell-phone' and 'Throw the trash in the bin' be executed at the same time for a case?

- Y - Yes
- N – No
- X - I don't know

Q63 (Exclusiveness) Can the activities 'Take the stairs' and 'Turn on the TV' both be executed for the same case?

- Y - Yes
- N – No
- X - I don't know

Q64 (Exclusiveness) Can the activities 'Go take a walk' and 'Eat your breakfast' both be executed for the same case?

- Y - Yes
- N – No
- X - I don't know

Q65 (Order) - If activity 'Drink a cup of coffee' is executed for a case, must then activity 'Give the plants water' be executed for the same case?

- Y - Yes
- N – No
- X - I don't know

Q66 (Order) - If activity 'Take your cell-phone' is executed for a case, must then activity 'Take the stairs' be executed for the same case?

- Y - Yes
- N – No
- X - I don't know

Q67 (Order) - Can this process be completed by executing nine or less activities?

- Y - Yes
- N – No
- X - I don’t know

Q68 (Repetition) Can the activity ‘Give the plants water’ be executed more than once for the same case?

- Y - Yes
- N – No
- X - I don’t know

Q69 (Repetition) Can the activity ‘Take the stairs’ be executed more than once for the same case?

- Y - Yes
- N – No
- X - I don’t know

Part 5 – Perceived Understandability test Model A

Q51 Select one of the below options which reflects the best your perception you have regarding the model processed in the previous section.

- A1 - Understanding the process model was very difficult.
- A2 - Understanding the process model was difficult.
- A3 - Understanding the process model was rather difficult.
- A4 - Understanding the process model was neither difficult nor easy.
- A5 - Understanding the process model was rather easy.
- A6 - Understanding the process model was easy.
- A7 - Understanding the process model was very easy.

Part 5 – Perceived Understandability test Model B

Q71 Select one of the below options which reflects the best your perception you have regarding the model processed in the previous section.

- A1 - Understanding the process model was very difficult.
- A2 - Understanding the process model was difficult.
- A3 - Understanding the process model was rather difficult.
- A4 - Understanding the process model was neither difficult nor easy.
- A5 - Understanding the process model was rather easy.
- A6- Understanding the process model was easy.
- A7 - Understanding the process model was very easy.

End message

Thanks for participating in this experiment! This will help me and the scientific community to generate new insights. I hope you enjoyed it and you were able to learn a few new things about process models. Please click on SUBMIT to finish this experiment.

If you would like to transfer any comment to the maker of this experiment please feel free to write it below. This is not mandatory.

Appendix XVI – Demographic details of participants

The demographic distribution from the participants per group can be visualized in Table 20. 47% of the participants are females and 53% are males. These are similarly distributed between the groups, except for group 3 having significantly more males than females.

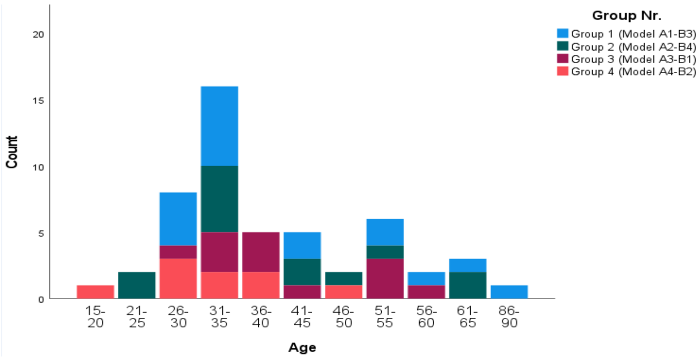


Figure 30 - Age distribution of participants in the experiment

Most of the participants (90.1%) have an age between 21 and 60 years. The biggest group have are between 31 and 35 years old. The age distribution of the participants can be visualized in Figure 30.

88,2% of the participants work (main occupation) and three participants (5,9%) are students. The remaining 5,9% (3 participants) expressed having work experience (not working at the moment of the experiment). With exception of the three participants still studying at school, all the participants have a school, bachelor, or master certificate (33.3%, 31.4%, and 29,4% respectively).

60,4% of the working participants have more than 10 years of work experience, followed by 25% of participants with work experience between 5 and 10 years. The rest have less work experience. Group 2 is mostly composed of participants with 10 or more years of work experience, while the other groups are distributed similarly. The detail can be consulted in Table 20.

Table 20 - Demographic details of participants in the experiment

		Group Nr.									
		Group 1 (Model A1-B3)		Group 2 (Model A2-B4)		Group 3 (Model A3-B1)		Group 4 (Model A4-B2)		Total	
		Count	Layer Column N %	Count	Layer Column N %	Count	Layer Column N %	Count	Layer Column N %	Count	Layer Column N %
Age	15-20	0	0,0%	0	0,0%	0	0,0%	1	11,1%	1	2,0%
	21-25	0	0,0%	2	15,4%	0	0,0%	0	0,0%	2	3,9%
	26-30	4	23,5%	0	0,0%	1	8,3%	3	33,3%	8	15,7%
	31-35	6	35,3%	5	38,5%	3	25,0%	2	22,2%	16	31,4%
	36-40	0	0,0%	0	0,0%	3	25,0%	2	22,2%	5	9,8%
	41-45	2	11,8%	2	15,4%	1	8,3%	0	0,0%	5	9,8%
	46-50	0	0,0%	1	7,7%	0	0,0%	1	11,1%	2	3,9%
	51-55	2	11,8%	1	7,7%	3	25,0%	0	0,0%	6	11,8%
	56-60	1	5,9%	0	0,0%	1	8,3%	0	0,0%	2	3,9%
	61-65	1	5,9%	2	15,4%	0	0,0%	0	0,0%	3	5,9%
	86-90	1	5,9%	0	0,0%	0	0,0%	0	0,0%	1	2,0%
Education	I'm still studying at school or I have no school certificate	1	5,9%	1	7,7%	0	0,0%	1	11,1%	3	5,9%
	School certificate	7	41,2%	1	7,7%	4	33,3%	1	11,1%	13	25,5%
	One or more years of university without a certificate	1	5,9%	2	15,4%	1	8,3%	0	0,0%	4	7,8%
	Bachelor certificate (University)	4	23,5%	5	38,5%	3	25,0%	4	44,4%	16	31,4%
	Master certificate (University)	4	23,5%	4	30,8%	4	33,3%	3	33,3%	15	29,4%
Ocupation	Student	1	5,9%	1	7,7%	0	0,0%	1	11,1%	3	5,9%
	Working	14	82,4%	12	92,3%	11	91,7%	8	88,9%	45	88,2%
	None	2	11,8%	0	0,0%	1	8,3%	0	0,0%	3	5,9%
Work_experience	Less than 1 year working	0	0,0%	1	8,3%	0	0,0%	0	0,0%	1	2,1%
	1 year or more but less than 2 years	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0	0,0%
	2 years or more but less than 5 years	2	12,5%	0	0,0%	2	16,7%	2	25,0%	6	12,5%
	5 years or more but less than 10 years	6	37,5%	0	0,0%	3	25,0%	3	37,5%	12	25,0%
	10 years or more	8	50,0%	11	91,7%	7	58,3%	3	37,5%	29	60,4%
Computer_office_level	Low	0	0,0%	1	8,3%	1	9,1%	0	0,0%	2	4,4%
	Medium	1	7,1%	3	25,0%	1	9,1%	2	25,0%	7	15,6%
	High	13	92,9%	8	66,7%	9	81,8%	6	75,0%	36	80,0%
Mental_ordening	BOTTOM to TOP	0	0,0%	0	0,0%	0	0,0%	0	0,0%	0	0,0%
	LEFT to RIGHT	15	88,2%	12	92,3%	11	91,7%	8	88,9%	46	90,2%
	RIGHT to LEFT	2	11,8%	1	7,7%	1	8,3%	0	0,0%	4	7,8%
	TOP to BOTTOM	0	0,0%	0	0,0%	0	0,0%	1	11,1%	1	2,0%

