Bringing Rigor to the Qualitative Evaluation of Process Mining Findings: An Analysis and a Proposal

Jelmer J. Koorn
Utrecht University
Utrecht, the Netherlands
j.j.koorn@uu.nl

Iris Beerepoot
Utrecht University
Utrecht, the Netherlands
i.m.beerepoot@uu.nl

Vinicius Stein Dani
Utrecht University
Utrecht, the Netherlands
v.steindani@uu.nl

Xixi Lu
Utrecht University
Utrecht, the Netherlands
x.lu@uu.nl

Inge van de Weerd
Utrecht University
Utrecht, the Netherlands
g.c.vandeweerd@uu.nl

Henrik Leopold
Kühne Logistics University
Hamburg, Germany
henrik.leopold@the-klu.org

Hajo A. Reijers
Utrecht University
Utrecht, the Netherlands
h.a.reijers@uu.nl

***Abstract*—Before the findings of a process mining project can be turned into actionable insights or recommendations, it is essential to make sure that the findings are actually valid. Therefore, the evaluation of the findings is a crucial part of a successful process mining project. Current process mining methodologies, however, fall short in providing actionable support to perform such an evaluation. This is especially true when domain experts are involved. To close this gap, we performed a literature study considering all process mining case studies published in the last two decades. In total, we identified 244 candidate papers of which we analyzed 80 in depth. Based on this literature study, we found a need for a more systematic approach for qualitative evaluations in process mining projects where domain experts are involved. Therefore, we build on these results to propose six validation strategies, which originate from qualitative research. We hope that this proposal for more rigor in the evaluation phase of process mining projects helps to move the discipline forward.**

***Index Terms*—Process Mining, Evaluation, Methodology, Domain Expert, Qualitative Research**

I. Introduction

Process mining is widely used to discover, analyze, and improve business processes in various industries. Given its popularity, several methodologies have been developed to guide both practitioners and academics in performing process mining projects [1]. An important factor in process mining methodologies is the interaction between process analysts and domain experts. This interaction often takes place at the start of a project, during data extraction and pre-processing, and at the end of a project, i.e., during the *evaluation*. The evaluation is an important step in any process mining project as it is concerned with making sure that the findings are actually valid. The domain experts are essential in this context since they are able to assess and interpret the findings, and translate them into actionable insights and recommendations. In this way, they make sure that a process mining project eventually results in organisational value [2].

This research was supported by the NWO TACTICS project (628.011.004).

While existing process mining methodologies generally recognize the importance of evaluating with domain experts [1], they do not provide specific guidelines as to how such an evaluation should be performed. The main evaluation focus of most existing process mining research is determining the effectiveness of proposed techniques using established metrics such as precision, recall, etc. Projects with this evaluation aim typically use a *quantitative* research approach, which commonly uses methods such as surveys and experiments. These are methods that aim to numerically test theories and models by examining the relation between variables [3]. While this is often sufficient to evaluate the technique itself, the translation of the findings of a process mining project into actionable insights and recommendations requires an additional evaluation step that involves domain experts [2]. In these evaluations often a *qualitative* research approach is used in which methods such as semi-structured interviews and focus groups are common. This approach is used to explore the 'why' and 'how' behind discovered models or theories [4].

Overall, we observe an abundance of projects that involve organizational partners where evaluations take place at the end of a project, but the field varies widely in their approach to the evaluation with these partners. Based on this observation we hypothesize that informal ways of evaluating have been applied over time. With that in mind, we perform a literature study to describe current practices in process mining projects. Additionally, we look into strategies from the qualitative research field, which provide more guidance during qualitative evaluations with domain experts in process mining projects.

This research aims to help move process mining research forward by offering support for process mining experts that seek to perform a qualitative evaluation in their project. More concretely, we propose a list of six validation strategies from the qualitative research field. These strategies should be considered when performing process mining evaluations in which domain experts are involved.

The rest of the paper is structured as follows. First, we describe the methodology behind the literature study. Then, the results section describes the goals and methods of these case studies. In the proposal section we describe six existing validation strategies from the qualitative research domain. Finally, we conclude our work, discuss the limitations, and sketch directions for future research.

II. Research method

We performed an in-depth, systematic literature review aimed at reviewing the existing literature on process mining projects in which domain experts are involved. We followed the guidelines presented by [5]. We illustrate the search and selection process in Fig. 1. In total, four of the authors actively performed the literature study, we refer to this as the literature team. We explain the process in more detail below.

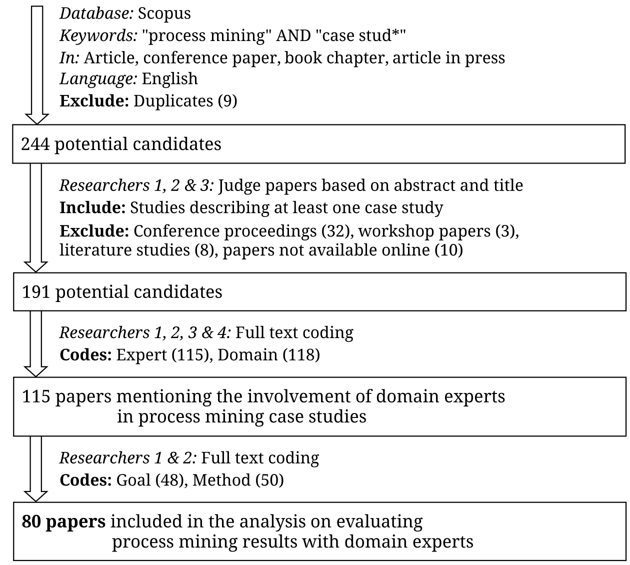

Fig. 1. Literature study process

A. Extraction and Abstract Screening

Our objective was to include papers that describe the process of performing a process mining project in practice. Therefore, we used the Scopus database to collect a broad sample of papers. The authors in [6] showed that Scopus provides the best balance between relevance and quantity when researching process mining papers. The literature team extracted the papers in October 2020 using the keywords “process mining” AND “case stud*”. The search resulted in 244 potential candidates to be judged on abstract and title. Three members of the literature team were involved with this next phase of the extraction process. To ensure uniformity amongst the team members, the first five papers were individually screened after which the decision to include/exclude was verified by one other team member. The result of this phase was a set of 191 papers that described one or more process mining case studies.

