Auditors' categorization of process deviations

Marzie Hosseinpour[1], Mieke Jans

Abstract

In addressing control deficiencies, auditors increasingly rely on data analytics. Despite the need to align information presentation with auditors' cognitive structures, scant scholarly attention is given to how auditors internally categorize process deviations. This study investigates experienced auditors' categorization of 62 deviations, revealing three primary categories: missing, reordered, and duplicated activities. These insights inform the development of active-learning algorithms, aligning with auditors' knowledge structures to mitigate redundant processing risks. Blindly adopting process management research outcomes, however, poses a risk to auditing quality, impacting both effectiveness and efficiency in risk assessment and control testing. This research highlights the importance of validating and aligning deviation categories with auditors' nuanced interpretations to enhance audit tools' efficacy.

Keywords: Auditing, process deviations, process mining, deviation categories

1. Introduction

Being in control of the core business processes does not only support an organisation in creating customer value in an efficient way, it is also the key to managing risk. That is why an auditor has a keen interest in understanding the client's processes as a way to understand the client's environment and associated risks. This has been recognized in international standards (example given in ISA 315 of the International Auditing and Assurance Standards Board

[1]Corresponding author: Marzie Hosseinpour

(IAASB, 2019) and Auditing Standard No. 12 of the Public Company Accounting Oversight Board (PCAOB, 2010)).

When analyzing an organization's functioning, the auditor is interested in identifying deviations. "To which extent are employees deviating from the prescribed procedures?" In order to formulate an answer to this question, process mining techniques can be used. Process mining refers to the collection of tools and techniques to analyze processes in an evidence-based way (van der Aalst, 2016). Starting from the logged transactions in an information system (like an Enterprise Resource Planning (ERP) system), it reconstructs the true process enactments. This way, deviations from the normative procedure can be identified, signaling potential control deficiencies (Werner et al., 2021; Chiu and Jans, 2019; Jans and Hosseinpour, 2019).

Notwithstanding the promising added value of process mining for auditing purposes, there is a key limitation of applying these techniques that prohibits a full integration in audit engagements. Creating such a high level of transparency has the downside of revealing an immense number of deviations, which makes follow-up cumbersome (Jans et al., 2013, 2014; Chiu and Jans, 2019). This phenomenon of 'alarm floods' is not unique to process mining, but emerges when applying data analysis in general (Alles et al., 2008). The framework of Jans and Hosseinpour (2019), the 'Catalyst approach', presents a theoretical solution that incorporates an artificial intelligence algorithm with a human-in-the-loop. However, a prerequisite of this solution is to present the deviations to the auditor in different categories. The auditor will then feed the algorithm with expert knowledge by labeling the presented deviation as anomalous or cleared.

To date, we have no knowledge on how process deviations are processed by auditors, or even by humans in general. Some studies present process deviations categories, but these are (mainly theoretical) suggestions and have never been validated through an independent study (Weber et al., 2008; Adriansyah et al.,

2011; García-Bañuelos et al., 2015). This paper provides empirical insights into the categories that auditors use when they are confronted with process deviations. These insights are imperative for further advancement of human-in-the-loop solutions to deal with the numerously identified deviations. Namely, we know from previous studies on information presentation, that it is essential to present information in such a way that it fits best with the task at hand (Vessey and Galletta, 1991; Nelson et al., 1995; Bonner et al., 1997; Kopp and O'Donnell, 2005; Davern et al., 2012; Dunn et al., 2017).

This paper reports on a study that answers the following research question: "Which categories do auditors use for grouping process deviations?". To address this question, a field study is conducted. Fourteen auditors are presented a total of 62 recorded process executions, along with a normative process model against which to compare the deviations. They are invited to elaborate on what they are confronted with. A verbal content analysis of their responses is executed to identify the categories that auditors apply. The identified categories are mapped to the theoretical categories that were found in the literature. In general, there are three process deviation categories that auditors use: activities are *missing*, *reordered*, or *duplicated*. These categories overlap three theoretical categories from literature (*skipping*, *replacing*, and *repeating*), but they are not a direct translation. The remaining categories from literature in other fields (business process management and process mining) are either not used (such as *resorting to* or *inserting*) or they are included in another category (*swapping* and *executing in loop*). Our findings are a validation of the categories that were used by Chiu and Jans (2019) when they applied process mining for the evaluation of internal control effectiveness and intuitively grouped process deviations as well.

