Enterprise Information Systems

ISSN: 1751-7575 (Print) 1751-7583 (Online) Journal homepage: https://www.tandfonline.com/loi/teis20

Assessing performance of mined business process variants

Lucas Van Den Ingh, Rik Eshuis & Sarah Gelper

To cite this article: Lucas Van Den Ingh, Rik Eshuis & Sarah Gelper (2020): Assessing performance of mined business process variants, Enterprise Information Systems, DOI: 10.1080/17517575.2020.1746405

To link to this article: https://doi.org/10.1080/17517575.2020.1746405

Published online: 10 Apr 2020.

ENTERPRISE INFORMATION SYSTEMS
https://doi.org/10.1080/17517575.2020.1746405

ARTICLE

Assessing performance of mined business process variants

Lucas Van Den Ingh[a], Rik Eshuis[b] and Sarah Gelper[b]

[a]Celonis, The Netherlands; [b]Department of Industrial Engineering & Innovation Sciences, Eindhoven University of Technology, Eindhoven, MB, The Netherlands

ABSTRACT

Process mining can be used to discover different variants of a business process. To optimise the performance of the process, the best variant needs to be identified. However, process performance is multi-dimensional and how these dimensions weigh out against each other to find the best performing process is often unclear. This paper proposes to use conjoint analysis and regression analysis to assess the overall performance of different process variants discovered by process mining. The approach has been implemented in a visual tool and has been applied to an industrial case study.

ARTICLE HISTORY
Received 22 October 2018
Accepted 19 March 2020

KEYWORDS
Process mining; performance measurement; process performance

1. Introduction

Business processes in organisations constantly generate event data logged in the organisation's information systems. Many organisations have adopted process mining tools that use these event data for the discovery and analysis of the actual execution of their business processes (van der Aalst 2016). A wide range of process mining techniques have been developed in the past years (Maita et al. 2018). Process mining helps organisations to identify systematic problems in their business processes, e.g. bottlenecks or undesired loops. Such analysis can be used to improve the performance of mined business processes by redesigning them (Maruster and van Beest 2009).

In practice, a process is typically performed in many different ways, leading to many different process variants. For instance, Figure 1 shows a simple Purchase-to-Pay (P2P) process where regular execution follows the solid lines but four variants are possible that follow the solid and dotted lines. Variant 1 occurs when a supplier is unable to deliver the ordered goods in one shipment, variant 2 when the procurement system contains an outdated price list, variant 3 when the purchase order was approved by the wrong person, and variant 4 when a basic office supply that is not in stock is bought directly from a shop and afterwards the purchase is created. More complex processes have many more variants.

To improve the performance of a process, organisations should determine which variants have the best performance and should be selected to be used as future standard. Process mining algorithms do not answer this question. The most commonly used tool for performance assessment in the process mining literature is conformance checking (Rozinat and van der Aalst 2008; van der Aalst 2016). Conformance checking

CONTACT Rik Eshuis h.eshuis@tue.nl Department of Industrial Engineering & Innovation Sciences, Eindhoven University of Technology, Eindhoven, MB 5600, The Netherlands

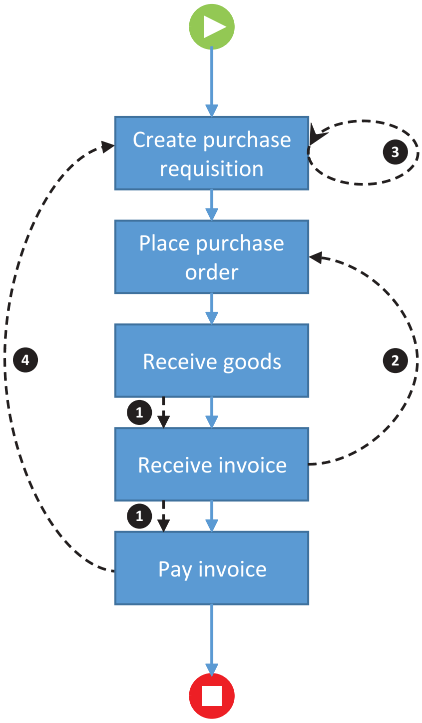


Figure 1. P2P process with four execution variants.

compares the logged behaviour of a process with a known benchmark process. For the performance valuation of business processes, however, conformance checking has two main drawbacks. First, the performance of business processes is typically assessed on multiple dimensions such as cost, time or quality of the process. Conformance checking does not allow for such a multi-dimensional evaluation of process performance. Second, the benchmark process is not necessarily the best performing process, even though it has been designed as standard process within the organisation. Selecting the best performing process is not possible using conformance checking. Therefore, we propose a performance assessment technique for mined business processes that accounts for the multi-dimensional nature of business process evaluation and does not assume the best performing process is known upfront.

The scientific literature on performance measurement of business processes only partially addresses this problem (Neely, Gregory, and Platts 2005; Kaplan and Norton 1992; Neely, Adams, and Kennerley 2002; Jansen-Vullers, Kleingeld, and Netjes 2008). Different performance dimensions and performance indicators have been proposed, but there is no method how to use these to choose the best performing process variant. Methods have been proposed to structure performance indicators in a hierarchy (Bititci, Suwignjo, and Carrie 2000; Han, Kang, and Song 2009; Rodrguez, Saiz, and Bas 2009) where lower level indicators contribute to indicators at an adjacent higher level. Having such a hierarchy would make the selection straightforward. However, constructing this hierarchy takes a lot of effort and is therefore not feasible for most companies. Moreover, for the problem of identifying the best performing variant, selecting the right variant is more important than knowing or predicting the exact performance of each variant, though naturally the latter feature makes the selection problem trivial.

The main contribution of this paper is an approach to select the best performing process variants from a set of mined process variants when the ideal benchmark process is unknown and it is not clear how performance indicators and overall performance are related. It is applicable to any organisation that logs processes that have variations, as it allows an organisation to identify the best process variant from a set of observed variants. This significantly advances the state of the art in performance analysis using process mining, since related work targets analysis of single processes only, as we explain in Section 2. This significantly advances the state of the art in performance analysis using process mining, since related work targets analysis of single processes only, as we explain in Section 2.