B. Full Text Coding

Next, each member of the literature team coded five papers in NVivo [7], a qualitative analysis software by QSR. The code ‘expert’ was used whenever we came across a piece of text in which a domain expert was mentioned in the context of a process mining project. We also coded information on the domain in which the process mining case study was performed.

In the second part of the coding phase, two members of the literature team, we refer to them as the core team, performed another round of full text reading, each on half of the papers mentioning in any way the involvement of domain experts in process mining case studies. We started with five papers before synchronizing and discussing the proposed sub-codes (Goal and Method), after which we read another ten papers before synchronizing a second time. The result of this phase was a set of quotations related to the two topics, collected from a total of 80 papers.

C. Analysis and Synthesis

In the analysis and synthesis phase, the core team revisited their quotations for the two sub-codes. Each team member created an overview with a reflection on the information found for both the goal and method of the evaluation, for each of the papers. After both team members were finished with creating the overview and discussing common themes together, the overviews were merged. Finally, the core team performed two rounds of abstraction, the first round analyzed the lower level codes. These were grouped together in higher level codes that were given by both team members. For example, the description ‘conformance checking’ and ‘identifying non-conforming cases’ were combined, this resulted in 14 description codes. Then, the second round of abstraction took a more holistic perspective. In the second round the core team considered fundamental differences between the higher level codes. The outcome of these rounds of abstractions are presented in the results.

III. Results

In what follows, we present the results from the coding of the literature review. Recall that we look into process mining case studies in which a domain expert is involved in the evaluation. In these projects, we refer to the authors as *process analysts* to avoid confusion. We coded in two main themes: (1) goals of the evaluation, and (2) method of evaluation. Having established this, we can cross-reference the results to see the overlap between the codes. Below, we first go into detail on each of the themes, followed by the cross-reference analysis.

A. Goal of evaluation

By studying the various process mining case studies in which domain experts are involved, we found that there were substantial differences in the objectives that were set out in the evaluation phase of the individual papers. In Table I we

TABLE I
TABLE CONTAINING ALL CITATIONS PER GOAL CATEGORY

Goal	Sub-goal	Papers
Artifact	Understandability	[8]–[22]
	Usability	[9], [11], [14], [15], [17], [19], [23]–[25]
	Quality	[9], [12], [14], [16], [17], [22], [23], [26]–[42]
Insights	Findings confirmation	[2], [17], [19], [24], [27], [30]–[34], [34], [43]–[70]
	Relevance	[18], [24], [25], [28], [46], [56], [68], [69]
	Generalizability	[51], [61], [62], [64]
	Conformance checking	[48], [49], [71], [72]

TABLE II
TABLE CONTAINING ALL CITATIONS PER METHOD CATEGORY

Method type	Methods	Papers
Quantitative	Survey	[11], [17], [24], [38], [49], [68]
	Manual annotation	[8], [9], [12], [28], [29], [35], [39], [41], [69], [73]–[75]
	Experiment	[9], [35], [38], [68]
Qualitative	Focus group	[21], [56]
	Undefined discussion	[2], [12], [13], [15], [16], [19], [22], [23], [25], [27], [34]–[36], [42]–[44], [47], [49], [50], [52], [54], [55], [57]–[60], [63]–[65], [67], [68], [70], [72], [76]–[81]
	Interviews	[2], [8], [9], [12], [14], [17], [28], [35], [38]–[41], [48], [61], [69], [72]–[75]
	Workshop	[10], [14], [62], [71]

present the goal and sub-goal categories found in the literature study. In particular, we found the most fundamental difference in the focus of the evaluation. Here, two main foci can be distinguished: (1) focusing on the evaluation of the artefact itself, or (2) focusing on the evaluation of the insights that can be generated from the artefact.

The first type of evaluation focuses on evaluating the artefact that is created. Slightly less than half of the articles belong to this type of evaluation. We find three measures that highlight which aspects of the artefact are evaluated: (1) understandability (n = 15), (2) usability (n = 9), and (3) quality (n = 25). Usually, a process analyst creates an artifact which is consequently presented to a business domain expert. The *understandability* of the artefact refers to the level of comprehension with which the business expert can examine the artifact. Concrete examples of evaluation criteria for understandability are complexity of the artefact (e.g. [12]) or readability of an artefact (e.g. [44]). The *usability* captures the ease of use of the artefact. In one of the case studies the process analyst collaborated with domain experts to determine specific KPIs for the process, also in light of conformance checking [61]. Finally, the *quality* of the artefact describes how well the artefact is constructed. The largest part of the studies focus on this latter aspect (n = 25). Typically, the quality of an artifact consists of several dimensions: correctness (the artefact is true and correct), completeness (the artefact contains all appropriate elements), conciseness (the artefact represents the same information in a similar fashion repeatedly), and consistency (the artefact does not contain contradictions). Note that quality here only refers to the internal quality of an artefact; the external quality (i.e. clarity) is captured in the usability and understandability features.