The remainder of this paper is organized as follows. Section 2 provides some background on deviation detection in the context of auditing and on process deviation categories. The methodology and research design of our study are

described in section 3. Sections 4 and 5 present the results and a discussion of our study. The paper is concluded in section 6.

2. Background

In this section, some background is provided on deviation detection in the context of auditing. Next, conformance checking is shortly introduced. The section concludes with describing process deviation categories as encountered in the fields of process mining and business process management.

2.1. Deviation detection in a continuous auditing context

The advent of data analysis techniques supports the auditing profession in providing assurance in real-time –or at least on a frequent basis– and enables analysis of the full population of business processes and data, known as continuous auditing (Kuenkaikaew and Vasarhelyi, 2013; Chan and Vasarhelyi, 2011).[1]

Within the continuous auditing domain, most attention has been given to techniques that discover deviations and respective alarming systems. Examples of studies that analyzed the full population of transactions with deviation detection techniques can be found in Groomer and Murthy (1989); Vasarhelyi and Halper (1991); Kogan et al. (1999); Dull et al. (2006); Kim and Vasarhelyi (2012); Thiprungsri and Vasarhelyi (2011). As mentioned before, one of the key problems of those techniques is the immense set of deviations they detect, which makes the follow-up cumbersome for auditors (Kogan et al., 1999; Alles et al., 2006a; Debreceny et al., 2003; Alles et al., 2008; Perols and Murthy, 2012; Kim and Vasarhelyi, 2012). Moreover, most of those techniques present the deviations in a too fine-grained format (Russell et al., 2006; Hosseinpour and Jans,

[1]For more explanation of the need of more frequent auditing, the interested readers can refer to Debreceny and Rahman (2005); Flowerday and Von Solms (2005); Elliott (2002); Braun and Davis (2003); Vasarhelyi et al. (2002).

2016). Those characteristics make the investigation of deviations overwhelming. To deal with the immense number of detected deviations, some approaches have been proposed in the continuous auditing literature. In general, two types of approaches have been suggested. One way to approach the alarm flood is by setting a limit about when to raise an alarm. Only when a certain threshold is reached, the alarm is raised (Alles et al., 2006a, 2008; Kim and Vasarhelyi, 2012; Issa, 2013; Li et al., 2015). Another way to deal with the flood of alarms is by grouping deviations (Perols and Murthy, 2012; Hosseinpour and Jans, 2016). Grouping the alarms is a first step in a staged approach. Merely grouping the alarms won't reduce the number of alarms, but when grouped intelligently, facilitates the follow-up investigation. In particular, the 'Catalyst approach' that requires the auditor to interpret a deviation and transfer its knowledge to an intelligent learning algorithm, benefits from presenting the deviations in categories that are aligned with the knowledge structures of the auditor.

2.2. Conformance checking

Conformance checking is a type of process mining, a broader research field that generates techniques and tools to analyze processes in an evidence-based way. Process executions, as logged in the information system, are collected in a specific format, known as an event log (Carmona et al., 2022). Each process execution is captured as a sequence of activities which took place at a certain time, also known as the trace. By using conformance checking techniques, the process executions are compared against a normative process model or against a set of business rules. The goal of these techniques is to find the degree of agreement between process executions and a model, and to detect the differences between them.

The output of conformance checking tools is typically expressed as follows: "activity C was skipped in trace number 1365," or "activity G was inserted in trace number 1365." Depending on the tool, this information may also be aggregated. All cases that showcase a certain deviation, like "activity G was

inserted", are then grouped into one bucket to investigate later.

In the field of Business Process Management, discovering these differences is mostly used to evaluate or to improve the process models (e.g., De Weerdt et al. (2012); Mendling et al. (2007); Fahland and van der Aalst (2015)). However, it can also be used to discover the process executions that do not conform to the designated model, i.e., the deviating cases. For auditing purposes, only this latter application of conformance checking techniques is relevant Jans et al. (2013).

2.3. Cognitive information processing

While previous studies confirm that taking a process angle when performing audit tasks positively influences performance (O'Donnell and Schultz Jr, 2003; Kopp and O'Donnell, 2005; Bierstaker et al., 2009), other studies investigated process representation's importance in an auditing setting. They investigated whether differences in performance arise when diagrammatic or textual process representations are used (Dunn and Gerard, 2001; Kelton et al., 2010; Boritz et al., 2012; Ritchi et al., 2020). Presenting information on process deviations to an auditor who must decide whether they hold risks to material financial misstatements ties into the field of human information processing.