The approach leverages the knowledge of domain experts on ideal process performance. To extract this knowledge, the approach uses conjoint analysis (Green, Douglas Carroll, and Goldberg 1981; Hair et al. 2014) to estimate weights of different performance dimensions. Conjoint analysis is a standard technique used in new product development to identify how potential customers value the attributes of the new product. The advantage is that, by requesting the respondent to choose among fictitious products rather than rating each attribute separately, the relative importance of each attribute is unconsciously revealed. In our setting of process evaluation, domain experts are asked to choose among fictitious business processes in terms of performance according to their own judgement, which reveals the relative importance of the process dimensions. The weights that the experts implicitly assign to the performance dimensions are estimated from their choices. Once the weights are estimated, the overall performance of any process variant can be assessed.

The remainder of this paper is structured as follows. Section 2 discussed related work. Section 3 discusses the evaluation approach, which uses regression analysis and conjoint analysis. The approach has been implemented in a tool that interfaces with an existing process mining tool. Section 4 discusses a case study that we performed to evaluate the approach and the tool. Section 5 ends the paper with conclusions.

2. Related work

To understand and position the results from this paper, we discuss related work in performance measurement, process mining, and conjoint analysis.

2.1. Performance measurement

Performance measurement is a broad field that spreads both the strategic, tactical, and operational level of organisations. Several performance measurement frameworks have been proposed (Neely, Gregory, and Platts 2005), notably the Balanced Scorecard (Kaplan and Norton 1992) for the strategic level, and the performance prism (Neely, Adams, and Kennerley 2002) for the tactical and operational level.

While these approaches focus on organisational performance, other approaches focus on measuring performance of a single business process (Van Looy and Shafagatova 2016). Process performance involves both qualitative and quantitative aspects (Kueng 2000; Brand and Hans 1995). Since we focus in this paper on process mining, we are interested in quantifiable aspects of process performance (Jansen-Vullers, Kleingeld, and Netjes 2008). Typical quantifiable performance dimensions for business processes are time, cost, quality and flexibility (Jansen-Vullers, Kleingeld, and Netjes 2008).

The devil's quadrangle (Brand and Hans 1995; Reijers and Limam-Mansar 2005) relates these dimensions. The name of the quadrangle refers to the trade-off that needs to be made among the dimensions when optimising a process. The devil's quadrangle has been applied to business process redesign (Reijers and Limam-Mansar 2005; Limam-Mansar and Reijers 2005) and has also been used to quantify the impact of different redesign heuristics (Jansen-Vullers, Kleingeld, and Netjes 2008). Note that in business process redesign, redesigned processes are modelled manually and are not derived by mining and that conjoint analysis has never been applied before to select the most appropriate redesign, despite the merits of conjoint analysis as described below.

Performance indicators have been proposed to guide the search which processes to optimise to improve organisational performance (Han, Kang, and Song 2009). However, then the process models already exist and are not discovered by process mining. Furthermore, the performance indicators and their interdependencies are modelled in a hierarchical way. In contrast, we propose the use of conjoint analysis to discover the preferred constellation of the highlevel performance indicators.

2.2. Process mining

Process mining has developed very rapidly as a research field in the past years (van der Aalst 2016; Maita et al. 2018). The key focus is on discovering processes from event data, which record past process executions. One strand of research in process mining studies the derivation of variants of the same process by partitioning the analysed event data (Garca-Bañuelos et al. 2014; Greco et al. 2006). This paper studies the logical follow-up question: how to decide which variant performs best?

Another topic in process mining is checking conformance of event logs against a benchmark process (Rozinat and van der Aalst 2008). Conformance checking does not consider performance indicators and moreover assumes that the benchmark process, i.e. the best performing process, is known upfront. In this paper, we do consider performance indicators. Moreover, we analyse performance of process executions in order to identify the best performing process variant. This means there is no benchmark process. This motivates the use of conjoint analysis, which does not assume that a benchmark process exists.

Recent research in process mining for performance measurement has focused on different aspects, for instance analysing the impact of exceptions on process performance (Dijkman et al. 2019), analysing waste in logistic processes (Knoll, Reinhart, and Prueglmeier 2019) and predicting performance indicators of processes at the instance level (Verenich et al. 2019) and model level (Park and Song 2020). These papers analyse performance of a single process, while this paper analyses performance of a set of related process variants, with the aim to select the best one.

To the best of our knowledge the devil's quadrangle has not been applied to process mining before. This paper uses the devil's quadrangle to offer a common frame of reference for comparing the performance of the different process variants. There does exist work in process mining on performance analysis by clustering events (Song and van der Aalst 2007) or computing time-based performance indicators of processes (Hornix 2007) or computing performance indicators of business process redesigns (Cho et al. 2017). However, these works do not aim to select the best performing process variant based on performance analysis.

2.3. Conjoint analysis

Conjoint analysis is primarily known as a tool that designers use during new product development, in particular to explore different configurations of product attributes (Green, Douglas Carroll, and Goldberg 1981). The product designers use conjoint analysis to learn from customers how important the different product attributes are, and how one attribute weights against the other. The implicit preferences of customers are revealed while making choices in the conjoint study. An important feature of conjoint analysis is that no benchmark product configuration is needed. Instead, the preferences are inferred from comparisons of different configurations. Similarly, our approach starts from the premise that an organisation wants to know how important the different performance dimensions of the devil's quadrangle are, without having a benchmark quadrangle in mind. To the best of our knowledge, conjoint analysis has not been applied to process mining before.

Conjoint analysis can be used to rank a set of configurations (product configurations, or, in our case, process variant configurations) based on a latent utility model. Conjoint analysis has been integrated in decision support systems for product design (Luo et al. 2012) and positioning as well as for product recommendations (Scholz et al. 2015). Here, we integrate it in a tool for process evaluation based on process mining.

2.4. Conclusion

In sum, the main contribution of this paper is a new approach, implemented in a tool, to identify the best performing process variant from a set of process variants obtained with process mining. The approach does not require a benchmark 'best practice' process model to make the selection, but relies on conjoint analysis.