The slight majority of the studies focus on a second type of evaluation: the insights that one can generate based on the artefacts that are created during a process mining project; we refer to these evaluations as *insights evaluation*. The largest part of research that focuses on insights evaluation considers the *confirmation* of findings as the most important aim for the evaluation (n = 41). Typically, insights evaluations focus on the interpretation and explanation of findings represented in the artefacts. For example, some studies compare the expectations of business experts to the insights generated from the artefact (e.g. [50]). Another goal is to regard the *relevance* of the findings, where the process analysts try to determine how valuable the produced artefact is to the organization. In one case study, the evaluation focused on four questions, one of which concerned determining the relevance of their proposed framework for designing data visualization for process mining diagrams [14]. Some studies also consider the *generalizability* of the produced artefact. To exemplify, in one case study a multi-perspective approach was tested in one context, where the evaluation focused on gaining insights into the applicability of the approach in other contexts [62]. Finally, some studies focus on *conformance checking* based on the produced artefact. Here, the goal is to perform a ”reality check”. These studies compare the artefact that is created based on real data (event logs) to the existing protocols within an organization as indicated by the domain experts (e.g. [49]).

B. Method of evaluation

We now turn our attention to the *method* that is used in the evaluation of process mining case studies in which domain experts are involved. In Table II, we present the coding hierarchy for the method of the evaluation. We can distinguish two main types of method: qualitative and quantitative. Within each type of method, we identify a number of data collection techniques (e.g. survey). On an aggregate, level we observe a strong tendency to perform the evaluation using a qualitative method (n = 64).

What stands out within the qualitative methods is that we find a clear pattern as to the lack of structure in the evaluation phase of a process mining project. The vast majority of the studies perform an *undefined discussion* to evaluate their results (n = 39). These undefined discussions are usually described using the following terminology: “The results were presented to business expert.” or “We discussed the results of the project with domain expert.”. In these studies no guidelines, rules, or protocols are described that support the evaluation phase of the research.

Other data collection techniques are also employed in the qualitative method group, such as *interviews* (n = 19), *focus groups* (n = 2), and *workshops* (n = 4). These studies all

TABLE III
CROSS-REFERENCE RESULTS WHEN COMBINING THE HIGHEST LEVEL CODES IN METHOD (COLUMNS) AND GOAL (ROW) OF EVALUATION

	Qualitative	**Quantitative**	**Qualitative & Quantitative**	**Total**
Artefact	7	3	5	**15**
Insights	28	1	2	**31**
Insights & artefact	7	2	2	**11**
Total	**42**	**6**	**9**	**57**

describe a setting in which the domain expert and process analysts interactively discuss the study findings following some sort of protocol. Often, a rigorous approach and reflection on the chosen method and technique lack in these works as well.

By contrast, we observe that in the quantitative method group, studies follow a more rigorous approach. For example, a strand of studies collect data through the *annotation* of process mining results (n = 12). In one case study, the authors ask experts to manually annotate the complexity of so-called attack models for an intrusion detection system [8]. Furthermore, some studies use *surveys* to evaluate their findings (n = 6). This is generally done when the evaluation criteria are determined *a priori*. For example, conjoint analysis is used to determine the weights of various performance dimensions [68]. Finally, we see a handful of studies setting up an experiment (n = 4) to study their results. For example, the authors of [49] simulate a hospital setting to validate their algorithm.

C. Cross-reference analysis

As presented above, process mining project evaluations can differ fundamentally in two respects: (1) their goal, and (2) their method. It is important to note that the divisions presented above are not mutually exclusive: a study can pursue both types of goals or employ multiple types of method. The goal of the present study is not to provide a normative perspective on when to use which method or define what goal. Rather, we describe the current practices in process mining projects. There are two things that stand out when we do this: (1) a *qualitative approach* is often taken to evaluate process mining findings, and (2) the majority of the projects aim to generate insights based on their artefacts, i.e. *insights evaluation*.

To gain a better understanding of the relation between method and goal in the evaluation of a process mining project, we cross-referenced the higher level codes. In Table III the results are visualized for all 57 studies for which a goal and a method was defined. What we can see from the table is that most studies use a qualitative approach (n = 42), especially when the goal is to gain insights (n = 28). Another apparent trend shows that there are quite some studies that use a mixed-method approach (n = 9) or have mixed goals (n = 11).

What we can infer from this is that a qualitative approach with the goal of insights evaluation is most frequently used. Recall that the most frequently used method in the qualitative approach is an undefined discussion. In addition, note that almost a third (n = 23) of the original studies (n = 80) is not included in the cross-reference analysis (n = 57) as they do not define both a goal and a method for their evaluation. Thus, a large portion of studies lack a structured method to perform this insights evaluation. Most existing case studies do not follow specific guidelines to properly evaluate their findings in a qualitative manner. Thus, we observe that a systematic approach in these types of evaluations is missing. We observe that individual studies sometimes use good practices. We believe this requires a look into existing best practices from qualitative research literature.

IV. PROPOSAL

From the literature study we can conclude that a qualitative approach is often taken, but a systematic approach for determining the accuracy and meaning of findings are mostly lacking. In the wider scientific literature there are many ways in which normative support is offered to researchers. Such ways of support can be divided into different layers, which Saunders [82] illustrates by means of a so-called 'research onion'. When designing a study, researchers peel off the individual layers one by one, going from broad research philosophies all the way down to specific data collection and analysis techniques. A process mining study can be designed in a similar way. Researchers may start by deciding on a broad research paradigm, such as design science [83], before deciding on particular research methods, such as a case study. Within such a case study, they may choose specific data collection techniques such as interviews or focus groups. Up to this point, the literature on evaluation with domain experts in process mining is quite explicit. However, the *inner layers* are not as clear. There is little discussion on how data is best collected and analyzed or how accuracy is ensured. To propose a way to fill this gap, we look into literature from qualitative research. Based on this literature, we propose six validation strategies that should be considered in qualitative process mining evaluations. Below, we will go into detail for each strategy and reflect on the extent to which these strategies can be observed in current process mining practices.