In the "Paradigm of Cognitive Fit" mental representations play a crucial role (Davern et al., 2012). This theory links the (external) presentation of the problem and the problem-solving task to an (internal) mental representation (Vessey and Galletta, 1991). This connection, along with the interaction between them, serves as input for the mental representation to develop the solution of the task (Shaft and Vessey, 2006). The tighter the fit between the mental representations and the task, the better the performance in solving the task. The connection between representations and performance has been investigated in depth in the accounting and auditing domain (Libby and Luft, 1993; Bierstaker et al., 1999; Dunn and Grabski, 2001; Dunn et al., 2017; Ritchi et al., 2020). Based on this

literature we postulate that in order to increase the quality of an auditor's risk assessment and control testing, the process deviations must be presented in a way that best fits the auditor's judgment and expert knowledge.

Besides the presentation of information, the amount of information is important. Humans have limited cognitive processing capacity, so when confronted with a task that requires too much capacity, experts divide the task into subtasks (Sweller, 1988). In the context of processing thousands of process deviations, this corresponds to grouping those deviations into categories that are meaningful to the auditor (Hosseinpour and Jans, 2016). After grouping these deviations, follow-up investigations have to take place. This study investigates whether there are categories that auditors use to group process deviations, in order to start follow-up investigations with categories that are in synch with the auditor's mental model.

2.4. Process deviation categories in literature

In the research field of conformance checking and in the broader field of Business Process Management, three studies present a set of process deviation categories, albeit under different terminology. The first study proposes certain combinations of skipping and inserting activities that result in the deviation patterns *replacement*, *swapping*, and *repetition*,(Adriansyah et al., 2010).

A second study focuses on verbalizing the detected deviations in a set of natural language statements (García-Bañuelos et al., 2015). The authors propose nine deviation patterns.[2] To date, their approach is the only one that provides deviation patterns at a higher level in natural language. Four of those

[2]The nine deviation patterns are *immediate causality-concurrency, immediate concurrency-conflict, task skipping, unmatched repetition, task substitution, task relocation, task absence/insertion and unobserved acyclic interval*, and *unobserved cyclic interval*

patterns relate to characteristics of the model, and they are not dedicated to identifying process executions that deviate from the normative model. Consequently, only the remaining five patterns are considered relevant to our study: *task skipping*, *unmatched repetition*, *task substitution*, *task relocation*, and *task absence/insertion*. Both first and second study suggest these deviation groups from a theoretical point of view. There is no input from an application domain used to construct these, neither have they been validated in a follow-up study.

A third study, by Weber et al. (2008), proposes a comprehensive set of change patterns for comparing business processes. They suggest 18 change patterns from a control-flow perspective.[3] The sources for the suggested patterns are case studies that were executed before in the healthcare domain and the automotive domain. To our knowledge, these patterns were not validated in a different, more generic setting. Among the change patterns, only six are relevant for observable deviations in event log traces. They are: *insert process fragment, delete process fragment, replace process fragment, swap process fragment, copy process fragment*, and *embed process fragment in loop*.

In the context of process mining in auditing, Chiu and Jans (2019) manually investigated a complete set of process deviations. They present their findings according to the three categories of process deviations they identified: *missing activities*, *activities not in the right order*, and *redundant activities*. Since these categories were raised while investigating process deviations in the specific context that is subject of this study, it is of interest to check whether the categories that auditors use resonate with this suggestion of deviation categories.

[3]The 18 change patterns are *insert process fragment*, *delete process fragment*, *replace process fragment*, *swap process*, *copy process fragment*, *extract sub-process, in line sub-process, embed an existing fragment in a loop, parallelize a process fragment, embed an existing process fragment in a conditional branch*, *add control dependencies, remove control dependencies, and update transition conditions*.

Theoretical category	Deviation Pattern by Adriansyah et al. (2010)	Mismatch by García-Bañuelos et al. (2015)	Change Pattern by Weber et al. (2008)	Non-Standard Variant Category by Chiu and Jans (2019)
skip an activity	suppression	task skipping	delete process fragment	missing activity
insert an extra activity	insert	task absence/insertion	insert process fragment	
replace one activity by another	replacement	task substitution	replace process fragment	
swap two activities	swapping	task relocation	swap process fragment	activities not in the right order
repeat an activity	insert	unmatched repetition	copy process fragment	redundant activity
execute an activity in a loop	repetition	–	embed process fragment in loop	