Most of the participants are used to work with a computer, matching the experiment setup (understand process models in a computer). 80% of the participants work more than 20 hours per week with a computer and seven participants (15.6%) between 5 and 20 hours per week. Group 4 has fewer participants having a high level of the computer-office job. Group 1 and 4 have no participants with low computer-office level.

22 participants (out of the total of 51 participants) expressed to have experience and knowledge of process modeling. These participants were discretized into two groups: Low and High experience/knowledge, as described in Table 19. Thirteen participants (59.1%) were classified as having *low* experience/knowledge and nine participants (40.9%) were classified as having *high* experience/knowledge. Group 1 has more participants with low experience/knowledge in process modeling than other groups, the rest are distributed similarly among the groups.

When asking the participants in which direction they order an object in their imagination which is 'before' other object, 90.2% responded having a mental ordering 'first-then' in a direction from left-to-right (following the read direction). Four participants (7.8%) arrange the objects mentally in a right-to-left direction, which is less common. These participants have different demographic details not being able to identify a common characteristic among them. One participant answered to order the objects in a top-to-bottom direction. The latter might be a participant with a relation with an Asiatic culture where text is written from top to bottom.

Appendix XVII – Spearman correlation matrix – output SPSS

Table 21 - Correlation matrix - Spearman (rho)

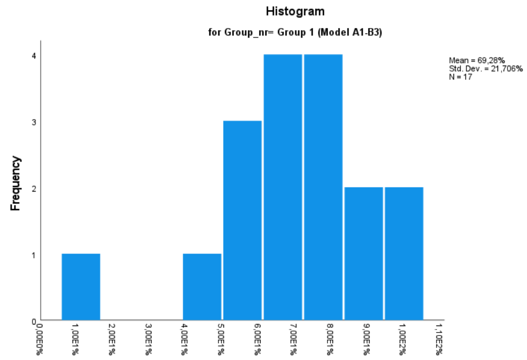
		Flow consistency Model A	Task Efficiency Model A	Perceived Undersandability Model A	Task Effectiveness Model A	Flow consistency Model B	Task Effectiveness Model B	Task Efficiency Model B	Perceived Undersandability Model B	Work experience	Computer office level	Education
Task Efficiency Model A	Corr. Coef.	0,01	1									
	Sig. (2-tailed)	0,94										
Perceived Undersandability Model A	Corr. Coef.	0,22	0,11	1								
	Sig. (2-tailed)	0,12	0,43									
Task Effectiveness Model A	Corr. Coef.	0,11	**,669****	0,16	1							
	Sig. (2-tailed)	0,45	**0,00**	0,26								
Flow consistency Model B	Corr. Coef.	**,712****	0,05	0,06	0,16	1						
	Sig. (2-tailed)	**0,00**	0,74	0,66	0,27							
Task Effectiveness Model B	Corr. Coef.	0,08	**,429****	0,27	**,744****	0,19	1					
	Sig. (2-tailed)	0,56	**0,00**	0,06	**0,00**	0,17						
Task Efficiency Model B	Corr. Coef.	0,13	**,667****	0,19	**,432****	0,11	**,505****	1				
	Sig. (2-tailed)	0,38	**0,00**	0,19	**0,00**	0,42	**0,00**					
Perceived Undersandability Model B	Corr. Coef.	0,18	0,24	**,575****	**,298***	0,10	**,361****	0,21	1			
	Sig. (2-tailed)	0,21	0,10	**0,00**	**0,03**	0,50	**0,01**	0,13				
Work experience	Corr. Coef.	-0,05	-0,06	0,09	0,14	-0,02	0,05	-0,22	-0,07	1		
	Sig. (2-tailed)	0,72	0,66	0,56	0,35	0,89	0,75	0,13	0,65			
Computer office level	Corr. Coef.	0,21	0,07	0,11	0,08	0,13	0,23	0,24	0,17	0,10	1	
	Sig. (2-tailed)	0,16	0,65	0,49	0,62	0,40	0,12	0,12	0,26	0,50		
Education	Corr. Coef.	-0,18	**,381****	-0,07	**,279***	-0,09	**0,27**	**,277***	0,06	-0,16	-0,06	1
	Sig. (2-tailed)	0,21	**0,01**	0,64	**0,05**	0,52	**0,05**	**0,05**	0,67	0,27	0,71	
Experience /intensity level	Corr. Coef.	-0,12	0,26	**-,544****	0,12	-0,02	-0,04	0,26	-0,35	0,38	0,24	0,09
	Sig. (2-tailed)	0,61	0,25	**0,01**	0,60	0,92	0,84	0,25	0,11	0,10	0,33	0,69
****. Correlation is significant at the 0.01 level (2-tailed).**												
***. Correlation is significant at the 0.05 level (2-tailed).**												

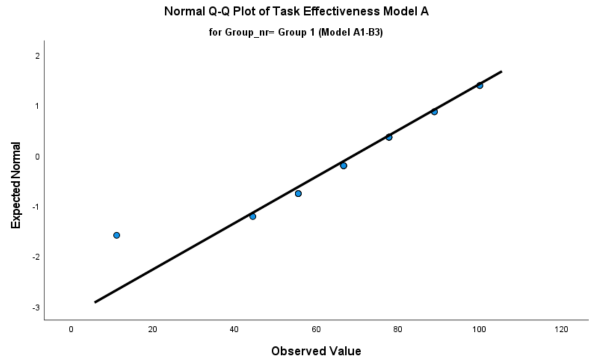
Appendix XVIII – Histograms and Q-Q plot charts

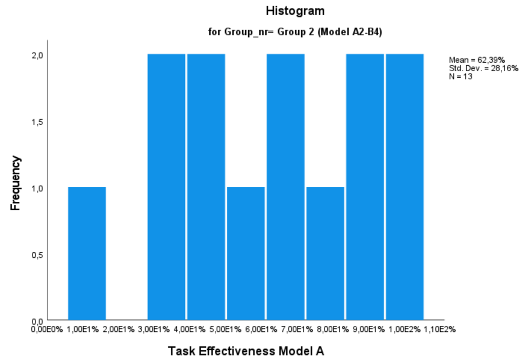
Table 22 - Evaluation of Normal Distribution per variable