3. Domain-dependent evaluation approach

This section presents an approach to evaluate the performance of different variants of a process according to the devil's quadrangle. The evaluation approach depends on the context of the process, as assessed by domain experts. Part of the approach is an automated method that computes for a discovered process variant its performance score for each dimension of the devil's quadrangle and that contrasts it with the preferred

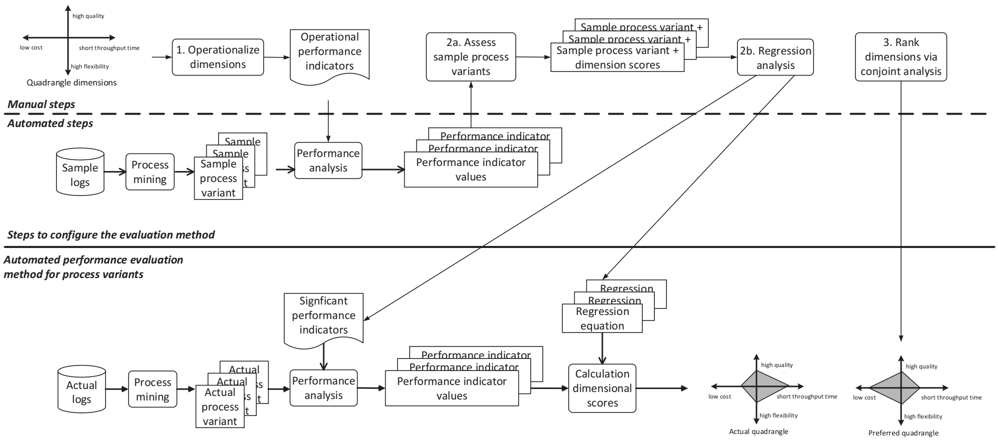


Figure 2. Evaluation approach.

performance scores, as identified by domain experts (bottom of Figure 2). The automated evaluation method is configured by inputs provided by a manual configuration method (top of Figure 2). The configuration method uses regression analysis and conjoint analysis to extract knowledge from domain experts into machine readable input for the evaluation method. Together, these two methods comprise the evaluation approach.

This section explains the three steps of the configuration method needed to configure the evaluation method. Next, it describes a tool that implements the evaluation method.

3.1. Step 1: operationalisation of the dimensions

The different dimensions of the devil's quadrangle represent abstract aspects of performance, which cannot be measured directly. However, these dimensions can be measured indirectly in many different ways by aggregating concrete performance indicators. To operationalise the performance measurement, first concrete performance indicators that can be measured directly need to be identified for each dimension. These indicators are used in subsequent steps to derive the overall performance for each dimension. Generic, concrete performance indicators for each dimension have been proposed in the literature (Jansen-Vullers, Kleingeld, and Netjes 2008). For many domains, there are standardised sets of relevant performance indicators, e.g. supply chains (Chae 2009). Alternatively, domain experts can be consulted to identify the concrete performance indicators. In most case, the dimension of each concrete indicator is immediately clear, for instance throughput time belongs to the time dimension. In case it is not clear to which dimension a concrete indicator belongs, domain experts can be consulted.

3.2. Step 2: analyse performance

In step 2, the performance of process variants is analysed. We do not consider how these process variants are selected, but it makes sense to select variants that are frequently occurring and have a high case coverage, i.e. cover a lot of cases. Domain experts can more easily assess performance of such frequent variants than of infrequent, exceptional variants.

Though each of the concrete performance indicators identified in step 1 can be measured for the selected variants, not all of them are relevant for measuring the actual performance. In step 2, domain experts are consulted to identify a subset of performance indicators that can be used to analyse the actual performance of each process variant with respect to the dimensions of the devil's quadrangle. This is done in two substeps.

First, domain experts manually assess the performance of different process variants for each dimension of the devil's quadrangle. The domain experts get as input the performance measurements of all the performance indicators identified in step 1, as well as a visual representation of the process variant. The assessment results in a ranking of the process variants for each of the dimensions. The measurement scale needs to be metric to support the next substep.

Second, a regression analysis is performed to find the relation between the performance measurements and the assessment of the dimensions. The performance indicators are independent variables whereas the variable for each dimension is a dependent variable. Regression analysis requires that all variables are metric, i.e. they use ratio or interval scales. Outcome of this substep is a set of regression formulas that define for each dimension the relation between the value of the performance indicators and the score on that dimension. Regression analysis is based on the ordinary least squares method, whose time complexity is quadratic in the number of features (Hastie, Tibshirani, and Friedman 2009).

3.3. Step 3: identify the preferred performance

The first two steps allow to measure and interpret the actual performance of each process variants against the dimensions of the devil's quadrangle. More precisely, the measured values for the performance indicators identified in step 1 can be translated using the regression formulas of step 2 into a score for each dimension of the devil's quadrangle. The third step is about identifying the best performance in terms of the shape of the devil's quadrangle. The output of this step provides the frame of reference for interpreting the actual performance of a business process.

Identifying the preferred performance can be done with conjoint analysis (Hair et al. 2014). This is an analysis technique that provides insight in the implicit preferences respondents have for a certain attribute (dimension), as well as the preferred level of that attribute and the relative importance of each attribute. In our setting, respondents are asked to choose among a set of different sample profiles of the devil's quadrangle according to their own judgement. By applying conjoint analysis, the implicit preferences for the dimensions are inferred. This allows us to calculate the performance of any possible process configuration and, thus, compare processes in terms of their performance. Since the data in this step are rank orders, the conjoint analysis algorithm is based on monotonic regression, which has quadratic time complexity (Burdakov, Grimvall, and Sysoev 2006).

3.4. Automated evaluation tool

We implemented the automated evaluation method specified in Figure 2 in a tool. It automatically evaluates different process variants, and thus gives users quick advice which process variant is best. The tool works with the process mining tool Celonis. This section introduces the highlevel requirements and architecture of the tool.

We identified three main requirements for the tool, which follow logically from the evaluation approach as sketched in Figure 2. For each process variant, the tool should show

- the performance of the process variant on each of the dimensions, as visualised in the devil's quadrangle.
- the values of the performance indicators for the variant that explain the scores for each dimension.
- the preferred shape of the devil's quadrangle against which the shape of the process variant can be compared.

