On deriving business models from process models: An empirical study

Abstract

At least two business requirement perspectives of the digital ecosystem should be revisited in case such an ecosystem changes significantly: (1) the business process perspective (e.g. represented by a BPMN 2.0 model), and (2) the business value perspective (e.g. depicted by an e^{3}value *model). Although both perspectives differ largely and address different stakeholder concerns, there is also overlap between the two points of view. Moreover, often there is already an explicit understanding of how the parties in the ecosystem execute processes. However, the business value model is, in many cases, left implicit, whereas most disruptive technologies results in significant changes in the business value model. It is important to understand and analyze these using a model-based approach like* e^{3}value. *To speed up the elicitation process, it would be more efficient to elicit an* e^{3}value *value model using an already existing process model. We test a series of guidelines to derive an* e^{3}value *model from a given BPMN model. We conducted a controlled experiment through which we analyze the quality of a conceptual model –* e^{3}value *business value (*e^{3}value*) – derived from another conceptual model – Business Process Notation v.2.0 (BPMN). We measure model quality via validity and completeness with respect to a normative standard solution and an expert solution. Subjects are divided into two groups: the treatment group, which use the guidelines to derive one model (*e^{3}value*) from the other (BPMN), and the control group, which do not use the guidelines. Furthermore, we analyze and evaluate the implications of the experiments results to understand the limitations and to improve them in future research.*

Keywords: Ecosystems, Business models, Process models, *e^{3}value* model, BPMN

1. Introduction

Developing an ICT-enabled ecosystem requires many requirement-oriented viewpoints. Two of those viewpoints are the business model perspective and the cross-organizational business process, each of which has its own concerns. We use here the *e^{3}value* (Gordijn and Wieringa, 2021) language for the business model perspective and the BPMN 2.0 (OMG, 2011) notation for the cross-organizational process perspective. Despite the overlap between both, there are also substantial differences. To mention a few, *e^{3}value* has the notion of economic reciprocity and commercial bundling on the supplier and customer side. (see for more differences Gordijn et al., 2000). These concepts are not present in BPMN. Conversely, BPMN represents the concept of time ordering in which activities take place, whereas *e^{3}value* represents only causal dependencies.

Since there are overlaps, it might be possible to derive (partially) one model from the other. In case of digital ecosystem development, we often start with the design of the *e^{3}value* model and derive the corresponding BPMN. However, many ecosystem development projects are about *changes* to an already up-and-running ecosystem, e.g., to optimize such a system, or to use new chances as a result of technological innovations. Due to the increasing popularity of business process (re)design and analysis, many ecosystems have already a representation of a cross-organizational business process. We assume here that the process model is created using BPMN, but there is no principled argument why our approach could not be useful in conjunction with other notations for business processes. Since in many situations process models already exist, but the corresponding business (value) models do not yet, the question is whether we can use the existing process model(s) to derive the corresponding business model described by the *e^{3}value* language. *Guidelines* can help and speed up the process of an *e^{3}value* model, by using the already elicited and modelled requirements in the BPMN model.

Concretely, we want to use the information expressed in cross-organization business process models (BPMN) to design a corresponding e^3*value* model capturing the business value concerns. This is possible to some extent, as both techniques have some conceptual overlap. However, our experience has been that reusing BPMN models to construct a e^3*value* value is difficult to automate, as there is also a significant conceptual difference between the two techniques. Nevertheless, we expect to a certain extent reuse is possible by means of *a set of guidelines* (da Silva Torres et al., 2021).

We investigate the model quality generated by these *guidelines* in terms of how well they support modelers in extracting the business models from existing BPMN models. Our research questions in this paper are as follows:

RQ 1. How does the quality of a derived e^3*value* model from a BPMN model is affected when using a set of guidelines?

RQ 2. Are there other factors affecting the quality of the model derived by the guidelines?

We performed a controlled experiment. Subjects were briefed to derive an e^3*value* model starting from a BPMN model with and without the help of a set of guidelines. By defining a normative and an expert model, we measure the completeness and validity of the subjects' derived models. Our research not only provides further evidence on the factors affecting the derivation of e^3*value* models from BPMN models, our results also inspire other research on the effectiveness of alternative requirements notations and requirement engineering techniques for different requirements-related tasks.

The paper is distributed as it follows. In Sec. 2, we inform the literature gap and discuss the related work. In Sec. 3, we shortly present the controlled experiment. In Sec. 3.3, we interpret and discuss the results of the experiment. We conclude and set plans for future research in Sec. 4.

2. Background and related work

BPMN is a well-known language in conceptual, for which we assume the reader is familiar with, but to make the paper self-contained, we briefly introduce the e^3*value* method. For a more detailed explanation see Gordijn and Wieringa, 2021.