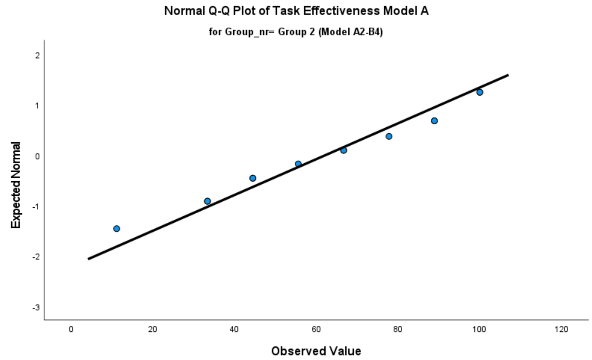
Variable	Version Model	Shapiro-Wilk p>0,05	Bell-shaped polygon	Normal as per Normal Q-Q plot	Conclusion
Task Effectiveness Model A	A1	Yes	Yes	No	Not Normal distributed
	A2	Yes	No	No	
	A3	No	No	Yes	
	A4	Yes	No	No	
Task Efficiency Model A	A1	No	Yes	Yes	Not Normal distributed
	A2	Yes	No	Yes	
	A3	Yes	Yes	Yes	
	A4	No	No	Yes	
Perceived Undersandability Model A	A1	Yes	Yes	No	Not Normal distributed
	A2	No	No	No	
	A3	Yes	No	No	
	A4	Yes	Yes	Yes	
Task Effectiveness Model B	B1	No	No	No	Not Normal distributed
	B2	Yes	No	Yes	
	B3	Yes	No	No	
	B4	No	Yes	No	
Task Efficiency Model B	B1	Yes	Yes	Yes	Not Normal distributed
	B2	No	No	Yes	
	B3	No	Yes	No	
	B4	Yes	No	No	
Perceived Undersandability Model B	B1	Yes	Yes	Yes	Normal distributed
	B2	Yes	Yes	Yes	
	B3	Yes	Yes	Yes	
	B4	Yes	Yes	Yes	