Table 1: Comparison of the deviation categories proposed in literature

Regardless of the different terminology that is used in the four studies, there is a general overlap in meaning. Table 1 presents an overview of all relevant deviation categories as suggested by Adriansyah et al. (2010); García-Bañuelos et al. (2015); Weber et al. (2008) and Chiu and Jans (2019) and how they relate to each other. Most relations are straightforward, except for a different interpretation of "insert" by Adriansyah et al. (2010) and Weber et al. (2008). The latter makes a distinction between inserting and copying a process fragment, where the former categorizes both these patterns as "insert." García-Bañuelos et al. (2015) is more in line with Weber et al. (2008), also making a distinction. Therefore, we include this more detailed split-up in inserting and repeating an activity in our overview. The first column in Table 1, Theoretical category, is the overarching description that we use throughout this paper when referring to these patterns.

3. Methodology and research design

In this section our methodology for investigating auditors' categories of process deviations is elaborated on. Both the methodological approach and the instrument design are described.[4]

3.1. Field research methodology

To reveal process deviation categories that are used by auditors, auditors are presented a normative process model, along with a set of process deviations, and were asked to interpret the deviations. Our study is set up along the guidelines of Glaser and Strauss (1968); Corbin and Strauss (1990) to conduct a rigorous analysis that includes the scrutinization of transcripts. It allows to identify patterns and themes and to modify and extend initial categories and subcategories in a hierarchical manner.

[4]Institutional approval for this study was received.

Starting from the verbal descriptions of the auditors, we discover the deviation categories that they rely on when assessing these deviations. This inductive (bottom-up) construction of a theory is in line with other field research performed in the auditing domain (Wilson, 2014; Malsch and Salterio, 2015).

3.1.1. Theoretical sampling and data saturation

To select the auditors to interview, a theoretical sampling approach is applied. In this method, some data are collected from the field, and they are analyzed and coded by the researcher. Based on the coding, more data are selected from the field (more interviews are taken) and analyzed as before (Willig, 2013). These back-and-forth steps between collecting and analyzing data continue until data saturation is achieved.

For our study, we target 20 auditors in our sampling plan, which is in line with comparable studies (Malsch and Salterio, 2015). An interview request is sent to more than 40 experienced auditors in Belgium and The Netherlands, with responses from 14 auditors. Data saturation occurs when more data from the sample does not lead to more information regarding the research question (Seale, 1999). Analyzing the interview data shows that no new interpretations (themes) emerged after the fourth interview (see Figure 1), indicating data saturation from the fifth interview. Consequently, we did not gather additional interviews after having analysed the interviews with the 14 auditors.

Table 2 provides information on the sample composition and interview statistics. Nine interviewed auditors are employed at Big Four auditing firms, and five are employed as internal auditors at multinational companies. In total 12, interviews are conducted (two interviews were with two auditors at the same time). Although the objectives of internal and external auditors are not fully equivalent to each other, there is a considerable overlap when it comes to test-

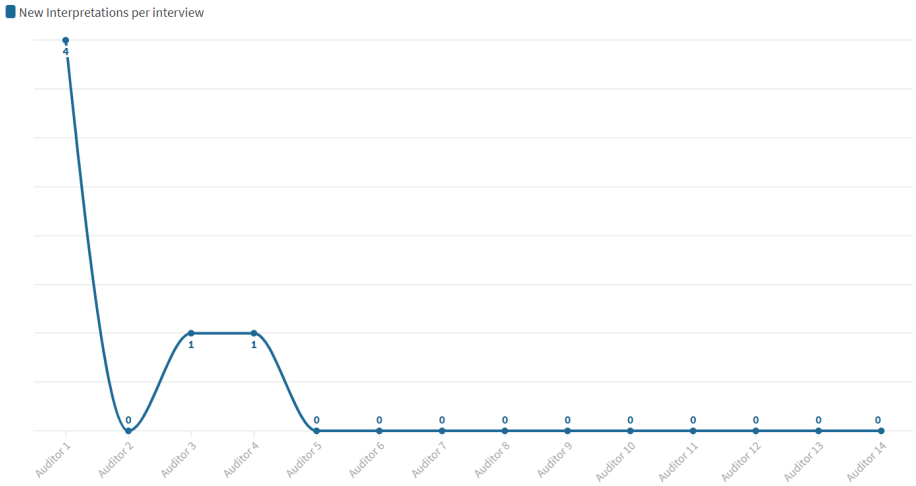

Figure 1: Data saturation is reached after the fourth interview, since no new concept is discovered after that interview.

ing internal controls. An internal auditor might primarily be concerned with operational efficiency; still, the effectiveness of the controls fall under his or her responsability. The external auditor is interested in the internal controls since a deficiency might have a severe impact on the financial statements. Provided the common interest in internal controls, we included both types of auditors in our sample and ran some post-analysis tests to see whether differences in interpretations exist.