2.1. e^3 value

The e^3 value language (Gordijn and Wieringa, 2021) is an approach in the field of value. Other approaches include the Resource Event Agent (REA) (McCarthy, 1982) ontology and value stream mapping (Hines and Rich, 1997). E^3value is particularly powerful in representing and analyzing the ecosystem or, as it is called in E^3value, the networked value constellation (Normann and Ramırez, 1994) of enterprises that collectively satisfy one or more needs of an end-user.

We briefly explain, by means of an educational example (Fig. 1) the e^3 value constructs relevant for this paper. *Actors* such as Amazon.com are profit-and-loss responsible and often legal entities. In many cases, it is useful to talk about multiple actors of the same kind, *market segments*, who assign economic value in the same way. Examples are readers (people who want to read a book) and publishers. Actors and market segments do *value transfer*s which each other; the subject of such a transfer is the *value object* (e.g., book, transportation, money), that has economic value for at least one of the actors/market segments. The latter transfer value object via *value ports*, which are grouped into *value interfaces*. These interfaces represent economic reciprocity; hence a value interface contains at least one *in-going* value port and one *out-going* value port. Actors/market segments perform *value activities* to earn money (companies) or increase economic utility (end-users). *Customer needs* (read book) indicate a state of felt deprivation by an actor that is satisfied by one or more value objects represented by *dependency connection elements*. At the end of the dependency chain, there are one or more *boundary elements* to show that further value transfers are not considered anymore. This does not imply they are not there (e.g., the publisher needs to do transfers with writers); they are only considered out-of-scope for the model. Hence, boundary elements specify the model boundary.

2.2. Related work

A number of researchers have paid attention to the derivation of process models from value models and vice versa. Also, work has been done to relate e^3*value* to the Resource Event Agent (REA) framework (McCarthy, 1982; Schuster and Motal, 2009), which was also the basis for a mapping to UN/CEFACT Methodology (UMM) (Hofreiter et al., 2006; Schuster, 2010) models. In previous work (da Silva Torres et al., 2021; da Silva Torres et al., 2020), we have proposed a set of guidelines to derive an e^3*value* model by using

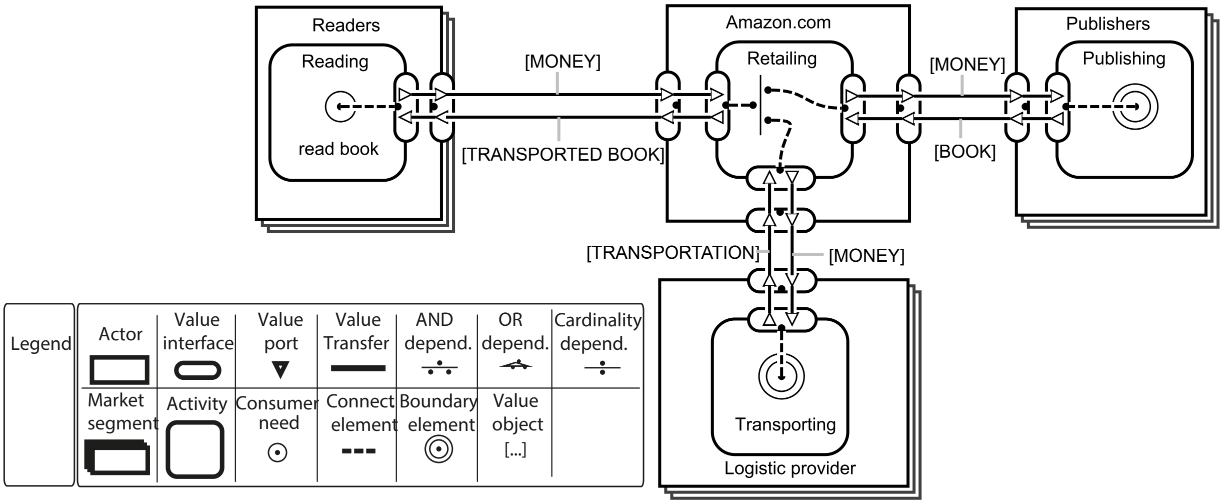

Figure 1. An educational e^3 value model

a BPMN model as input. The point of departure is the fundamental ontological difference between value and process models, e.g. explored in (Gordijn et al., 2000). Follow-up research investigated this in more detail, e.g. by (Bodenstaff, 2010), where formal consistency rules are defined between value and process models.