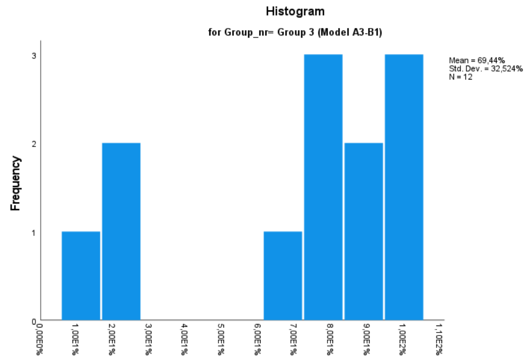
Model A – Task Effectiveness

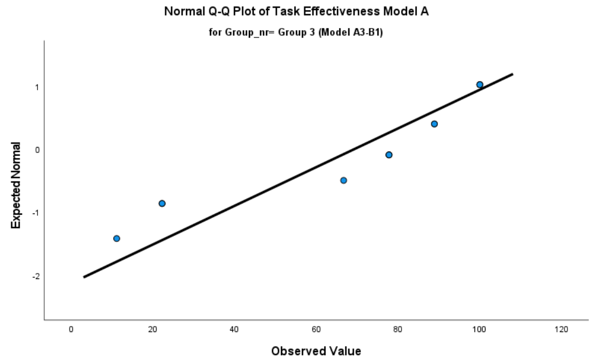


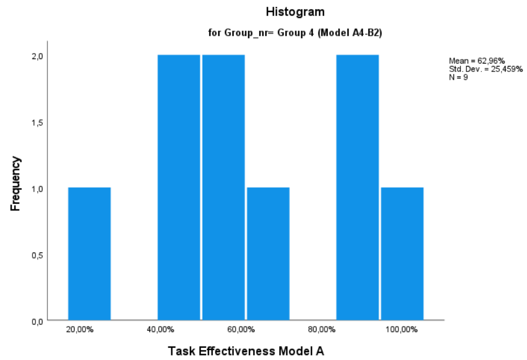


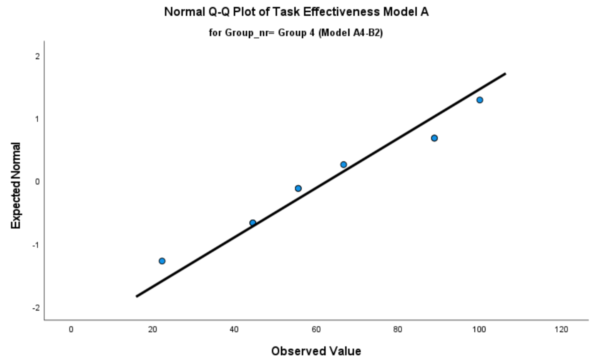




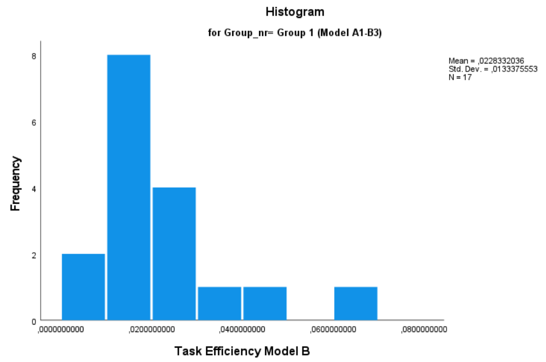


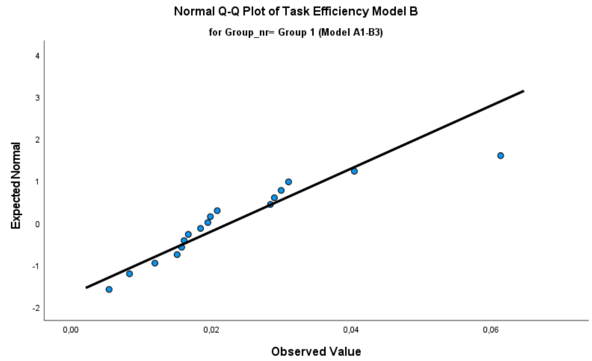


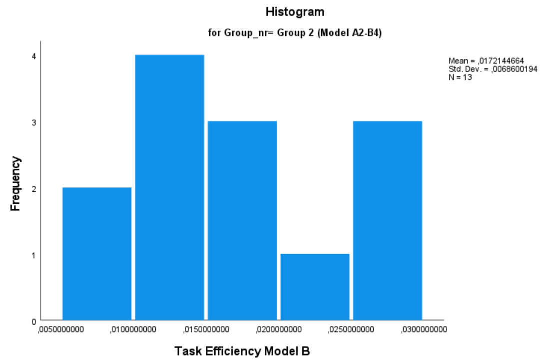


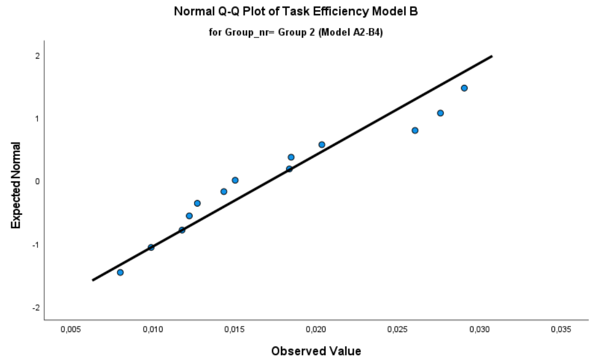


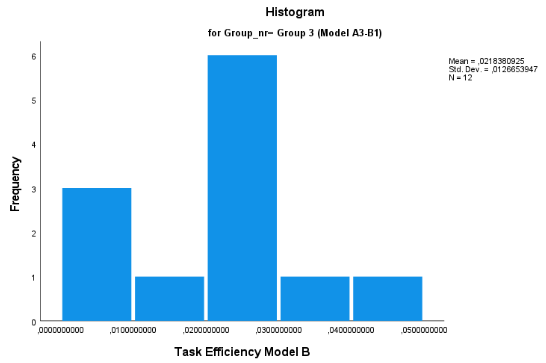
Model A – Task Efficiency

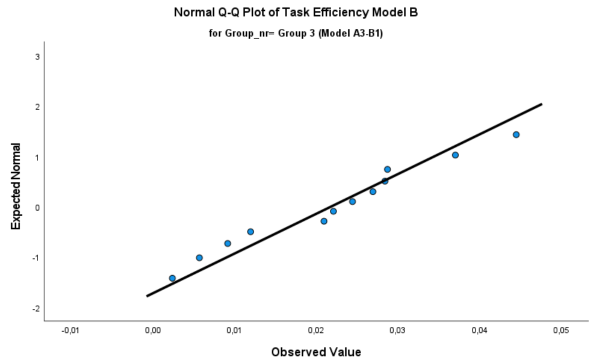


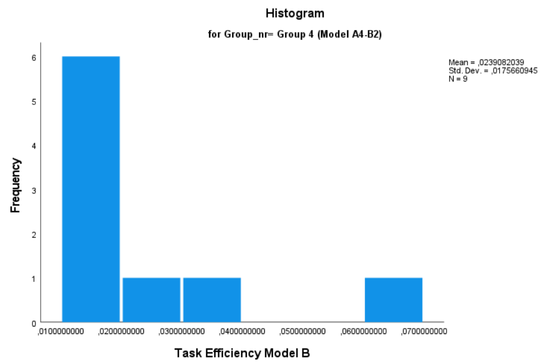


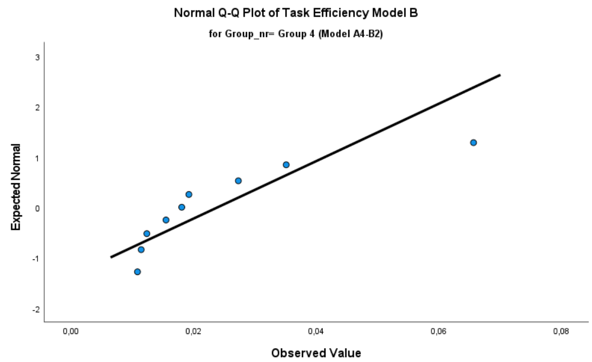




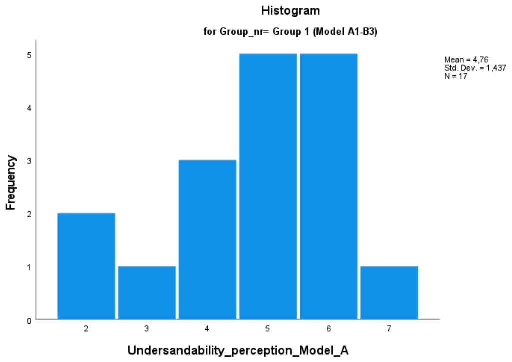


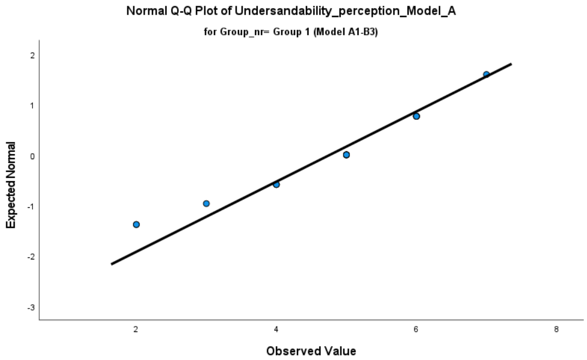


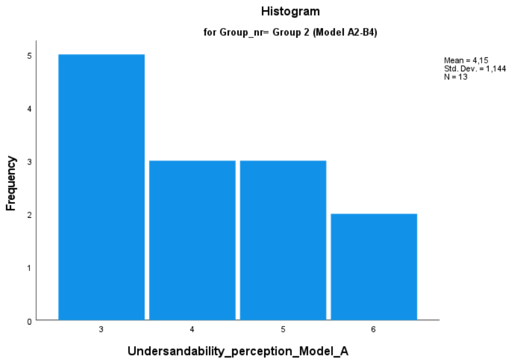


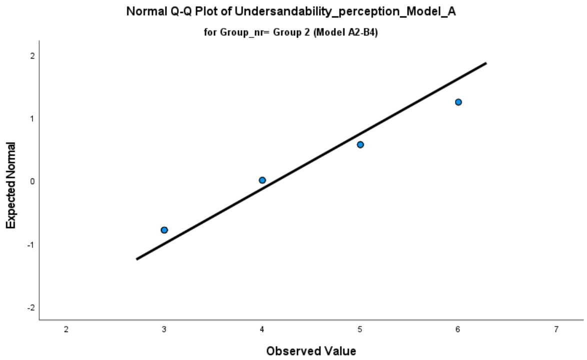


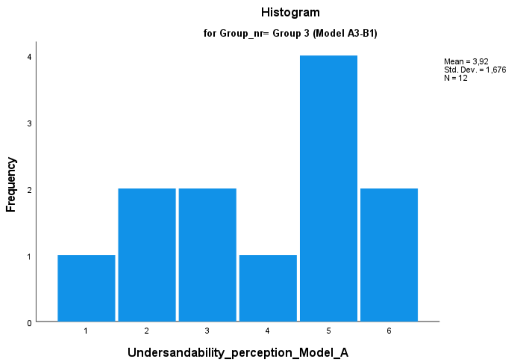
Model A – Perceived Understandability

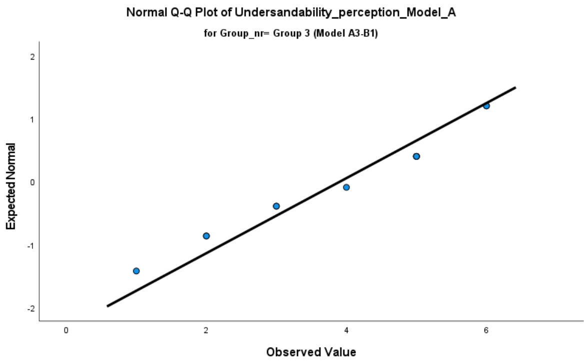


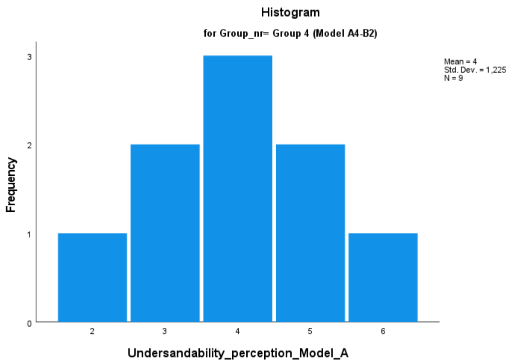


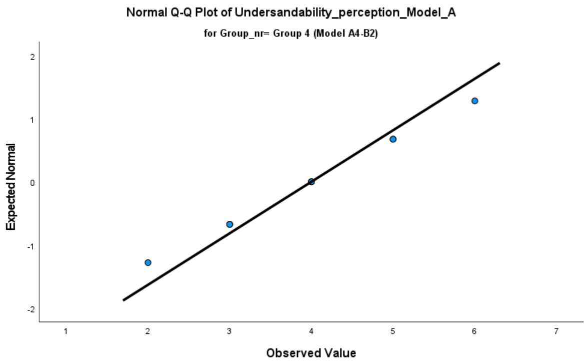




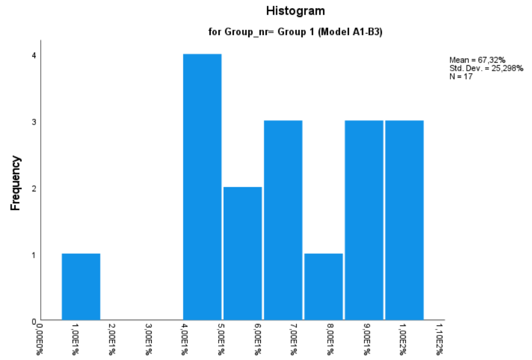


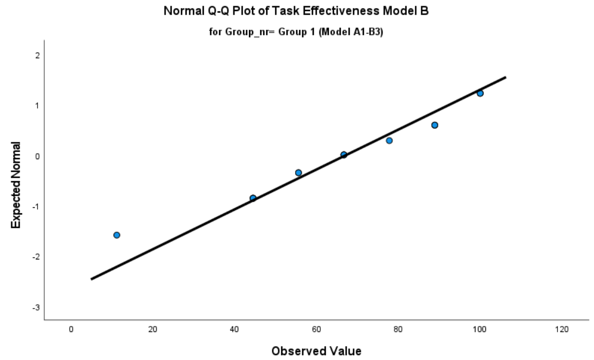


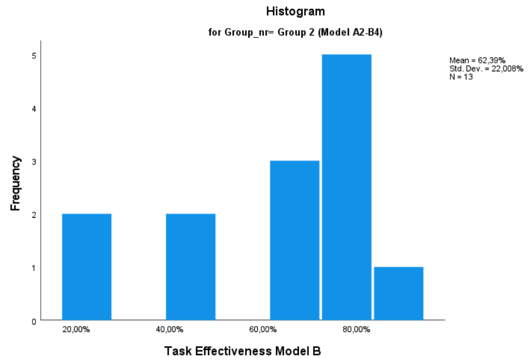


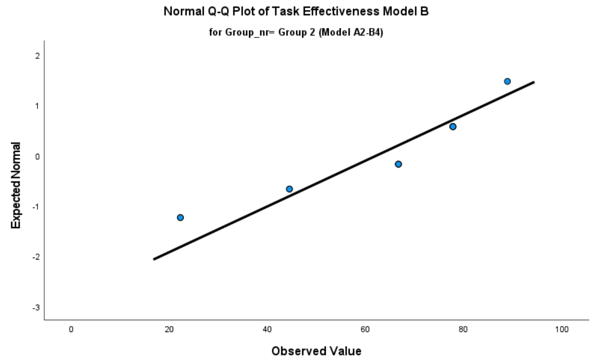


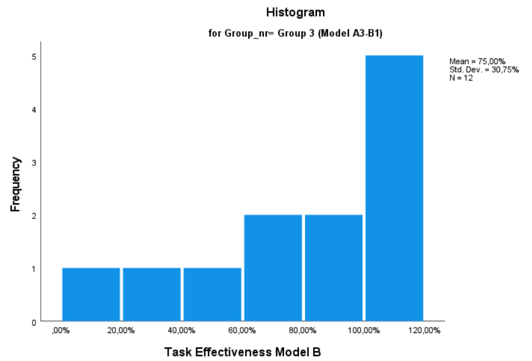
Model B – Task Effectiveness

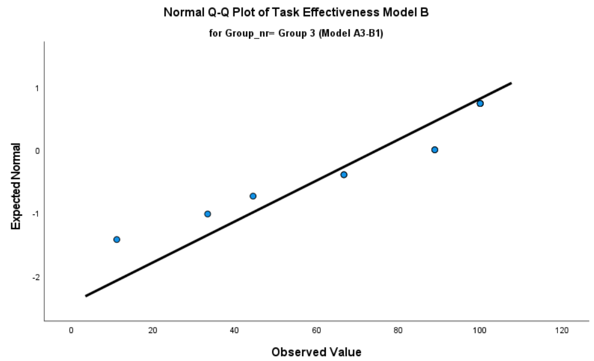


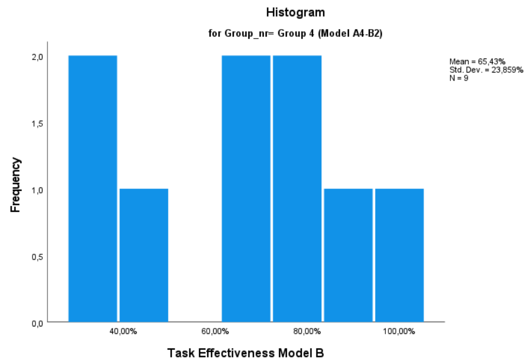