The process architecture of the tool is based on the evaluation method in the bottom of Figure 2. In the following description, we list in parenthesis the related steps of the configuration method that produce the required input for the tool. The tool takes as input process variants obtained via a process mining tool and the list of concrete performance indicators (from step 1) that predict performance. Next, the tool calculates based on the regression equations (from step 2) for each variant from the values of the performance indicators its performance on each dimension of the quadrangle. The calculation results in scores for each dimension, that are visualised in a devil's quadrangle. The actual performance thus obtained can be compared with the preferred performance, also visualised in a devil's quadrangle (from step 3).

4. Case study

To illustrate the method and show its feasibility, we discuss how we applied the evaluation method and the tool in a case study performed at SAP Netherlands, part of the ERP supplier SAP. A full description of the case study can be found elsewhere (Van den Ingh 2016). SAP Netherlands gives advice to its clients how to improve their processes supported by the SAP ERP solution. SAP Netherlands is interested in applying process mining to improve their recommendations. They use Celonis for this purpose, which by their policy is the only allowed process mining tool within SAP.

For the case study, we analysed the purchase-to-pay (P2P) processes for four clients of SAP. In addition, we interviewed SAP consultants that are experts in these processes. We next describe the outcome of applying each step of the configuration method of Section 3. We used the output of these steps to configure an Excel-based analysis tool. Next, we evaluate the results of applying the tool to new process variants that were not considered in the configuration method.

4.1. Operationalisation of the dimensions

In order to operationalise the devil's quadrangle, a brainstorm session with SAP consultants was held, aimed at discovering which performance indicators are relevant in measuring the performance of a P2P process. In this session, first the research and devil's quadrangle were introduced to the participants, next the participants were asked to think of all possible performance indicators for a P2P that could be related to P2P process performance. Then, the performance indicators were assigned to the four dimensions of the devil's quadrangle. Participants of the brainstorm sessions were selected from the SAP-team (the intended users of the result) based on their experience with P2P processes.

The session started with brainstorming on any possibly relevant performance indicator for a P2P process. After all participants had written down a list of performance indicators, the performance indicators were assigned to the four dimensions of the devil's quadrangle. The participants were first asked to think of performance indicators regardless of the dimensions to ensure that performance indicators whose relation with the dimensions is not immediately clear would be overlooked. Afterwards, we allocated each performance indicator to one dimension.

Finally, we evaluated the list with experts from Celonis in order to remove performance indicators that cannot be measured. Reasons for removal are either that the required data is not available in SAP systems or that the required calculation is not available in Celonis. Table 1 shows the list of identified and measurable performance indicators.

4.2. Analyse performance

We measured the performance indicators listed in Table 1 for real P2P process data of four different companies. For each company, we selected the most frequently occurring process variants. As explained in Section 3.2, domain experts understand frequently occurring variants better and can therefore more easily quantify their performance. The percentage of cases covered by the selected variants for each of the four companies ranges between 35 and 60%.

Since the processes are industry-specific and subject to compliance regulations from their specific geographical region, company demographics have to be collected and included in the analysis. Because the anonymity of companies providing data needed to be guaranteed, not too many demographics could be used, since these could be use to trace the involved companies. The following demographics were recorded:

(1) Geographic region.
(2) Type of industry.
(3) The operations strategy.
(4) The type of sourcing.

The geographic region is recorded since it can influence the way business processes are executed (especially P2P processes) by e.g. specific tax rules. Next, type of industry is important, since the way processes are executed depends on the type of industry, due to for instance industry-specific regulations or industry-specific processes. The operations

Table 1. Measurable performance indicators for P2P processes.

	#	Performance indicator	Definition
Time	1	# of handover activities	The number of arrows in a process variant
	2	# of activities	The number of activities that have been executed
	3	# of no touch activities	The number of activities that are executed without any personal handling
	4	Duration (days)	The time between the first and last activity in the process
	5	Internal lead time (days)	The cumulative time between all internally executed activities (without e.g. waiting for an order to be delivered)
	6	Time before/after purchase discount deadline	The deviation from the purchase discount deadline (positive is payment before pdd). If there is no PDD then 0.
	7	Deviation from confirmed delivery date (- is late)	The deviation from the confirmed delivery date (positive is payment before cdd)
	8	% of orders within 2σ of avg duration	The percentage of orders that has a lead time within the range [$\mu \pm \sigma$]
Cost	9	% rework	The % of activities that was executed more than once
	10	Possible PD (% of PO value)	Percentage of potential purchase discount that could have been realised
	11	Missed purchase discount (% of PO value)	Percentage of purchase discount that has not been realised, relative to the PO value
	12	Purchase discount realised (% of PO value)	Percentage of realised purchase discount relative to the PO value
	13	Lost interest on capital (based on 1% interest)	The interest that is not earned by paying invoices before their purchase discount deadline
	14	Return goods present?	Boolean: is 'reverse goods receipt' present in this variant?
	15	# of users per €bln spent	The number of different resources that is used to process 1 billion worth of POs
Quality	16	Avg # of orders per suppliers	The average number of orders that per supplier
	17	Avg spend/supplier	The average PO value per supplier
	18	% catalogue spend (via SRM)	% of purchase value that is spent via SRM
	19	Days payable outstanding	The number of days between receiving and paying an invoice
	20	Deviation of payment term	Sum of all absolute deviations/number of deviating payments
	21	% payment done too early (vs contract conditions)	The number of payments that have been done before the purchase discount deadline date/total number of payments
	22	% payment done on time (vs contract conditions)	The number of payments that have been done on the purchase discount deadline date/total number of payments
	23	% payment done late (vs contract conditions)	The number of payments that have been done after the purchase discount deadline date/total number of payments
	24	Does this variant handle wrong master data?	Is wrong master data the cause for rework?
	25	# suppliers/bln spent	The number of suppliers/total order value in billions
	26	Compliance with payment blocks	Boolean: does this process execute payment while a payment block is present?
	27	Payment activity present?	Boolean: does process involve payment?
	28	Unplanned activities?	Does this process variant include any activities that do not add value?
	29	% not first time right	Percentage of total arrows going from activity n to activity n-1, n-2, etc
	30	# of duplicated process steps	# of activities that have been executed more than once
	31	# of errors	# of rework activities in a variant
	32	# of touches	# of unautomated activities (manually executed activities)
	33	# of automated activities	# of automatically processed activities
	34	Payment block present?	Boolean: does this variant include 'set payment block'?
	35	Vendor timely delivery performance	The number of orders that meet CDD/number of orders
	36	Double payments?	Boolean: does this variant include more than one payment handling?
Flexibility	37	% of materials/products processed	# of different items that is purchased/# total items
	38	# of vendors that can be processed	# of different vendors that are processed/total vendors
	39	# of order types that can be processed	# of different document types that are processed/total types
	40	% of cases handled in variant	The number of cases in variant/total number of cases (= coverage)
	41	# of changes	The number of 'change *' activities
	42	# of processes	Total number of variants
	43	Lead time/coverage	Lead time in days/case coverage of that variant

strategy of an organisation tells whether for instance low cost or high flexibility is aimed for. The process can source either direct or indirect materials. Direct materials are used to produce a product (e.g. raw materials) or service while indirect materials are not directly traceable to a product or service (e.g. printer supplies).