We argue that the two models are too different to allow for such automatic translation (see, e.g., (Gordijn et al., 2000) for important differences). Therefore, we propose a design-oriented approach, e.g., in (Hotie and Gordijn, 2019), we have presented a method using intermediate models to derive a BMPN model from an *e^{3}value* model. As for empirical testing of this kind of derivation, similar work has been done when comparing two different conceptual models, such as: user stories and use cases (Dalpiaz et al., 2021a; Dalpiaz and Sturm, 2020); BPMN against textual use cases (Ottensooser et al., 2012); textual, semi-structured, diagrammatic (Hoisl et al., 2014). As we are using subjects (students) i.e. users; we take into account System Quality (SQ) which considers desired characteristics of a system/tool (i.e. our guidelines) that produces the information output which relates to the perceived ease of understanding and the perceived usefulness of the tool to the user (Maes and Poels, 2006).

3. The controlled experiment

To answer the research questions, we conducted a controlled experiment in which we analyzed the quality of a conceptual model – *e^{3}value* business model (*e^{3}value*) – derived from another conceptual model – Business Process Notation v.2.0 (BPMN) – using two different methods: with the guidelines (GL) and without the guidelines (WG); e.g. simply using common sense. The guidelines are summarized in Table 1. We measure model quality via validity and completeness with respect to a normative standard solution that we created ourselves in an extensive elicitation process with the case study providers, and a solution created by a BPMN *e^{3}value* expert using the guidelines. The test subjects use the guidelines to derive one model (*e^{3}value*) from the other (BPMN). In addition, we collect and compare the experience of the test subjects in terms of the usefulness of the tool for the tasks performed.

The starting point of our research is the following null hypothesis, in which we do not assume a difference exists between the methods:

H_0: *deriving an* e^{3}value *model based on a BPMN process model using the guidelines or not are equally good.*

3.1. Experiment design

3.1.1. Independent variables. The first variable is the treatment (IV1) according to which the method used is specified. It has two possible values: with guidelines (GL) and without guidelines (WG). The second independent variable is the case study used (IV2). It also has two possible values: Music Case (MC) and Fintech Bank Case (FC). This setting is adapted from (Dalpiaz et al., 2021b).

3.1.2. Dependent variables. There are two dependent variables we use for measuring the quality of the generated conceptual models. These variables are specified by comparing the elements in the subject

Table 1. Guidelines – from BPMN model to e^{3}value model

ID	BPMN element	e^3*value* element	Guideline description
G1	Start event, End event	Consumer need, Boundary element	BPMN start/end events may correspond to e^3*value* consumer needs and boundary elements.
G2	Pool	Actor, Market segment	BPMN pools may correspond to e^3*value* actors or market segments.
G3	Lane 1, Lane 2	Value activity	BPMN lanes may correspond to e^3*value* value activities.
G4	Activity, Subprocess	Value activity	BPMN activities and sub-processes may correspond to e^3*value* value activities.
G5	Message flow	Value transfer	BPMN message flows may correspond to e^3*value* value transfers.
G6	Activity, Sequence flow, Subprocess	Value transfer	BPMN activities and sub-processes and their sequence flows may correspond to e^3*value* value transfers.
G7	Message flow, Sequence flow	Value interface	Following a BPMN sequence/message flow may lead to an e^3*value* value interface.
G8	Message flow, Sequence flow	Value offering	Following a BPMN sequence/message flow may lead to an e^3*value* value offerings.
G9	Message flow, Sequence flow	Dependency path	Following a BPMN sequence/message flow may lead to an e^3*value* dependency path.
G10	AND Gateway	AND dependency	BPMN AND gateways may correspond to e^3*value* AND dependencies.
G11	XOR Gateway	OR dependency	BPMN XOR gateways may correspond to e^3*value* OR dependencies.
G12	OR Gateway	AND dependency, OR dependency	BPMN OR gateways may correspond to a combination of e^3*value* AND/OR dependencies.
G13	Task 2, Task 3, Loop	Cardinality dependency	BPMN loops may correspond to e^3*value* cardinality dependencies.

solution (i.e. the conceptual model derived by a subject, which in our case are the students from our courses) against the normative solution (the conceptual model derived by interviewing the domain experts) and, the expert solution (the conceptual model derived by two conceptual modelers specialized in BPMN and e^3*value*):

- *Validity (DV1)*: the relation of elements in a student model (subject solution) that also appear in the expert model or the normative model, over the number of elements that are correctly or incorrectly represented in the student model. In other terms, validity equates to preciseness.

$$Validity(precision) = \frac{|TP|}{|TP| + |FP|}, \quad (1)$$

where TP = True Positives, and FP = False Positives.

- *Completeness (DV2)*: the relation of elements in a student model correctly or incorrectly represented, over the number of elements in the expert model or the normative model. In other terms, completeness equates to recall.

$$Completeness(recall) = \frac{|TP|}{|TP| + |FN|}, \quad (2)$$

where TP = True Positives, and FN = False Negatives.