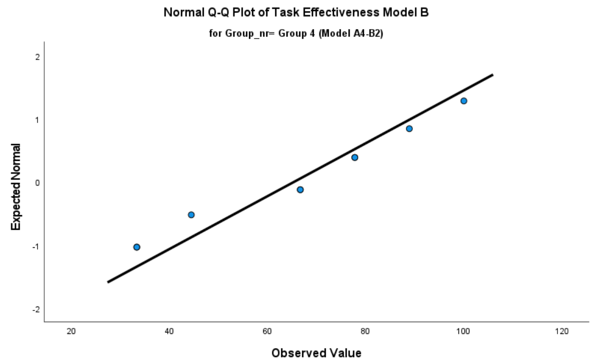




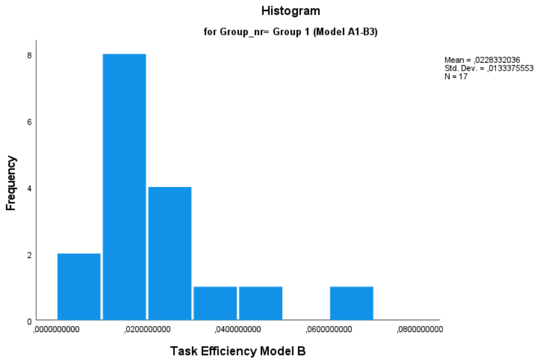


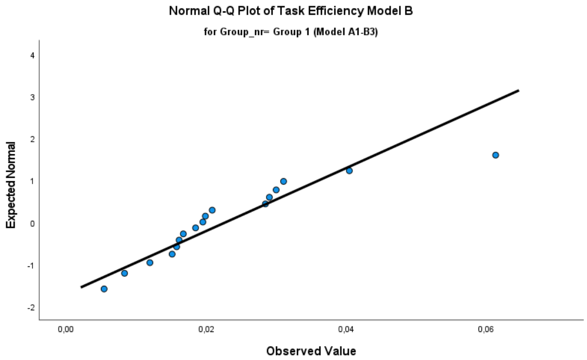


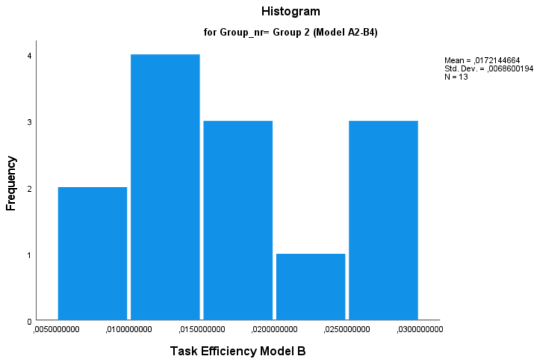


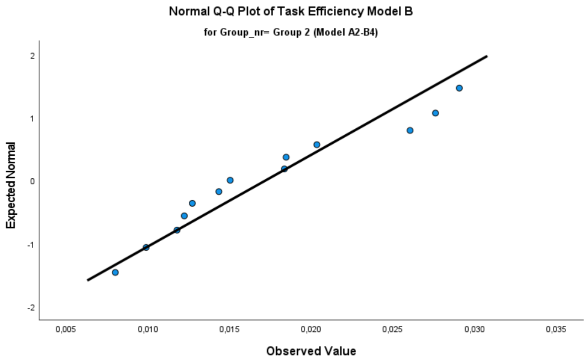


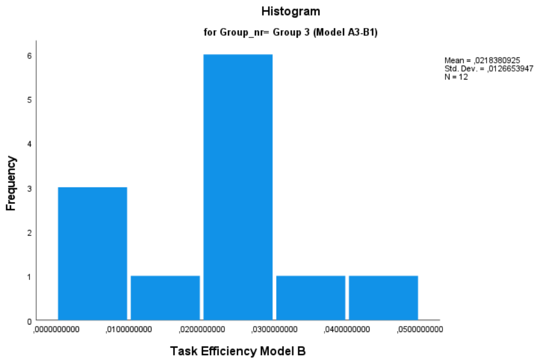
Model B – Task Efficiency

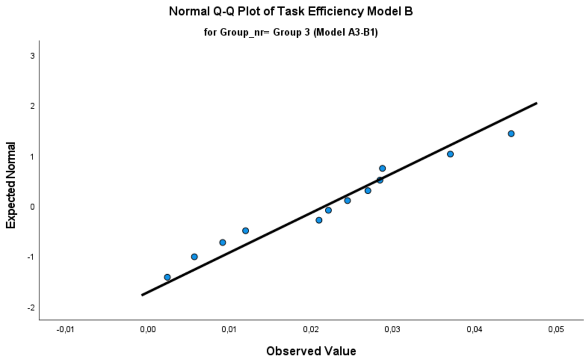


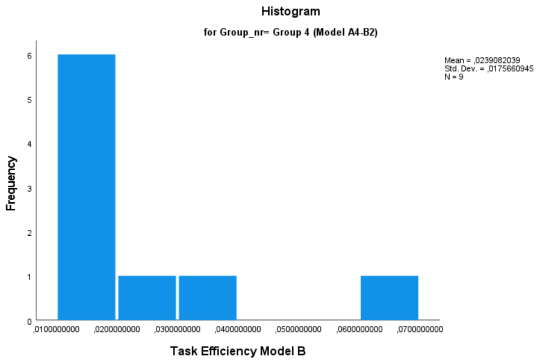


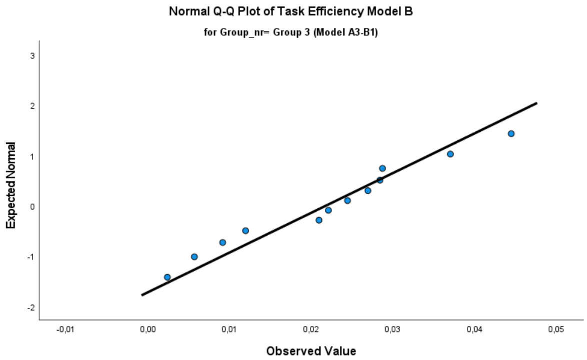




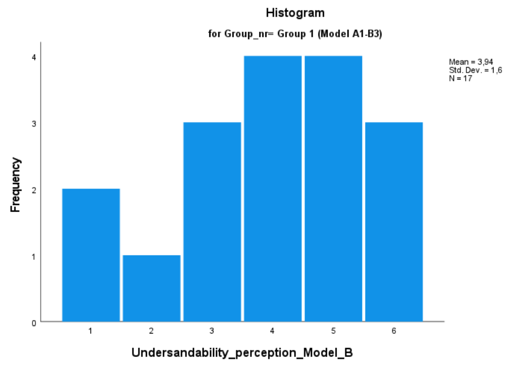


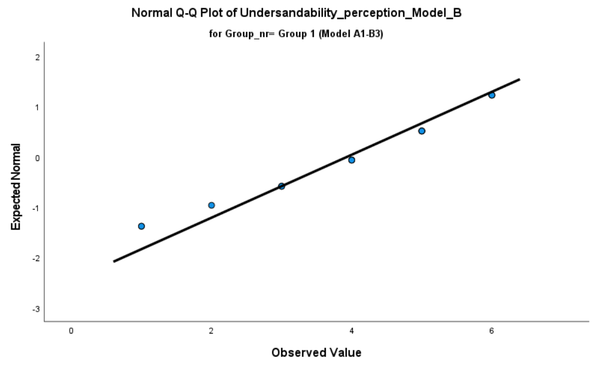


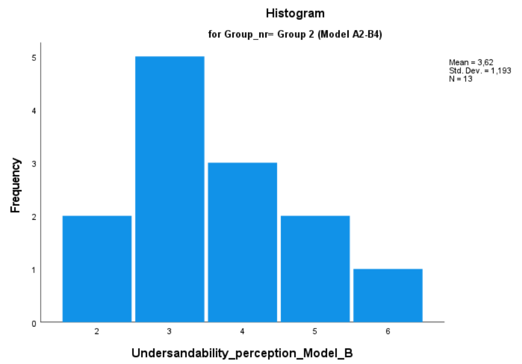


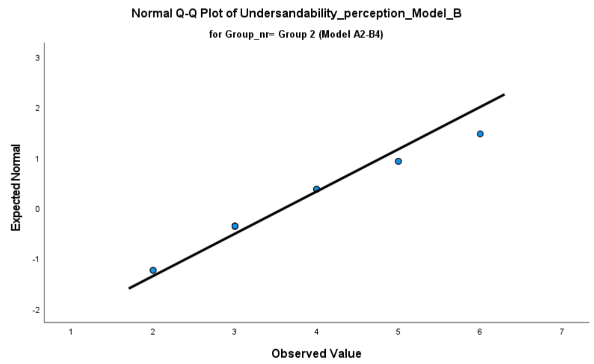


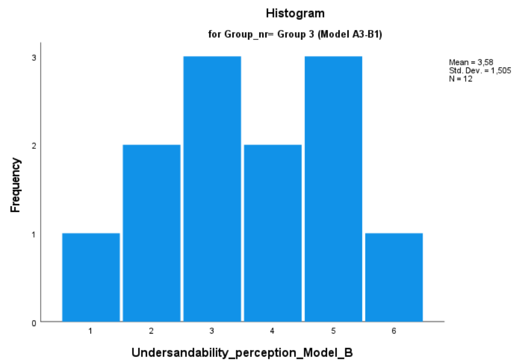
Model B – Perceived Understandability

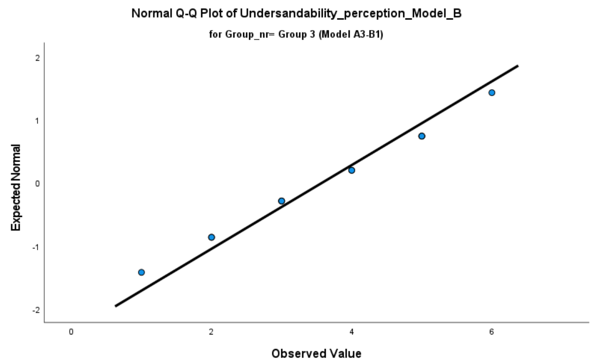


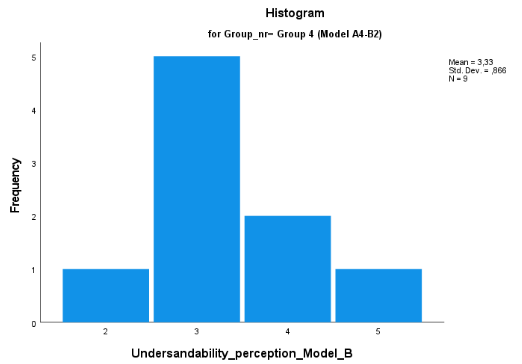


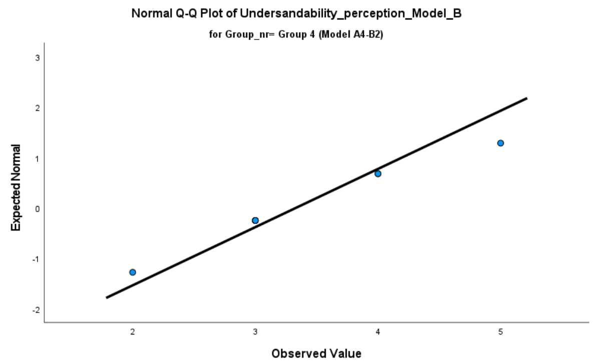












Appendix XIX – Outputs - IBM SPSS Statistics

Kruskal Wallis test for H1 - Model A

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.[a,b]	Decision
1	The distribution of Task Effectiveness Model A is the same across categories of MODEL_A_Version.	Independent-Samples Kruskal-Wallis Test	,736	Retain the null hypothesis.
2	The distribution of Task Efficiency Model A is the same across categories of MODEL_A_Version.	Independent-Samples Kruskal-Wallis Test	,966	Retain the null hypothesis.