4.2.1. Assessment of process variants

The mined process variants were assessed by consultants that had analysed the processes themselves at the client. Therefore, the consultants are familiar with the both the processes and the companies owning the processes. All the selected participants had broad experience in P2P processes, and therefore their responses were regarded as equally important.

The assessment was done through a survey that included the following aspects:

- Introduction of the devil's quadrangle to ensure each assessor has a similar understanding of the framework and to minimise interpretation bias.
- For each process variant (the survey includes five variants) a graphical representation of the process and the score of all performance measures for that particular variant are shown.
- Fields to note the score for each variant according to the assessor, for all four dimensions. These scores are used as dependent variables in the subsequent analyses.

The surveys were either handed over personally, with an explanation and walk-through of the survey, or sent to the assessors by email, with an introduction and the statement that whenever anything in the survey is unclear, the assessor was requested to ask for clarification of this issue before proceeding with the survey.

Respondents were asked to rank the performance on all four dimensions on a ratio scale, ranging from 1 to 10, in which 1 represented the worst and 10 the best possible performance. Next to the identified performance indicators that were shown per dimension, the survey contained a graphical representation of the process variant, showing which activities in what order were executed. A set of dummy variables was created to include all aspects of the visualised process in the analysis. A metric measurement scale was preferred over a non-metric one, since a regression analysis can only be executed with a metric dependent variable. Data from four companies was available to be researched, and for all but one company two consultants that worked with the processes were willing to participate in the research, leading to seven participants. All surveys that were sent out, were returned with all questions answered.

4.2.2. Regression analysis

To discover which of the performance indicators from Table 1 are significant predictors of performance according to the expert assessments in the surveys, and to determine how these performance indicators are related to the overall performance, we performed a regression analysis in SPSS. These significant performance indicators, combined with the theoretical grounding of the performance indicators for their specific dimension, lead to models that calculate the performance on the different dimensions. The analysis assessed data from seven surveys, all consisting of five variants, resulting in 35 observations.

The information that was used in the analysis consisted of two parts: a dimension-specific part, consisting of the identified performance indicators for that dimension, and a generic part, containing company demographics and information on the activities that are executed in that variant (with dummy variables for each activity; since the respondents could see these activities on the image they had to be included in the analysis). The regression was performed in SPSS, and the models were created by different methods of adding variables: Enter, Stepwise, Forward and Backward, with the specific and generic performance indicators divided into two so called 'blocks'. Since the specific performance indicators are based on previous research steps, all methods are suitable for the specific block. For the generic block, all methods but Enter were used as a theoretical reason for including these indicators was not present.

For each dimension, we selected the three best models for that dimension based on the highest adjusted R^2. We show for one dimension, time, the three models in Table 2. The models satisfy all criteria for internal validity and the assumptions of normality, heteroscedasticity and linearity stated by Field (Field 2009). The values for (adjusted) R^2 show that a larger portion of the variation in the experts' evaluation is captured by the variables in the model. However, the model fit is not perfect, indicating that the opinion of the experts cannot be fully captured by the process variables, and the expert opinions contain unique information (see for instance (Blattberg and Hoch 1990)).

For each dimension the three best models (based on the highest adjusted R^2) are shown. Column B contains the unstandardised coefficients of the performance indicator in that model. The constant is obviously not influenced by the process, but the other values of B's should be multiplied by the value for that performance indicator, for that specific process variant. An empty cell indicates that the performance indicator is not present in that particular model. So, the assessment of the time dimension according to model 1 consists of a constant with value 10,032 minus 0,04998 times the end-to-end time (in days), minus 0,0004 times the number of execution variants minus 0,000054 times the average PO value (in).

4.3. Identify the preferred performance

To determine what values a well performing P2P process variant should have on the dimensions of the devil's quadrangle, a conjoint analysis was executed. This is an analysis that provides insight in the preference a respondent has for a certain attribute (dimension), as well as the preferred level of that attribute, according to Hair et al..(Hair et al. 2014). There are three types of conjoint analysis: choice-based conjoint, traditional

Table 2. The three best models for the time dimension (dependent variable).

	Model 1		Model 2		Model 3	
	B	Sig.	B	Sig.	B	Sig.
Constant	10,032	0,000	10,015	0,000	11,234	0,000
End-to-end time (days)	−0,04998	0,000	−0,058	0,000		
Total number of execution variants	−4,00E-04	0,003	−4,00E-04	0,005	−5,24E-04	0,020
Average PO value (€)	−5,40E-05	0,049			−1,1E-04	0,007
# of no touch activities					−1,676	0,018
Deviation from confirmed delivery date (days)					0,029	0,070
F	11,138	0,000	13,289	0,000	4,016	0,017
N (observations)	35		35		23	
Adjusted R^2 (R^2)	0,472 (0,519)		0,420 (0,454)		0,354 (0,472)	

conjoint and adaptive choice. Choice-based conjoint can handle a maximum of six attributes and has the advantage over other conjoint techniques that it creates a realistic choice task (thanks to a 'no choice' option) and can measure the interaction effect between attributes (Hair et al. 2014), and is therefore the conjoint type that was chosen. If respondents answer more than 30 choice tasks, the quality of the answers decreases (Hair et al. 2014).