Given the number of students, experts and normative models, the measurements in Table 2 are used to calculate validity *DV1* and completeness *DV2* as follows:

$$Validity(precision) = \frac{|AL|}{|AL + SO| + |WR|} \quad (3)$$

Table 2. Measurement to calculate DV1 and DV2. Adapted from (Dalpiaz et al., 2021b)

State Analyzed	Description
Aligned (AL)	An element is represented in both models, either with the same name or using synonyms or clearly connectable names (e.g. “client” instead of “customer”).
Wrongly represented (WR)	An element in the expert model or the normative model is incorrectly represented in the student model, either (i) via different groups, elements, or (ii) using a generic term (e.g., “fee” instead of “service”).
Service-oriented (SO)	An element in any model that denotes a service delivery aspect, e.g., payment service. Elements that represent this kind of transaction are considered.
Omitted (OM)	Elements in the expert model or the normative model that does not appear in any way in the student model.
Missing (MI)	Elements in the student model that does not appear in any way in the expert model or normative model.

$$Completeness(recall) = \frac{|AL + WR|}{|AL| + |WR + OM|} \quad (4)$$

Although we count missing elements, we do not invest our metrics on them. Since them, most of the times, represent something the subjects decided to put in the model thinking it would make it more complete.

3.1.3. Subjects. We involved master students taking the course on Ecosystems Disrupted by IT at Vrije Universiteit Amsterdam (VU), and also third year bachelor students taking the course on Business Process Engineering at Universidad Rey Juan Carlos (URJC) in Madrid. The course Ecosystems Disrupted by IT at VU teaches how to analyze, design, and implement various techniques in conceptual inside ecosystems that suffer drastic changes due to technology. In the course, the students-subjects learned the notion of and, in particular,e^3*value* and BPMN. The instructors of the course was the first and fifth author of this paper. They learned e^3*value* and BPMN to have a more holistic view of complex ecosystems. They also practiced e^3*value* and BPMN through homework assignments, in which they achieved good results, showing they understood the notations wellv. The course Business Process Engineering at URJC teaches how to discover, identify, analyse, redesign, implement and monitor processes in a business context. The second and fourth author of this paper were the instructors. They learn how to model and analyze BPMN models. They were instructed in different tools for , analyzing and simulating process models. They practice in class and were graded via assignments and an exam. They also receive instruction on e^3*value* constructs and .

The test subjects are not precisely the same as the intended user base of the guidelines (e.g. consultants). Nevertheless, we argue that is acceptable to use students for the experiments for a number of reasons. First, they have extensively trained on BPMN and e^3*value* . This is difficult to accomplish with consultants, as it would take a serious amount of time of them. Second, it allows to have quite a number of test subjects, which is more difficult to reach with consultants. And finally, because two universities in two different EU countries (The Netherlands and Spain) are involved, it allows to replicate the experiment.

3.1.4. Task. We designed the experiment so part of the subjects would experience the derivation through the treatment (with the guidelines - GL) and the other without it (WG). For that purpose, we designed two forms in which we alternate the treatment and the case study. The form had 4 parts: (1) a pre-test questionnaire that checks the subjects' background and knowledge (on BPMN and e^3*value*); (2) the first task, in which subjects receive the requirements of the Music ecosystem case or the Fintech ecosystem case, specified in BPMN, and were asked to derive an e^3*value* model using or not the guidelines. In order to accomplish this, we consider two real-world scenarios: the financial securities trading and the Intellectual Property Rights (IPR) music clearance. To understand and model the problem in both domains – financial securities trading and International IPR clearing – we have worked with persons affiliated with the Dutch National Bank (De Nederlandsche Bank – DNB) and the IPR society NL (SENA; exploitation on neighboring rights), respectively; (4) a post-test with questions about the subjects' perception of the guidelines and its usefulness. The latter was only part of the groups that applied the guidelines.

3.1.5. Execution. The experiment took place in a dedicated time slot of two hours (2hr). All subjects, however, finished in one hour and half minutes (1h30m). The distribution of the cases as well as the groups participants (39) were done randomly. Since there were students from two different courses, we needed to give an extra lecture on e^3*value* to the students of Universidad Rey Juan Carlos, because the focus of their course was mainly BPMN . A two hours (2hr) lecture was given regarding the main e^3*value* concepts and . The authors of this paper played the role of the model experts, one per case. The students were then given the pre-test, most of the students finished their knowledge test in twenty (20) minutes. After they were randomly allocated in groups of 4-6 persons and given the specifications (in BPMN) of each case (music case - MC; and fintech case - FC), along with the guidelines. In total six (6) groups derived the e^3*value* model with the guidelines (GL) and, another 6 groups derived the e^3*value* model whithout it (WG). This took around an hour, with that concluded they handed their derived models to the researchers. The GL groups received at the end the post-test which they finished in less than ten (10) minutes.