a. The significance level is ,050.

b. Asymptotic significance is displayed.

Independent-Samples Kruskal-Wallis Test Summary

Total N	51
Test Statistic	1,270[a,b]
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	,736

a. The test statistic is adjusted for ties.

b. Multiple comparisons are not performed because the overall test does not show significant differences across samples.

Independent-Samples Kruskal-Wallis Test Summary

Total N	51
Test Statistic	,270[a,b]
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	,966

a. The test statistic is adjusted for ties.

b. Multiple comparisons are not performed because the overall test does not show significant differences across samples.

Kruskal Wallis test for H1 - Model B

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.[a,b]	Decision
1	The distribution of Task Effectiveness Model B is the same across categories of MODEL_B_Version.	Independent-Samples Kruskal-Wallis Test	,481	Retain the null hypothesis.
2	The distribution of Task Efficiency Model B is the same across categories of MODEL_B_Version.	Independent-Samples Kruskal-Wallis Test	,596	Retain the null hypothesis.

a. The significance level is ,050.

b. Asymptotic significance is displayed.

Independent-Samples Kruskal-Wallis Test Summary

Total N	51
Test Statistic	2,467[a,b]
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	,481

a. The test statistic is adjusted for ties.

b. Multiple comparisons are not performed because the overall test does not show significant differences across samples.

Independent-Samples Kruskal-Wallis Test Summary

Total N	51
Test Statistic	1,887[a,b]
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	,596

a. The test statistic is adjusted for ties.

b. Multiple comparisons are not performed because the overall test does not show significant differences across samples.

Kruskal Wallis test for H2 - Model A

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.[a,b]	Decision
1	The distribution of Undersandability_perception_Model_A is the same across categories of MODEL_A_Version.	Independent-Samples Kruskal-Wallis Test	,349	Retain the null hypothesis.

a. The significance level is ,050.

b. Asymptotic significance is displayed.

Independent-Samples Kruskal-Wallis Test Summary

Total N	51
Test Statistic	3,290[a,b]
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	,349

a. The test statistic is adjusted for ties.

b. Multiple comparisons are not performed because the overall test does not show significant differences across samples.