A. Validation Strategies

In their seminal book on qualitative research methods, the authors of [4] discuss validation strategies, see Table IV. They propose a number of strategies to perform in qualitative studies: (1) engage with the field of research, (2) triangulation, (3) peer review or external audit, (4) refine work hypothesis, (5) clarify bias, (6) perform member checking. These strategies are all potentially relevant for process mining projects that qualitatively evaluate their findings with domain experts. In their work, the authors of [4] recommend researchers to always follow at least two strategies when engaging in qualitative research. Below, we first elaborate on each strategy, explaining how they aim to improve qualitative research in general. Second, we provide guidelines to how the strategy can be applied in the context of an evaluation with domain experts in a process mining project.

1) Engagement and understanding of the field: This strategy refers to the relation a researcher builds with the study participants, the understanding the researcher builds of the (organizational) culture, and the ability of the researcher to spot how misinformation might influence the study [4], [84]. In a process mining project, the study participants are the domain experts involved in the study. Building a trust relation with the experts ensures that there is an honest and open evaluation at the end of the project. Honesty and openness contribute largely to the value generation for both the research and the organisation. Next to that, the researcher gains an understanding of the culture of the organization to interpret the data correctly. Finally, it is of vital importance that a researcher gets a feeling for where misinformation, such as bad data quality, can stem from to account for this during the interpretation and generation of the findings. This prevents the researcher from drawing wrongful conclusions.

A good practice is presented in the process mining literature by [66]. The authors start the qualitative evaluation by presenting the final findings, assumptions made, and the interpretation of findings. In light of the strategy of engagement and understanding of the field, this can be expanded in two ways. First, to *carefully select* the domain experts that are involved in the qualitative evaluation. The domain experts must possess the required knowledge of the project and hold a central position in the organization to collaboratively interpret the results. Second, to *Standardize and discuss the presentation*. Presentations might often be held in process mining projects, but are rarely discussed explicitly in the research article. As a result, the content of such presentations can vary widely. Therefore, we propose to explicitly mention if a presentation is given to domain experts in the evaluation and discuss the content of the presentation. We propose to include at least three points in the presentation as proposed by [66]: (1) final findings, (2) assumptions made, and (3) interpretation of findings. In addition, we advocate to include a fourth part: discuss data quality issues that arose and were tackled during the project. Making this an explicit part of the presentation allows for the domain experts attending the qualitative evaluation to check if all potential data quality issues are addressed, this in line with the identification of misinformation as discussed previously.

TABLE IV
TABLE CONTAINING ALL VALIDATION STRATEGIES, EFFECTS, AND THE PRACTICAL GUIDELINES

Strategy	Effect	Practical guideline
Engage with the field of research	An open and honest evaluation	Carefully select domain experts; include data quality issues in presentation of results
Triangulation	Completeness and consistency of the results	Use multiple quantitative (e.g. simulations) and qualitative evaluation methods (e.g. interviews)
Peer review or external audit	Credibility of the analysis and interpretation	Plan peer reviews to reflect on research design, approach, and results on a regular basis and keep notes of these meetings
Refine work hypothesis	Transparency and soundness of the results	Keep detailed notes on hypotheses, how they are tested, and the final results and use these to guide the evaluation with domain experts
Clarify and normalize bias	Transparency and reliability of the results	Discuss different types of biases in the evaluation or limitation section
Perform member checking	Credibility of the results	Ask interviewees to check the correctness and authenticity of a summarized report of the interview results and interpretation

2) Triangulation: This strategy refers to the use of multiple data sources to study the research problem [4], [85]. Ideally, a *mixed-methods* approach using qualitative and quantitative techniques is taken to increase the validity of the findings. Another possibility is to use multiple data sources within one type of method. In our literature study sample, no example could be found that applied the triangulation strategy to one evaluation goal. The authors of [38] do apply a mixed-methods approach, but they do so by applying one method for each goal. Their quantitative evaluation describes how a confusion matrix, as proposed by [86], is used to measure the performance of the proposed process mining algorithm. The qualitative evaluation aims to validate the quality of the recommendations that the algorithm produces through structured interviews. In order to apply the triangulation strategy in full, the interviews would need to be complemented with another data source to evaluate the quality of recommendations.

3) Peer review or external audit: This strategy refers to reviewing the research process with a reviewer or auditor [4], [87]. The difference between the peer review and the external audit is the connection to the research. An external auditor cannot have any connection to the research, whereas a peer reviewer can have some connection to research. This strategy was proposed by [84] who describe the reviewer as a 'devil's advocate' that checks with the researcher (in process mining, process analyst) how the research is performed. The authors stress the importance of doing these peer reviews on a regular basis during the research and to keep notes of each meeting. In process mining this would be a good strategy to: (1) critically reflect on the research design, (2) to ensure the data is handled in a compliant manner, and (3) to offer opportunities to discuss interpretations of results.

In practice, peer reviewing or external auditing would not be included in an academic article. As such, it is hard to determine the extent to which this strategy is already applied in the process mining community. We propose that the content produced through this strategy can serve as complementary material that should be provided upon request. The external audit works similarly to a peer review of an academic article. However, this audit only focuses on the result generation and interpretation. The external auditor is given access to the data and notes, including the data collected from domain experts. The audit focuses on answering the question: are the findings, conclusions, and interpretations, supported by the data? This is particularly useful when domain experts are involved and data generated by them is interpreted by the process analyst. Performing such an audit or review increases the validity of the research as it gives an outsider a chance to check the interpretation of the process analyst.

4) Refine work hypothesis: This strategy describes how a researcher can use negative case analysis [4], [84], [85]. This is an analysis in which a researcher formulates a hypothesis, and changes it every time a case is encountered that the hypothesis cannot explain. Keeping detailed notes on each hypothesis, the way it is tested, and the final result increases the transparency and soundness of the final results, and insights generated. These notes can help guide the qualitative evaluation in the sense that it can help scope and structure the process of translating the findings into insights with domain experts.