We analyzed the collected data. We created the stakeholders model (normative model) i.e. the model created interviewing the domain experts of each case; starting from the specifications and without accessing the students' e^3*value* diagrams. This activity required some iterations with the domain experts to ensure uniformity across the researchers. Then, after an informal analysis of the student models, we compared the models: *student model* vs *normative model*; and *student model* vs *expert model*, using the measurements listed in Table 2, to populate the spreadsheet. The comparison between the *normative model* vs *expert model* was already object of our analysis in previous research (da Silva Torres et al., 2020).

That spreadsheet was then used to execute most of the statistical analyses. We use T-Tests to analyze the significance of validity and completeness when splitting the data by each independent variable (IV1 and IV2).

3.2. Experiment results

Prior to the data analysis, we checked the subjects pre-test results and found them with similar competencies, though there was a difference in their distribution, we address this on Sec. 3.4.

We first analyzed the data when comparing the subjects' models with the normative model. Table 3 presents descriptive statistics regarding this comparison. Each row represents the outcomes of a single subject's group, showing the treatment used (IV1), the case applied (IV2), the size ratio (i.e., the number of elements within the student model divided by the number of elements within the normative or expert model), the model validity (DV1), and the model completeness (DV2). A relevant initial observation is all the models created by the students were smaller in size to those created by the domain experts (normative model) and the model experts (expert model), since the size ratio is < 1.

The average size ratio from GL is way greater than WG. Although this metric only shows relative size, without focusing on the alignment between student and normative models, we can observe some differences. In particular, we see how the music case (IV2) seems the most challenging to fully represent: many students had a considerably lower number of elements than in the corresponding normative models.

Another interesting data from the comparison between the student model and the normative model is validity (DV1) and completeness (DV2) had way higher average when GL were used than with WG. This can be better seen in Fig. 2 with respect to both validity (0.81) and completeness (0.75) to GL and, validity (0.34) and completeness (0.42) to WG. These differences show to have an e^3*value* model that represents the domain with quality, the guidelines would show better results than without it.

We then compared the students' with the experts' models. The results were even better than the previous comparison, see Table 4. The model quality achieved an average of 0.90 on validity and 0.93 on completeness when using the guidelines. In contrast, an average of 0.34 and 0.42 on validity and completeness, respectively, for WG. The results show the student model is more similar to the expert model than the normative model. This relates to how the expert model uses the modeler understanding of the domain to draft the model the same way the students had to do, which does not happen when during an interview with the stakeholders. The level of subjectivity is lower when with domain experts, which reflects on the quality of the derived models.

When separating the quality per case (IV2), the MC was surprisingly more complete than FC with a difference of only (0.06). This was unexpected since the MC is more complex than the FC. However, both cases received the same score for validity (0.90). In Fig. 3 we can see how the treatment (IV1) relates to each dependent variable (DV1 and DV2).

In all metrics in both case studies (IV2), the conceptual models derived from the set of guidelines (GL) outperforms the models derived without using it (WG), as it can be seen in Table 5. Based on the results, we can conclude we can reject the hypotheses on the equality of both treatments (IV1), since the quality of the derived conceptual model, in particular with respect

Table 3. Results from comparing students model against the normative model

Group	IV1: Treatment	IV2: Case	AL	WR	SO	OM	MI	Size Ratio	DV1: Validity	DV2: Completeness
1	GL	MC	28	5	1	10	20	0.52	0.82	0.77
2	GL	MC	26	5	0	12	20	0.49	0.84	0.72
3	GL	MC	23	3	1	13	24	0.41	0.85	0.67
4	GL	FC	19	4	1	8	19	0.46	0.79	0.74
5	GL	FC	20	5	0	6	19	0.50	0.80	0.81
6	GL	FC	18	4	1	5	23	0.44	0.78	0.81
7	WG	MC	5	11	2	14	33	0.25	0.28	0.53
8	WG	MC	3	10	3	15	35	0.21	0.19	0.46
9	WG	MC	4	3	0	13	43	0.11	0.57	0.35
10	WG	FC	3	6	1	13	30	0.18	0.30	0.41
11	WG	FC	2	3	1	8	37	0.10	0.33	0.38
12	WG	FC	3	5	0	12	31	0.16	0.38	0.40