As the aim of the research was to find the preferred ratio between the different dimensions, the conjoint analysis used a scale of 1–3 for each dimension (translated into a low, average and high performance on a dimension), that provides sufficient insight in this preference. The choice-based conjoint-tool in software package XLSTAT-Premium was used to generate the profiles and corresponding choice tasks, and to analyse the response, which showed the importance of each dimension. The software generated 10 choice tasks that had to be assessed by the participants, well below the stated upper limit of 30.

The respondents of the choice tasks were selected based on their earlier involvement of this research, as SAP consultants with knowledge about P2P processes previously participated in one of the brainstorm sessions or contributed by answering the survey about process performance. Consultants from Celonis that have broad experience in P2P processes and had received an introduction of the devil's quadrangle also responded to the choice tasks, leading to a group of 13 respondents that have an understanding of the devil's quadrangle. Figure 3 shows an example of a conjoint task, where respondents have to select a well performing process or select none, if all processes do not perform well in their opinition.

The XLSTAT conjoint-tool was used to analyse the responses and find the preferences for the dimensions. As a goodness-of-fit measure, Nagelkerke's R^2 (Nagelkerk 1991) was used. Since a conjoint analysis is an adapted regression, and this research is opinion-based, this value should again be higher than 0,25. The value for Nagelkerke's R^2 from the analysed data is 0,292, so above the threshold. The utilities per dimensions shown in Table 3 are normalised, meaning they add up to 100. Although the relative importance for cost is higher than the importance for time, the importance is so close that both time and cost are

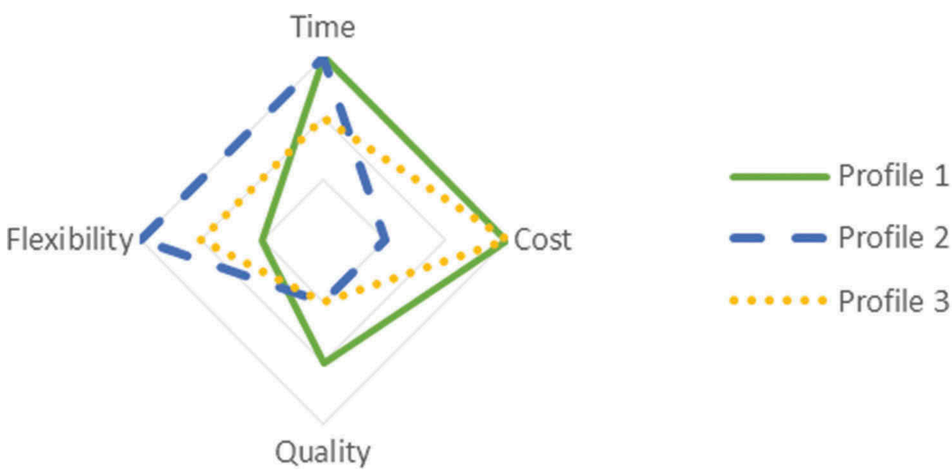


Figure 3. Example of a conjoint choice task: choosing the best profile.

Table 3. Importances of the dimensions.

Dimension	Utility	Importance
Cost	31,188	High
Time	30,168	High
Quality	25,457	Medium
Flexibility	13,186	Low

the dimensions that should have the highest performance. Since Time and Quality have almost the same utility, they are scored as high. Flexibility has the lowest score and quality has a medium score. These qualitative scores can be be translated into the preferred shape of the devil's quadrangle (Figure 4).

4.4. Validation

We implemented an Excel-based analysis tool that is configured with the output of the steps. Figure 5 shows a screenshot. The shapes of the devil's quadrange for variants 3 and 4 resemble most the ideal shape and could therefore be identified as being closest to the ideal.

In order to evaluate whether the tool is capable of giving a valid performance score on all dimensions, two consultants assessed a new P2P data set. Both consultants took part in identifying the list with performance indicators in Table 1 (one during the brainstorm, one by checking the list for completeness), and also have experience with the company of the new data set. The setup of the assessment was the same as the one used for the first companies (Section 4.2.1), i.e. the consultants received a survey that contained the five most frequent process variants of the new data set, and they were asked to rank the performance of the five variants for each of the four dimensions. The performance values that they assigned to each dimension were then tested to see whether they fit into the 95% confidence intervals of the three models that were selected for that dimension during regression analysis (Section 4.2.2). Because this validation consists of 5 observations per participant, so 10 in total, the expected number of observations outside the confidence interval is 0,5 observation, so both null or one observation outside the confidence interval meet the expectation, translating to 90% or 100% of observations within the confidence interval.

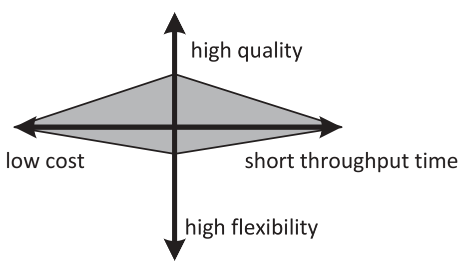
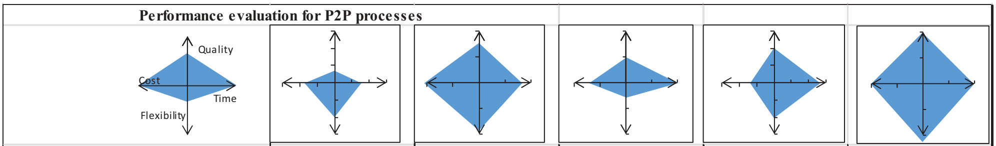


Figure 4. Preferred quadrangle according to Table 3.

Performance evaluation for P2P processes



Ideal shape	Variant 1	Variant 2	Variant 3	Variant 4	Variant 5
Cost	8,6	10,5	7,1	8,4	10,1
LOG(Users per €bln spent)	1,00	0,50	100	15	1
Time	7,9	8,3	10,0	9,7	9,8
Duration (days)	54,00	44,00	18	22	17,44
Quality	3,4	7,6	5,0	9,0	10,0
% payments done late	18%	1%	12%	30%	99%
% unautomated activities	90%	10%	70%	25%	10%
Vendor delivery performance	18%	30%	45%	97%	55%
Flexibility	10,1	9,2	2,8	9,1	11,3
Case coverage	7%	5%	3%	2%	1%
% of order types	51%	13%	3%	20%	15%
Other					
Relative percentage of PO value in this variant	27%	32%	6%	14%	21%
Total number of execution variants	2514	2514	2514	2514	2514
Average PO value (€)	€ 8.514,00	€ 10.000,00	€ 1.858,00	€ 4.235,00	€ 6.583,00
'Create PR' activity present?	1	1	1	1	0

Figure 5. Excel-based analysis tool.