The cyclical nature of hypothesis refinement is a well known and established approach in process mining analysis (see for example [2]) and in certain paradigms (see for example [83]. The concept is usually applied throughout the project. However, to the best of our knowledge, no research has described how to use it during the qualitative evaluation. This strategy can be applied to the work of [66]. In their work, they define two hypotheses for the analysis. An additional hypothesis for the qualitative evaluation would be a good first step to help guide this phase of the project. In addition, transparency into the evolution of the hypotheses provides valuable insights to the research. It makes the considerations of the process analyst explicit. This allows domain experts to validate the process of insight generation that the process analyst has conducted.

5) Clarify biases: Furthermore, another strategy describes the standard practice of reflecting on the possible *biases* from a qualitative perspective in process mining research. Many biases can be present in a study, especially when domain experts are involved. One illustrative example is the familiarity of domain experts with process mining [2], [14], [35]. Depending on the goal of the process mining project, this might influence the qualitative evaluation in different ways. For example, in recommendation systems the experience of the domain experts influences the quality of the output of an algorithm, a domain expert more experienced with process mining can better understand recommendations displayed in a process models, whereas a less experienced domain expert might benefit from recommendations in text form. We know that the more experienced domain experts are with process mining, the better they can interpret the models and derive insights. It should be standard practice to discuss these types of bias in the evaluation or limitations of a process mining project. To exemplify, consider the authors of [35] who already do so: "the assessment of this study would probably be carried out through surveys involving people that possibly do not have the same expertise level about business process management, which may imply biased results.".

6) Member checking: This strategy originated in studies that use interviews as a data collection technique [4], [84], [88]. It describes how researchers, after they have interviewed the participants and analyzed the data, return to the interviewees to confirm that their findings and the interpretation thereof are credible. Looking at the literature study, we can see that a comparable approach is partially deployed by some researchers who aim to confirm their findings through interviews. The advantage of introducing member checking as an approach in process mining is that it introduces rigor by providing a set of guidelines and procedures. Consider for example a specific member checking technique proposed by [89]. The authors describe a five-step plan to increase validity of studies that use interviews: (1) prepare synthesized summary of raw data and the interpretation of the data, (2) formulate criteria to select participants for eligibility for a member check, (3) send member check data and feedback form, (4) gather response data, and (5) integrate response data with raw and interpreted data.

We can apply the member checking strategy on the work of the authors of [10] after the interviews have been performed and analyzed. The process analysts would return to (a part of) the group of interviewees (i.e. members) and provide them with a report on their findings of the interview. The members would then asked to check the authenticity and their comments can be used as input to check for the validity of the interpretation of the interview data. The strategy can also be used when focus groups are used as data collection technique.

V. Conclusion

The involvement of domain experts in process mining projects is essential in translating results into actionable insights. We performed a systematic literature study of recent process mining case studies where domain experts were involved in the evaluation. We found that such evaluations are performed widely, but that there is a lack of structure in *how* they are performed. With that in mind, we present six strategies from the qualitative research field. We show good examples of existing process mining projects where aspects of these strategies are already applied. These six strategies contribute to the process mining community by offering a set of coherent guidelines to perform a more rigorous qualitative evaluation of process mining results with domain experts.

The literature study has been performed using an established systematic review checklist and four researchers were involved in the process. However, the setup may pose limitations. First, our sample of papers purely consists of case studies in which the involvement of a domain expert is mentioned. Although different terms were used, it excludes studies in which an expert was involved but not mentioned explicitly. Second, the sample of papers was exclusively drawn from the Scopus database. Although Scopus has shown to contain the most relevant process mining papers, a small number of additional papers may be found in other databases. Last, a large number of studies do not explicitly describe how they evaluate their findings. As such, we cannot infer anything from their practices. Explicit reporting on the qualitative evaluation aspects is highly recommended to increase the transparency, replicability, and validity of a study.

In future work we want to focus on filling in each layer of research (i.e. from philosophies to data collection and analysis) as defined by [82] for qualitative evaluations in process mining projects. The insights on each of these layers contributes to a more rigorous, and ultimately better, qualitative evaluation in process mining projects.

REFERENCES

[1] F. Emamjome, R. Andrews, and A. H. ter Hofstede, "A case study lens on process mining in practice," in *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*. Springer, 2019, pp. 127–145.

[2] M. L. van Eck, X. Lu, S. J. Leemans, and W. M. van Der Aalst, "Pm2: a process mining project methodology," in *CAiSE*. Springer, 2015, pp. 297–313.

[3] J. W. Creswell, *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage, 2013.

[4] J. W. Creswell and C. N. Poth, *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications, 2016.

[5] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," 2007.

[6] M. Ghasemi and D. Amyot, "Process mining in healthcare: a systematised literature review," *International Journal of Electronic Healthcare*, vol. 9, no. 1, pp. 60–88, 2016.

[7] K. Jackson and P. Bazeley, *Qualitative data analysis with NVivo*. Sage, 2019.

[8] S. C. De Alvarenga, S. Barbon Jr, R. S. Miani, M. Cukier, and B. B. Zarpelão, "Process mining and hierarchical clustering to help intrusion alert visualization," *Computers & Security*, vol. 73, pp. 474–491, 2018.

[9] A. V. Deokar and J. Tao, "Orgminer: A framework for discovering user-related process intelligence from event logs," *Information Systems Frontiers*, pp. 1–20, 2020.

[10] E. V. Epure, P. Martín-Rodilla, C. Hug, R. Deneckère, and C. Salinesi, "Automatic process model discovery from textual methodologies," in *2015 IEEE 9th International Conference on Research Challenges in Information Science (RCIS)*. IEEE, 2015, pp. 19–30.

[11] R. Andrews, C. G. van Dun, M. T. Wynn, W. Kratsch, M. Röglinger, and A. H. ter Hofstede, "Quality-informed semi-automated event log generation for process mining," *Decision Support Systems*, vol. 132, p. 113265, 2020.