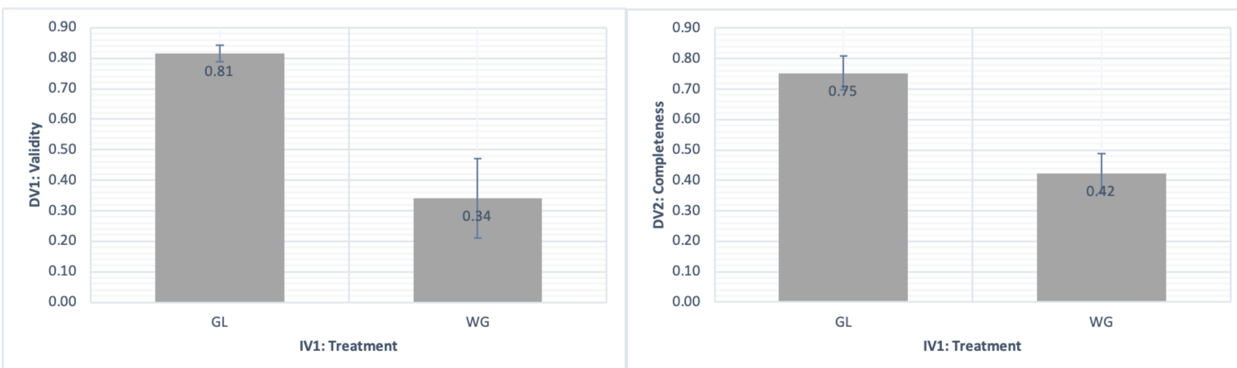

Figure 2. Validity and completeness in relation to students model vs normative model

Table 4. Results from comparing student model against the expert model

Group	IV1: Treatment	IV2: Case	AL	WR	SO	OM	MI	Size Ratio	DV1: Validity	DV2: Completeness
1	GL	MC	34	2	1	3	10	0.72	0.92	0.92
2	GL	MC	32	3	0	2	2	0.70	0.91	0.95
3	GL	MC	29	4	0	3	2	0.66	0.88	0.92
4	GL	FC	24	4	1	4	2	0.70	0.83	0.88
5	GL	FC	23	1	0	5	3	0.60	0.96	0.83
6	GL	FC	23	2	0	3	3	0.63	0.92	0.89
7	WG	MC	6	1	2	11	3	0.14	0.67	0.39
8	WG	MC	5	4	0	7	2	0.18	0.56	0.56
9	WG	MC	4	3	0	9	12	0.14	0.57	0.44
10	WG	FC	3	2	1	13	5	0.13	0.50	0.28
11	WG	FC	2	3	1	8	8	0.13	0.33	0.38
12	WG	FC	3	4	0	6	9	0.18	0.43	0.54

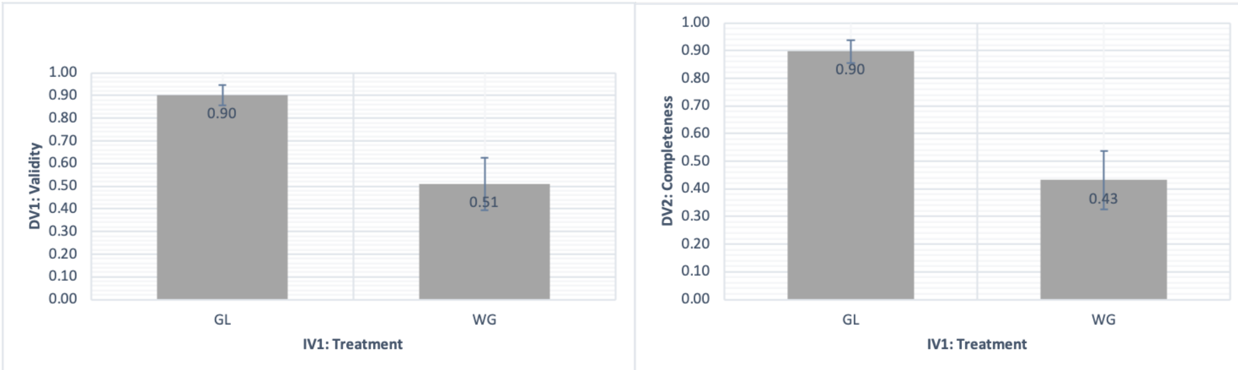

Figure 3. Validity and completeness in relation to students model vs expert model

to validity (DV1) and completeness (DV2), the data is significantly different ($p < 0.05$). For the other tasks, there was a consensus regarding the benefits of using the guidelines to derive the models as well.

3.3. Discussion

The conclusion we can draw from the experiment is the treatment (IV1) seems to affect the conceptual model derivation, since the quality of the derived models are drastically inferior when not using the guidelines. This conclusion, however, needs to be interpreted according to other factors, as we describe in the following.