For each of the three best models in each dimension, as selected with regression analysis, Table 4 shows the percentage of observations that are in the 95% confidence interval and the mean absolute prediction error (MAPE). For all dimensions, at least one model has a MAPE lower than two, indicating that the average prediction according to that model is less than 2 points from the value that was assigned by the consultants in the validation phase. As the rating takes place on a scale from 1 to 10, this deviation is quite large. Especially for the flexibility dimension, two models have a MAPE that is a lot wider than the 1–10 range on which respondents were asked to rate the process variants and one model produces scores that have little in common with those given by the participants. This indicates that to align the automated evaluation method with expert opinions, more data sets are needed, especially for the flexibility dimension.

As generalisability of the framework is an essential part of this research, it is more important that a model is capable of predicting the performance of a process outside the data sets that were used to generate the models than that it has a slightly higher adjusted R^2 (since this value only indicates the fit of the model on the data set that was analysed in the first phase). Therefore, the model with the smallest MAPE was selected as the most valid model based on the validation phase. For time, quality and flexibility model 2 is selected, for cost model 3. In Table in Table 4, these models are marked bold. Table 5 shows the equations for these models while Table 6 gives an overview of the significant performance indicators according to these models, as well as the abbreviations used in the formulas.

Table 4. Validation data.

	Time		Cost		Quality		Flexibility	
	MAPE	% cases in 95% conf int	MAPE	% cases in 95% conf int	MAPE	% cases in 95% conf int	MAPE	% cases in 95% conf int
Model 1	2,5089	90%	1,3730	90%	1,9467	70%	97,0247	10%
Model 2	**1,7221**	**90%**	**2,1376**	**90%**	**1,9457**	**100%**	**1,6364**	**90%**
Model 3	4,9407	70%	**1,2258**	**90%**	**2,6913**	**100%**	**75,4045**	**100%**

Table 5. Formulas for the models selected for each dimension.

$$E(Time) = 10.015 - 0.058 * E2E - 0.0004 * VAR$$

$$E(Cost) = 10.100 - 1.476 * \log(Users\_BLN)$$

$$E(Quality) = 6{,}770 - 3{,}263E^{-09} * Avg\_Sup + 2{,}158 * Paym\_Late + 1{,}934 * Vend\_Perf - 3{,}971 * Manual$$

$$E(Flexibility) = 4{,}497 - 23{,}334 * Cases^2 + 4.054 * Rel\_PO - 1{,}351 * Cr\_PR - 1{,}656 * Goods$$

Table 6. Significant performance indicators for the validated models.

Dimension	Performance indicator	Abbreviation
Time	Duration (days)	E2E
Cost	# of users per €bln spent	Users_BLN
Quality	% payment done late (vs contract conditions)	Paym_Late
	% of manual executed activities	Manual
	Vendor timely delivery performance	Vend_Perf
	Average spend per supplier	Avg_Sup
Flexibility	% cases handled	Cases
Other	Relative percentage of PO value in this variant	Rel_PO
	Total number of execution variants	VAR
	Goods receipt activity present?	Goods
	Create PR activity present?	Cr_PR

5. Conclusion

The main contribution of this paper is a novel approach to automatically evaluate the performance of mined process variants by using an operationalised version of the devil's quadrangle. We focussed especially on the different steps to configure the automated evaluation method, which use both regression analysis and conjoint analysis to translate knowledge from domain experts into machine readable input for the evaluation method. In particular, the approach does not require a benchmark process to determine which process variant performs best, which makes the approach suitable in cases a benchmark process is unknown or does not exist.

Application of the approach in a case study gives promising results: most of the scores provided by the domain experts on each dimension were in the 95% confidence interval of the different regression models that were developed. Given the limited amount of data, we expect that including more data sets for the regression analysis can significantly improve the accuracy of the method for the case study.

There are several directions for further research. An interesting topic is incorporating user feedback from actual assessments to improve the accuracy of the evaluation method. Next, the approach can be applied to processes that are less standardised than P2P to assess whether the approach is useful for business processes with more variability.

Disclosure Statement

No potential conflict of interest was reported by the authors.

References

Bititci, U., P. Suwignjo, and A. Carrie. 2000. “Quantitative Models for Performance Measurement Systems.” *International Journal of Production Economics* 64 (1–3): 231–241. doi:10.1016/S0925-5273(99)00061-4.

Blattberg, R. C., and S. J. Hoch. 1990. “Database Models and Managerial Intuition: 50% Model + 50% Manager.” *Management Science* 36 (8): 887–899. doi:10.1287/mnsc.36.8.887.

Brand, N., and V. D. K. Hans. 1995. *Werkstroomanalyse En -ontwerp: Het Logistiek Vriendelijk Ontwerpen Van Informatiesystemen*. Deventer: Kluwer Bedrijfswetenschappen.

Burdakov, O., A. Grimvall, and O. Sysoev. 2006. “Data Preordering in Generalized PAV Algorithm for Monotonic Regression.” *Journal of Computational Mathematics* 24 (6): 771–790.

Chae, B. (Kevin). 2009. “Developing Key Performance Indicators for Supply Chain: An Industry Perspective.” *Supply Chain Management: An International Journal* 14 (6): 422–428. doi:10.1108/13598540910995192.

Cho, M., M. Song, M. Comuzzi, and S. Yoo. 2017. “Evaluating the Effect of Best Practices for Business Process Redesign: An Evidence-based Approach Based on Process Mining Techniques.” *Decision Support Systems* 104: 92–103. doi:10.1016/j.dss.2017.10.004.

Dijkman, R., O. Turetken, G. R. van IJzendoorn, and M. de Vries. 2019. “Business Processes Exceptions in Relation to Operational Performance.” *Business Process Management Journal* 25 (5): 908–922. doi:10.1108/BPMJ-07-2017-0184.

Field, A. 2009. *Regression*. London: Sage.