[12] E. Benevento, P. M. Dixit, M. F. Sani, D. Aloini, and W. M. van der Aalst, "Evaluating the effectiveness of interactive process discovery in healthcare: A case study," in *BPM*. Springer, 2019, pp. 508–519.

[13] T. Lehto, M. Hinkka, and J. Hollmén, "Focusing business improvements using process mining based influence analysis," in *BPM*. Springer, 2016, pp. 177–192.

[14] M. Sirgmets, F. Milani, A. Nolte, and T. Pungas, "Designing process diagrams–a framework for making design choices when visualizing process mining outputs," in *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*. Springer, 2018, pp. 463–480.

[15] A. Pini, R. Brown, and M. T. Wynn, "Process visualization techniques for multi-perspective process comparisons," in *Asia-Pacific Conference on Business Process Management*. Springer, 2015, pp. 183–197.

[16] M. Song, C. Günther, and W. van der Aalst, "Trace clustering in process mining. in (ardagna, d., mecella, m., yang, j., hrsg.): Business process management workshops. bpm 2008. lnbip 17," 2009.

[17] P. Gorissen, "Business process reporting using process mining, analytic workflows and process cubes: A case study in education," in *Data-Driven Process Discovery and Analysis: 5th IFIP WG 2.6 International Symposium, SIMPDA 2015, Vienna, Austria, December 9-11, 2015, Revised Selected Papers*, vol. 244. Springer, 2017, p. 28.

[18] A. Jalali, "Aspect mining in business process management," in *International Conference on Business Informatics Research*. Springer, 2014, pp. 246–260.

[19] E. Rojas, A. Cifuentes, A. Burattin, J. Munoz-Gama, M. Sepúlveda, and D. Capurro, "Performance analysis of emergency room episodes through process mining," *International journal of environmental research and public health*, vol. 16, no. 7, p. 1274, 2019.

[20] W. Abo-Hamad, "Patient pathways discovery and analysis using process mining techniques: An emergency department case study," in *International Conference on Health Care Systems Engineering*. Springer, 2017, pp. 209–219.

[21] O. A. Johnson, T. B. Dhafari, A. Kurniati, F. Fox, and E. Rojas, "The clearpath method for care pathway process mining and simulation," in *BPM*. Springer, 2018, pp. 239–250.

[22] J. De Weerdt, A. Schupp, A. Vanderloock, and B. Baesens, "Process mining for the multi-faceted analysis of business processes—a case study in a financial services organization," *Computers in Industry*, vol. 64, no. 1, pp. 57–67, 2013.

[23] R. van Langerak, J. M. E. van der Werf, and S. Brinkkemper, "Uncovering the runtime enterprise architecture of a large distributed organisation," in *CAiSE*. Springer, 2017, pp. 247–263.

[24] S. Suriadi, R. Andrews, A. H. ter Hofstede, and M. T. Wynn, "Event log imperfection patterns for process mining: Towards a systematic approach to cleaning event logs," *Information Systems*, vol. 64, pp. 132–150, 2017.

[25] J. Swinnen, B. Depaire, M. J. Jans, and K. Vanhoof, "A process deviation analysis–a case study," in *BPM*. Springer, 2011, pp. 87–98.

[26] M. Dees, M. de Leoni, and F. Mannhardt, "Enhancing process models to improve business performance: A methodology and case studies," in *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*. Springer, 2017, pp. 232–251.

[27] C. D. Ciccio and M. Mecella, "On the discovery of declarative control flows for artful processes," *ACM Transactions on Management Information Systems (TMIS)*, vol. 5, no. 4, pp. 1–37, 2015.

[28] D. Thabet, S. A. Ghannouchi, and H. H. B. Ghezala, "A process mining-based solution for business process model extension with cost perspective context-based cost data analysis and case study," in *IFIP International Conference on Computer Information Systems and Industrial Management*. Springer, 2018, pp. 434–446.

[29] S. Yang, W. Ni, X. Dong, S. Chen, R. A. Farneth, A. Sarcevic, I. Marsic, and R. S. Burd, "Intention mining in medical process: A case study in trauma resuscitation," in *2018 IEEE International Conference on Healthcare Informatics (ICHI)*. IEEE, 2018, pp. 36–43.

[30] H. A. Reijers, M. Song, and B. Jeong, "Analysis of a collaborative workflow process with distributed actors," *Information Systems Frontiers*, vol. 11, no. 3, pp. 307–322, 2009.

[31] T. Baier, J. Mendling, and M. Weske, "Bridging abstraction layers in process mining," *Information Systems*, vol. 46, pp. 123–139, 2014.

[32] D. Ferreira, M. Zacarias, M. Malheiros, and P. Ferreira, "Approaching process mining with sequence clustering: Experiments and findings," in *BPM*. Springer, 2007, pp. 360–374.

[33] D. Knoll, G. Reinhart, and M. Prüglmeier, "Enabling value stream mapping for internal logistics using multidimensional process mining," *Expert Systems with Applications*, vol. 124, pp. 130–142, 2019.

[34] R. S. Mans, M. Schonenberg, M. Song, W. M. van der Aalst, and P. J. Bakker, "Application of process mining in healthcare–a case study in a dutch hospital," in *International joint conference on biomedical engineering systems and technologies*. Springer, 2008, pp. 425–438.

[35] R. Pérez-Castillo, J. A. Cruz-Lemus, I. G.-R. de Guzmán, and M. Piattini, "A family of case studies on business process mining using marble," *Journal of Systems and Software*, vol. 85, no. 6, pp. 1370–1385, 2012.

[36] F. Marazza, F. A. Bukhsh, O. Vijlbrief, J. Geerdink, S. Pathak, M. van Keulen, and C. Seifert, "Comparing process models for patient populations: Application in breast cancer care," in *BPM*. Springer, 2019, pp. 496–507.