The complexity of the case studies and the case itself (IV2) does not seem to significantly affect the results. Thus, when introducing equivalent requirements in different case, we expect an modeler would be able to derive an e^3*value* model of comparable quality. The FC study was less complex than the MC study (50 vs. 63 elements). The conceptual models derived fit better the expert model solution. For MC, the complexity is due to some key factors: number of elements, the introduction of a bank service to each transaction (bank service fee), the multiple interactions among the actors and market segments, and the need to duplicate some actors and market segments. The results may also be affected by the courses environment: since they were taught by different instructors in different countries, there might be some impact in the learning outcome.

When referring to the qualitative aspects of the results, we used the post-test as base to derive our conclusions. Most of the students when asked what was the best aspect of the guidelines answered with: ‘It serves as a good starting point’ or ‘It made the derivation more tangible’ . Also, when asked what should be improved answered with: ‘Examples of use’ or ‘how each guidelines work through concrete examples would be better than text’. This led us to believe a semi-automated tool that guides the user could be done. Showing exactly which derivation process can be selected, if even possible.

The subjects also perceived the BPMN was complex enough which led them to believe their derivation was not right due to differences in size from one model to the other. This is caused because e^3*value* models only elucidates value whereas a process model in BPMN is a collection of activities which sometimes does not add value from one to the other. The level of granularity or deescalation was reported by the subjects.

On the other hand, the treatment (IV1) seems to have a stronger impact on the derivation of the e^3*value* model, as visible by the results in Fig. 2 and Fig. 3. Our study showed, in particular, models created usinn the guidelines (GL) led to significantly higher validity (DV1) and completeness (DV2) than the one that did not use (WG).

The research reveals multiple factors may affect the quality of the derived conceptual model. This leads to further hypotheses:

H1: Automated tooling that implements guided derivation helps to derive more complete conceptual models.

H2: The model’s size affects the quality of the derived conceptual model.

H3: The model’s style affects the quality of the derived conceptual model.

Which should be covered in subsequent research.

3.4. Threats to validity

This section details how our results need consider several threats to validity (Wohlin et al., 2012).

Construct validity. The selection of the real-world case scenarios may affect the results. We used domains that were accessible to us and not necessarily that would

Table 5. Significance for DV1 and DV2 per treatment (IV1)

Comparison type	Measure	$\overline{GL}$	σ	$\overline{WG}$	σ	p-value
Student Model vs. Expert Model	Validity	0.90	0.04	0.51	0.12	0.002%
Student Model vs. Expert Model	Completeness	0.90	0.04	0.43	0.11	0.0002%
Student Model vs. Normative Model	Validity	0.81	0.03	0.34	0.13	0.001%
Student Model vs. Normative Model	Completeness	0.75	0.06	0.42	0.07	0.0003%

be easy to understand by the subjects. Our input used BPMN 2.0, which is one of the most used languages for process , and although we know there are other process notations, we used BPMN due to familiarity. The models used as normative models were created much earlier than the experiment execution, which might lead to some outdated data.

Internal validity. We mitigate external factors that would impact in the dependent variables (such as domain knowledge, willingly commitment by the subjects, and the training level of the subjects to perform the tasks). The subjects had no prior knowledge to neither domains, and consequently, were not affected. The pre-test done to test their knowledge should eliminate various kinds of external factors. Even though the experiment was done on a voluntary basis, the subjects from the course at URJC in Madrid were told they would earn bonus points upon participation and the subjects from the course at VU Amsterdam were offered gift cards containing 35 euros. This was done to increase the motivation and commitment of the subjects. Moreover, although we tried to make sure the real-world scenarios had similar complexity, the MC was larger than the FC. Another threat concerns the order of presentation of the domains within the experiment, which would also affect the results. However, they were randomly distributed to avoid that. We already mentioned that students are not precisely the same as the intended user base of the guidelines (consultants), and argued that we consider the use of students as a reasonable choice to do the experiment. Lastly, another threat would be tiresome (fatigue), but the experiment was relatively short. In fact, all subjects delivered their outcomes way before the 2 hours of experiment, which excludes this threat.

Conclusion validity. We assumed and inferred on top of our statistical tests (such as the significance of the data) when analyzing the results. Our solutions were defined prior to the experiment, which lead to our grading system.

External validity. Our subjects were chosen from courses that had affinity with the tasks at hand. The subjects were bachelor and master students with limited experience in business and process . Caution should be taken when generalizing the results.

4. Conclusion

We conducted a controlled experiment to investigate the factors that potentially affect the quality of a derived conceptual model, in specific an e^3*value* model derive from a BPMN model. The experiment had 39 subjects, they were distributed in 12 groups. Each group were required to perform the derivation with or without the guidelines.