Garca-Bañuelos, L., M. Dumas, M. La Rosa, J. De Weerdt, and C. C. Ekanayake. 2014. “Controlled Automated Discovery of Collections of Business Process Models.” *Information Systems* 46: 85–101. doi:10.1016/j.is.2014.04.006.

Greco, G., A. Guzzo, L. Pontieri, and S. Domenico. 2006. “Discovering Expressive Process Models by Clustering Log Traces.” *IEEE Transactions on Knowledge and Data Engineering* 18 (8): 1010–1027. doi:10.1109/TKDE.2006.123.

Green, P. E., J. Douglas Carroll, and S. M. Goldberg. 1981. “A General Approach to Product Design Optimization via Conjoint Analysis.” *Journal of Marketing* 45 (3): 17–37. doi:10.1177/002224298104500302.

Hair, J. F., W. C. Black, B. J. Babin, and R. E. Anderson. 2014. “Conjoint Analysis.” In *Multivariate Data Analysis*, 341–415. 7th ed. Harlow: Pearson Education.

Han, K. H., J. G. Kang, and M. Song. 2009. “Two-stage Process Analysis Using the Process-based Performance Measurement Framework and Business Process Simulation.” *Expert Systems with Applications* 36 (3): 7080–7086. doi:10.1016/j.eswa.2008.08.035.

Hastie, T., R. Tibshirani, and J. H. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. *2nd* ed. Springer Series in Statistics. New York: Springer.

Hornix, P. T. G. 2007. “Performance Analysis of Business Processes through Process Mining.” Master’s thesis, Eindhoven University of Technology.

Jansen-Vullers, M. H., P. A. M. Kleingeld, and M. Netjes. 2008. “Quantifying the Performance of Workflows.” *Information Systems Management* 25 (4): 332–343. doi:10.1080/10580530802384589.

Kaplan, R. S., and D. P. Norton. 1992. “The Balanced Scorecard-Measures That Drive Performances.” *Harvard Business Review* 70 (1): 71–79.

Knoll, D., G. Reinhart, and M. Prueglmeier. 2019. “Enabling Value Stream Mapping for Internal Logistics Using Multidimensional Process Mining.” *Expert Systems with Applications* 124: 130–142. doi:10.1016/j.eswa.2019.01.026.

Kueng, P. 2000. “Process Performance Measurement System: A Tool to Support Process-based Organizations.” *Total Quality Management* 11 (1): 67–85. doi:10.1080/0954412007035.

Limam-Mansar, S., and H. A. Reijers. 2005. “Best Practices in Business Process Redesign: Validation of a Redesign Framework.” *Computers in Industry* 56 (5): 457–471. doi:10.1016/j.compind.2005.01.001.

Luo, X. G., C. K. Kwong, J. F. Tang, and Y. L. Tu. 2012. "Optimal Product Positioning with Consideration of Negative Utility Effect on Consumer Choice Rule." *Decision Support Systems* 54 (1): 402–413. doi:10.1016/j.dss.2012.06.002.

Maita, A. R. C., L. C. Martins, C. R. López Paz, L. Rafferty, P. C. K. Hung, S. M. Peres, and M. Fantinato. 2018. "A Systematic Mapping Study of Process Mining." *Enterprise IS* 12 (5): 505–549. doi:10.1080/17517575.2017.1402371.

Maruster, L., and N. R. T. P. van Beest. 2009. "Redesigning Business Processes: A Methodology Based on Simulation and Process Mining Techniques." *Knowledge and Information Systems* 21 (3): 267–297. doi:10.1007/s10115-009-0224-0.

Nagelkerk, N. J. D. 1991. "A Note on A General Definition of the Coefficient of Determination." *Biometrika* 78 (3): 691–692. doi:10.1093/biomet/78.3.691.

Neely, A., C. Adams, and M. Kennerley. 2002. *The Performance Prism: The Scorecard for Measuring and Managing Business Success*. Harlow: Financial Times.

Neely, A., M. Gregory, and K. Platts. 2005. "Performance Measurement System Design: A Literature Review and Research Agenda." *International Journal of Operations & Production Management* 25 (12): 1228–1263.

Park, G., and M. Song. 2020. "Predicting Performances in Business Processes Using Deep Neural Networks." In *Decis. Support Syst*, 129. doi: 10.1016/j.dss.2019.113191.

Reijers, H. A., and S. Limam-Mansar. 2005. "Best Practices in Business Process Redesign: An Overview and Qualitative Evaluation of Successful Redesign Heuristics." *Omega* 33 (4): 283–306. doi:10.1016/j.omega.2004.04.012.

Rodrguez, R., J. J. A. Saiz, and Á. O. Bas. 2009. "Quantitative Relationships between Key Performance Indicators for Supporting Decision-making Processes." *Computers in Industry* 60 (2): 104–113. doi:10.1016/j.compind.2008.09.002.

Rozinat, A., and W. M. P. van der Aalst. 2008. "Conformance Checking of Processes Based on Monitoring Real Behavior." *Information Systems* 33 (1): 64–95. doi:10.1016/j.is.2007.07.001.

Scholz, M., V. Dorner, M. Franz, and O. Hinz. 2015. "Measuring Consumers' Willingness to Pay with Utility-based Recommendation Systems." *Decision Support Systems* 72: 60–71. doi:10.1016/j.dss.2015.02.006.

Song, M., and W. M. P. van der Aalst. 2007. "Supporting Process Mining by Showing Events at a Glance." In *In Seventeenth Annual Workshop on Information Technologies and Systems (WITS'07)*, edited by K. Chari, A. Kumar, 139–145.

Van den Ingh, L. 2016. "Evaluating Business Process Performance Based on Process Mining."

van der Aalst, W. M. 2016. *Process Mining - Data Science in Action*. *Second* ed. Berlin, Heidelberg: Springer.

Van Looy, A., and A. Shafagatova. 2016. "Business Process Performance Measurement: A Structured Literature Review of Indicators, Measures and Metrics." *SPRINGERPLUS* 5: 24. doi:10.1186/s40064-016-3498-1.

Verenich, I., M. Dumas, M. La Rosa, and H. Nguyen. 2019. "Predicting Process Performance: A White-box Approach Based on Process Models." *Journal of Software: Evolution and Process* 31: 6.