[37] D. Duma and R. Aringhieri, "Mining the patient flow through an emergency department to deal with overcrowding," in *International Conference on Health Care Systems Engineering*. Springer, 2017, pp. 49–59.

[38] E. V. Epure, C. Hug, R. Deneckere, and S. Brinkkemper, "What shall i do next?" in *CAiSE*. Springer, 2014, pp. 473–487.

[39] M. L. van Eck, N. Sidorova, and W. M. van der Aalst, "Enabling process mining on sensor data from smart products," in *2016 IEEE Tenth International Conference on Research Challenges in Information Science (RCIS)*. IEEE, 2016, pp. 1–12.

[40] L. T. Ly, C. Indiono, J. Mangler, and S. Rinderle-Ma, "Data transformation and semantic log purging for process mining," in *CAiSE*. Springer, 2012, pp. 238–253.

[41] C. Lee, K. L. Choy, G. T. Ho, and C. H. Lam, "A slippery genetic algorithm-based process mining system for achieving better quality assurance in the garment industry," *Expert systems with applications*, vol. 46, pp. 236–248, 2016.

[42] P. Taylor, M. Leida, and B. Majeed, "Case study in process mining in a multinational enterprise," in *International Symposium on Data-Driven Process Discovery and Analysis*. Springer, 2011, pp. 134–153.

[43] V. Denisov, D. Fahland, and W. M. van der Aalst, "Unbiased, fine-grained description of processes performance from event data," in *BPM*. Springer, 2018, pp. 139–157.

[44] A. Syamsiyah, A. Bolt, L. Cheng, B. F. Hompes, R. J. C. Bose, B. F. van Dongen, and W. M. van der Aalst, "Business process comparison: A methodology and case study," in *International Conference on Business Information Systems*. Springer, 2017, pp. 253–267.