The analysis of the results shows the case seem to have a limited impact on the quality of the derived conceptual model, both with respect to validity and completeness. The most influential factor is the adopted treatment process. Furthermore, our results also show more complex domains or ecosystems lead to derived models of higher quality (RQ1), which corroborate with (RQ2) and elucidates other factors affect the quality on derived models, and should be topic for further research.

The results got call for further experimentation with the guidelines in other domains. The research community needs to build a corpus of evidence to assist practitioners in choosing and notations and techniques.

Acknowledgements. The authors are thankful to all the students that participated in this experiment setting.

References

Bodenstaff, L. (2010). *Managing dependency relations in inter-organizational models* (Doctoral dissertation). University of Twente.

Dalpiaz, F., Gieske, P., & Sturm, A. (2021a). On deriving conceptual models from user requirements: An empirical study. *Inform. and Softw. Techn.*, *131*, 106484.

Dalpiaz, F., Gieske, P., & Sturm, A. (2021b). On deriving conceptual models from user requirements: An empirical study. *Information and Software Technology*, *131*, 106484. https://doi.org/https://doi.org/10.1016/j.infsof.2020.106484

Dalpiaz, F., & Sturm, A. (2020). Conceptualizing requirements using user stories and use cases: A controlled experiment. *International Working Conference on Requirements Engineering: Foundation for Software Quality*, 221–238.

da Silva Torres, I., Fantinato, M., Branco, G. M., & Gordijn, J. (2021). Design guidelines to derive an e3value business model from a bpmn process model in the financial securities sector. *Proceedings of PoEM 2021 : IFIP Working Conference on The Practice of Enterprise Modeling*. https : / / dise - lab . nl / wp - content / uploads / 2021 / 11 / POEM2021FINTECHproject.pdf

da Silva Torres, I., Gordijn, J., Fantinato, M., & da Fountoura Vieira, J. F. (2020). Designing an ecosystem value model based on a process model – An empirical approach. *13th IFIP Working Conf. on The Practice of Enterprise Modeling*.

Gordijn, J., Akkermans, H., & Vliet, H. V. (2000). Business modelling is not process modelling. *Conceptual Modeling for E-Business and the Web, ECOMO 2000*, *1921*(40-51).

Gordijn, J., & Wieringa, R. (2021, February 12). *E3value user guide - designing your ecosystem in a digital world* (1st). The Value Engineers.

Hines, P., & Rich, N. (1997). The seven value stream mapping tools. *Int'l J. of Operations & Production Management*, *17*(1), 46–64.

Hofreiter, B., Huemer, C., Liegl, P., Schuster, R., & Zapletal, M. (2006). UN/CEFACT'S modeling methodology (UMM): A UML profile for B2B e-commerce. *2nd Int'l Works. on Best Practices of UML*, 19–31.

Hoisl, B., Sobernig, S., & Strembeck, M. (2014). Comparing three notations for defining scenario-based model tests: A controlled experiment. *2014 9th International Conference on the Quality of Information and Communications Technology*, 180–189.

Hotie, F., & Gordijn, J. (2019). Value-based process model design. *Bus. Inf. Syst. Eng.*, *61*(2), 163–180.

Maes, A., & Poels, G. (2006). Evaluating quality of conceptual models based on user perceptions. *4215*, 54–67. https : / / doi . org / 10 . 1007 / 11901181_6

McCarthy, W. E. (1982). The REA accounting model: A generalized framework for accounting systems in a shared data environment. *The Accounting Review*, *58*(3), 554–578.

Normann, R., & Ramırez, R. (1994). *Designing interactive strategy – From value chain to value constellation*. Wiley.

OMG. (2011). Business process model and notation, version 2.0 [Object Management Group (OMG)]. https://www.omg.org/spec/BPMN/2.0

Ottensooser, A., Fekete, A., Reijers, H. A., Mendling, J., & Menictas, C. (2012). Making sense of business process descriptions: An experimental comparison of graphical and textual notations. *Journal of Systems and Software*, *85*(3), 596–606.

Schuster, R. (2010). *Requirements management for b2b processes: A worksheet driven approach from e3-value and rea to umm* (Doctoral dissertation). Vienna University of Technology.

Schuster, R., & Motal, T. (2009). From e3-value to rea: Modeling multi-party e-business collaborations. *IEEE Conf. on Comme. and Enterpr. Compu.*, 202–208.

Wohlin, C., Runeson, P., Höst, M., Ohlsson, M. C., Regnell, B., & Wesslén, A. (2012). *Experimentation in software engineering*. Springer Science & Business Media.