Kruskal Wallis test for H2 - Model B

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.[a,b]	Decision
1	The distribution of Undersandability_perception_Model_B is the same across categories of MODEL_B_Version.	Independent-Samples Kruskal-Wallis Test	,628	Retain the null hypothesis.

a. The significance level is ,050.

b. Asymptotic significance is displayed.

Independent-Samples Kruskal-Wallis Test Summary

Total N	51
Test Statistic	1,739[a,b]
Degree Of Freedom	3
Asymptotic Sig.(2-sided test)	,628

a. The test statistic is adjusted for ties.

b. Multiple comparisons are not performed because the overall test does not show significant differences across samples.

Appendix XX – Averages per question and variable

In Table 23 the average on Task Effectiveness per question is being shown. In Table 24 the same is shown but per question type. In Table 25 the average on Task Efficiency per question is being shown. In Table 26 the same is shown but per question type. Some questions seem not to measure correctly when comparing against the average in question type. I.e. model version A4 has the highest average in Task effectiveness in questions about concurrency (Q1 and Q2) and exclusiveness (Q3 and Q4), but model version A4 only have the highest average in Q1 and Q4. Model version A3 scored the highest in average in Q2 and model version A2 scores the highest in Q3. The same in the repetition questions (Q8 and Q9) where model version A1 scores the highest in average but in Q9 is model version A3 the one who scores the highest. Similar inconsistencies are found in the averages of task Effectiveness in Model B where Q1, Q3, and Q5 are not following the average per question type.

In Task Efficiency in model A are Q3 and Q9 not following the average per question type and in model B is Q9 the only question where the model version B3 with the highest average in Repetition questions (Q8 and Q9) didn't score the highest average.

Table 23 - Task Effectiveness per question nr. per model

	MODEL A Version					MODEL B Version				
Question	A1 (61% East)	A2 (43% East)	A3 (46% East)	A4 (30% South)	Total Model A	B1 (69% East)	B2 (41% East)	B3 (48% East)	B4 (47% South)	Total Model B
Q1	76,50%	69,20%	75,00%	88,90%	**76,50%**	83,30%	66,70%	88,20%	76,90%	**80,40%**
Q2	70,60%	69,20%	75,00%	66,70%	**70,60%**	83,30%	44,40%	64,70%	53,80%	**62,70%**
Q3	47,10%	69,20%	41,70%	55,60%	**52,90%**	75,00%	77,80%	76,50%	84,60%	**78,40%**
Q4	52,90%	30,80%	66,70%	66,70%	**52,90%**	83,30%	77,80%	64,70%	69,20%	**72,50%**
Q5	64,70%	53,80%	66,70%	55,60%	**60,80%**	75,00%	55,60%	64,70%	15,40%	**52,90%**
Q6	88,20%	69,20%	75,00%	66,70%	**76,50%**	58,30%	88,90%	64,70%	76,90%	**70,60%**
Q7	52,90%	53,80%	58,30%	22,20%	**49,00%**	41,70%	33,30%	29,40%	23,10%	**31,40%**
Q8	82,40%	61,50%	75,00%	55,60%	**70,60%**	91,70%	88,90%	82,40%	76,90%	**84,30%**
Q9	88,20%	84,60%	91,70%	88,90%	**88,20%**	83,30%	55,60%	70,60%	84,60%	**74,50%**
Total	**69,30%**	**62,40%**	**69,40%**	**63,00%**	**66,40%**	**75,00%**	**65,40%**	**67,30%**	**62,40%**	**67,50%**

Table 24 - Effectiveness per question type per model

	MODEL A Version					MODEL B Version				
Question	A1 (61% East)	A2 (43% East)	A3 (46% East)	A4 (30% South)	Total Model A	B1 (69% East)	B2 (41% East)	B3 (48% East)	B4 (47% South)	Total Model B
Concurrency	73,6%	69,2%	75,0%	77,8%	**73,55%**	83,3%	55,6%	76,5%	65,4%	**71,55%**
Exclusiveness	50,0%	50,0%	54,2%	61,2%	**52,90%**	79,2%	77,8%	70,6%	76,9%	**75,45%**
Order	76,5%	61,5%	70,9%	61,2%	**68,65%**	66,7%	72,3%	64,7%	46,2%	**61,75%**
Order count	52,9%	53,8%	58,3%	22,2%	**49,00%**	41,7%	33,3%	29,4%	23,1%	**31,40%**
Repetition	85,3%	73,1%	83,4%	72,3%	**79,40%**	87,5%	72,3%	76,5%	80,8%	**79,40%**

Table 25 - Efficiency per question nr. per model

Question	MODEL A Version					MODEL B Version				
	A1 (61% East)	A2 (43% East)	A3 (46% East)	A4 (30% South)	Total Model A	B1 (69% East)	B2 (41% East)	B3 (48% East)	B4 (47% South)	Total Model B
Q1	0,038	0,044	0,037	0,077	**0,046**	0,029	0,035	0,040	0,032	**0,034**
Q2	0,026	0,032	0,032	0,039	**0,031**	0,022	0,013	0,025	0,012	**0,019**
Q3	0,020	0,028	0,023	0,019	**0,023**	0,023	0,044	0,023	0,032	**0,029**
Q4	0,023	0,013	0,024	0,047	**0,025**	0,023	0,052	0,034	0,019	**0,031**
Q5	0,031	0,013	0,019	0,013	**0,020**	0,026	0,039	0,027	0,003	**0,023**
Q6	0,032	0,021	0,023	0,020	**0,025**	0,021	0,032	0,028	0,023	**0,026**
Q7	0,016	0,019	0,019	0,007	**0,016**	0,008	0,015	0,012	0,015	**0,012**
Q8	0,046	0,033	0,036	0,054	**0,042**	0,053	0,040	0,056	0,041	**0,049**
Q9	0,034	0,043	0,056	0,049	**0,044**	0,027	0,019	0,026	0,032	**0,027**
Total	**0,030**	**0,027**	**0,030**	**0,036**	0,030	**0,026**	**0,032**	**0,030**	**0,023**	**0,028**

Table 26 - Efficiency per question type per model

Question	MODEL A Version					MODEL B Version				
	A1 (61% East)	A2 (43% East)	A3 (46% East)	A4 (30% South)	Total Model A	B1 (69% East)	B2 (41% East)	B3 (48% East)	B4 (47% South)	Total Model B
Concurrency	0,032	0,038	0,035	0,058	**0,0385**	0,026	0,024	0,033	0,022	**0,0265**
Exclusiveness	0,022	0,021	0,024	0,033	**0,024**	0,023	0,048	0,029	0,026	**0,03**
Order	0,032	0,017	0,021	0,017	**0,0225**	0,024	0,036	0,028	0,013	**0,0245**
Order count	0,016	0,019	0,019	0,007	**0,016**	0,008	0,015	0,012	0,015	**0,012**
Repetition	0,040	0,038	0,046	0,052	**0,043**	0,040	0,030	0,041	0,037	**0,038**