[45] E. Mahendrawathi, H. M. Astuti, and A. Nastiti, "Analysis of customer fulfilment with process mining: A case study in a telecommunication company," *Procedia Computer Science*, vol. 72, pp. 588–596, 2015.
[46] E. Mahendrawathi, H. M. Astuti, and I. R. K. Wardhani, "Material movement analysis for warehouse business process improvement with process mining: a case study," in *Asia-Pacific Conference on Business Process Management*. Springer, 2015, pp. 115–127.
[47] M. Rovani, F. M. Maggi, M. De Leoni, and W. M. van der Aalst, "Declarative process mining in healthcare," *Expert Systems with Applications*, vol. 42, no. 23, pp. 9236–9251, 2015.
[48] E. Mahendrawathi, S. O. Zayin, and F. J. Pamungkas, "Erp post implementation review with process mining: A case of procurement process," *Procedia Computer Science*, vol. 124, pp. 216–223, 2017.
[49] M. Cho, M. Song, and S. Yoo, "A systematic methodology for outpatient process analysis based on process mining," in *Asia-Pacific Conference on Business Process Management*. Springer, 2014, pp. 31–42.
[50] N. Mueller-Wickop and M. Schultz, "Erp event log preprocessing: timestamps vs. accounting logic," in *International Conference on Design Science Research in Information Systems*. Springer, 2013, pp. 105–119.
[51] S. Hessey and W. Venters, "Sensing distress–towards a blended method for detecting and responding to problematic customer experience events," in *International Conference on HCI in Business, Government, and Organizations*. Springer, 2016, pp. 395–405.
[52] F. Prathama, B. N. Yahya, D. D. Harjono, and E. Mahendrawathi, "Trace clustering exploration for detecting sudden drift: A case study in logistic process," *Procedia Computer Science*, vol. 161, pp. 1122–1130, 2019.
[53] R. S. Mans, H. A. Reijers, D. Wismeijer, and M. van Genuchten, "A process-oriented methodology for evaluating the impact of it: A proposal and an application in healthcare," *Information Systems*, vol. 38, no. 8, pp. 1097–1115, 2013.
[54] M. Stuit and H. Wortmann, "Discovery and analysis of e-mail-driven business processes," *Information Systems*, vol. 37, no. 2, pp. 142–168, 2012.
[55] R. Andrews, M. T. Wynn, K. Vallmuur, A. H. Ter Hofstede, and E. Bosley, "A comparative process mining analysis of road trauma patient pathways," *International journal of environmental research and public health*, vol. 17, no. 10, p. 3426, 2020.
[56] B. R. Gunnarsson, S. K. vanden Broucke, and J. De Weerdt, "Predictive process monitoring in operational logistics: A case study in aviation," in *BPM*. Springer, 2019, pp. 250–262.
[57] B. F. Hompes, A. Maaradji, M. La Rosa, M. Dumas, J. C. Buijs, and W. M. van der Aalst, "Discovering causal factors explaining business process performance variation," in *CAiSE*. Springer, 2017, pp. 177–192.
[58] M. Park, M. Song, T. H. Baek, S. Son, S. J. Ha, and S. W. Cho, "Workload and delay analysis in manufacturing process using process mining," in *Asia-Pacific Conference on Business Process Management*. Springer, 2015, pp. 138–151.
[59] T. Vogelgesang and H.-J. Appelrath, "Pmcube: a data-warehouse-based approach for multidimensional process mining," in *BPM*. Springer, 2016, pp. 167–178.
[60] C. Li, M. Reichert, and A. Wombacher, "Mining business process variants: Challenges, scenarios, algorithms," *Data & Knowledge Engineering*, vol. 70, no. 5, pp. 409–434, 2011.
[61] Y. Wang, F. Caron, J. Vanthienen, L. Huang, and Y. Guo, "Acquiring logistics process intelligence: Methodology and an application for a chinese bulk port," *Expert Systems with Applications*, vol. 41, no. 1, pp. 195–209, 2014.
[62] A. P. Kurniati, C. McInerney, K. Zucker, G. Hall, D. Hogg, and O. Johnson, "A multi-level approach for identifying process change in cancer pathways," in *BPM*. Springer, 2019, pp. 595–607.
[63] H. A. Reijers, M. Song, and B. Jeong, "On the performance of workflow processes with distributed actors: does place matter?" in *BPM*. Springer, 2007, pp. 32–47.
[64] A. Partington, M. Wynn, S. Suriadi, C. Ouyang, and J. Karnon, "Process mining for clinical processes: a comparative analysis of four australian hospitals," *ACM Transactions on Management Information Systems (TMIS)*, vol. 5, no. 4, pp. 1–18, 2015.
[65] A. Maris, R. Bijvank, and P. Ravesteyn, "The applicability of process mining to determine and align process model descriptions." in *Bled eConference*, 2016, p. 43.
[66] C. Alvarez, E. Rojas, M. Arias, J. Munoz-Gama, M. Sepúlveda, V. Herskovic, and D. Capurro, "Discovering role interaction models in the emergency room using process mining," *Journal of biomedical informatics*, vol. 78, pp. 60–77, 2018.
[67] S. Astromskis, A. Janes, and M. Mairegger, "A process mining approach to measure how users interact with software: an industrial case study," in *Proceedings of the 2015 International Conference on Software and System Process*, 2015, pp. 137–141.
[68] L. van den Ingh, R. Eshuis, and S. Gelper, "Assessing performance of mined business process variants," *Enterprise Information Systems*, vol. 15, no. 5, pp. 676–693, 2021.
[69] R. Pérez-Castillo, B. Weber, J. Pinggera, S. Zugal, I. G.-R. de Guzmán, and M. Piattini, "Generating event logs from non-process-aware systems enabling business process mining," *Enterprise Information Systems*, vol. 5, no. 3, pp. 301–335, 2011.
[70] E. A. Elhadjamor and S. A. Ghannouchi, "Analyze in depth health care business process and key performance indicators using process mining," *Procedia Computer Science*, vol. 164, pp. 610–617, 2019.
[71] E. Roubtsova and N. Wiersma, "Involvement of business roles in auditing with process mining," in *International Conference on Evaluation of Novel Approaches to Software Engineering*. Springer, 2018, pp. 24–44.
[72] S. Erdem and O. Demirörs, "An exploratory study on usage of process mining in agile software development," in *International Conference on Software Process Improvement and Capability Determination*. Springer, 2017, pp. 187–196.
[73] D. Calvanese, T. E. Kalayci, M. Montali, and S. Tinella, "Ontology-based data access for extracting event logs from legacy data: the onprom tool and methodology," in *International Conference on Business Information Systems*. Springer, 2017, pp. 220–236.
[74] T. Baier and J. Mendling, "Bridging abstraction layers in process mining: Event to activity mapping," in *Enterprise, Business-Process and Information Systems Modeling*. Springer, 2013, pp. 109–123.
[75] P. M. Dixit, H. Verbeek, J. C. Buijs, and W. M. van der Aalst, "Interactive data-driven process model construction," in *International Conference on Conceptual Modeling*. Springer, 2018, pp. 251–265.
[76] N. Martin, A. Martinez-Millana, B. Valdivieso, and C. Fernández-Llatas, "Interactive data cleaning for process mining: a case study of an outpatient clinic's appointment system," in *BPM*. Springer, 2019, pp. 532–544.
[77] R. Pérez-Castillo, M. Fernández-Ropero, and M. Piattini, "Business process model refactoring applying ibuprofen. an industrial evaluation," *Journal of Systems and Software*, vol. 147, pp. 86–103, 2019.
[78] M. Cho, M. Song, S. Yoo, and H. A. Reijers, "An evidence-based decision support framework for clinician medical scheduling," *IEEE Access*, vol. 7, pp. 15 239–15 249, 2019.
[79] J. Samalikova, R. J. Kusters, J. J. Trienekens, and A. Weijters, "Process mining support for capability maturity model integration-based software process assessment, in principle and in practice," *Journal of Software: Evolution and Process*, vol. 26, no. 7, pp. 714–728, 2014.
[80] M. Montali, F. M. Maggi, F. Chesani, P. Mello, and W. M. v. d. Aalst, "Monitoring business constraints with the event calculus," *ACM transactions on intelligent systems and technology (TIST)*, vol. 5, no. 1, pp. 1–30, 2014.
[81] R. Engel, W. M. van der Aalst, M. Zapletal, C. Pichler, and H. Werthner, "Mining inter-organizational business process models from edi messages: A case study from the automotive sector," in *CAiSE*. Springer, 2012, pp. 222–237.
[82] M. Saunders, P. Lewis, and A. Thornhill, *Research methods for business students*. Pearson education, 2009.
[83] A. R. Hevner, S. T. March, J. Park, and S. Ram, "Design science in information systems research," *MIS quarterly*, pp. 75–105, 2004.
[84] Y. Lincoln, Y. Guba, E. Guba, and S. Publishing, *Naturalistic Inquiry*. SAGE Publications, 1985. [Online]. Available: https://books.google.nl/books?id=2oA9aWlNeooC
[85] M. Q. Patton, "Enhancing the quality and credibility of qualitative analysis." *Health services research*, vol. 34, no. 5 Pt 2, p. 1189, 1999.
[86] F. Provost and R. Kohavi, "Guest editors' introduction: On applied research in machine learning," *Machine learning*, vol. 30, no. 2, pp. 127–132, 1998.
[87] M. B. Miles and A. M. Huberman, *Qualitative data analysis: An expanded sourcebook*. sage, 1994.
[88] R. E. Stake, *The art of case study research*. sage, 1995.
[89] L. Birt, S. Scott, D. Cavers, C. Campbell, and F. Walter, "Member checking: a tool to enhance trustworthiness or merely a nod to validation?" *Qualitative health research*, vol. 26, no. 13, pp. 1802–1811, 2016.