The focus of our auditor sample is on managers and senior managers with at least five years of experience. Those below this level may not have enough experience yet. We concur that partner profiles are presumably not the people (anymore) to look at low-level details of process deviations. However, their (and the manager's) expert level of domain knowledge is the subject of our study on how auditors interpret relevant process information. Following Agrawal et al. (1996), inexperienced auditors would not draw upon user schemas that structure their knowledge. They rather form mental representations based on the task and tool at hand. Given the objective of our study to uncover the knowledge

constructs of auditors such that new information can be presented as efficiently as possible, we exclude junior profiles from the study.

Each interview is designed to last approximately one hour. However, the interviews vary in length from 0:51" to 1:44". They are conducted over a 12-week period from April 2017 to June 2017. The period is chosen to be after the auditors' "busy period."

		Number of auditors	Median years of experience	Mean interview time
External auditors	Manager	4	6.75	1:02'
	Senior manager	3	11.33	1:04'
	Director	1	14	1:17'
	Partner	1	27	1:06'
Internal auditors	-	5	9	1:30'

Table 2: Interview statistics

3.1.2. Interview design

We organize our data collection through semi-structured interviews, since additional questions can be asked to probe further the line of responses where appropriate. An introduction with the objective and the context of the interview, along with the normative process model and an example of a process execution ($CreatePurchaseRequest \rightarrow Approve \rightarrow CreatePurchaseOrder \rightarrow Sign \rightarrow Pay \rightarrow GoodsReceipt \rightarrow InvoiceReceipt \rightarrow Pay$) is sent to the auditors prior to the interview. At the start of the interview, when the interviewer introduces herself and her research, she briefly explains the objectives of the study (e.g., the fact that only the order of activities is being considered).

The interviewees are told that the study is based on their experience, hence, there are no right or wrong answers. They are also told that the information they give will only be applied for academic purposes and their personal information and data will be treated confidentially. The interviewees are asked if there was anything confusing in the materials they received in advance. If something was unclear, further explanation is given. Then, general instructions on how the interview is structured are provided. Last, the auditors are asked for their permission to record the interview. The protocol of the interview can be found in appendix.

3.1.3. Conceptual coding

All interviews are recorded and transcribed by the interviewing researcher, using the *Otranscribe* tool.[5] The coding and analysis of interview data are performed Nvivo 11.[6]

The interview data are analyzed in two phases. In the first-order analysis, all data from the interviews is coded by a single researcher. In Vivo coding is applied, coding the actual phrases that are used by the interviewees themselves (Strauss, 1987) [7]. This approach supports the identification of the *concepts* used by auditors and avoids inserting existing theory into the analysis (Stringer, 2014; Willig, 2013). To reorganize the first-order codes (the concepts) into a list of categories that represent the auditor's deviationg groups, code mapping is applied. It is a transition between the first and the second coding phases (Saldaña, 2016).

[5] http://otranscribe.com/

[6] http://www.qsrinternational.com/nvivo/nvivo-products

[7] In Vivo coding is also known as literal coding, natural coding, inductive coding, verbatim coding, or indigenous coding (Saldaña, 2016). In the In vivo coding the label, which is a word or phrase, is taken from the data itself. In other words, the participant's own language is chosen as the label.

In the second-order analysis, similarities and differences between the concepts of the first-order coding are considered, such that the concepts are compressed into *themes* using conceptual coding (Gioia et al., 2013). This grounded theory approach, as described by Glaser and Strauss (1968); Corbin and Strauss (1990), allows revealing insights without tying them upfront to a theoretical framework.

3.2. Instrument design

To identify which deviation categories auditors use when provided with information on process deviations, we provide every auditor with a normative process model and a set of deviating process executions. This reflects the real-life setting of conformance checking as much as possible. Also in a real-life setting, the auditor starts from a normative process model -the desired situation- and compares deteced deviations -the real behaviour- against this model. To this end, we designed a purchase-to-pay process model and a set of deviating procurement transactions. In this section we explain how these instruments are developed.

3.2.1. Normative process model

To discuss process deviations, a normative process model is required. To make sure the process (model) itself does not become the subject of investigation, a traditional purchase-to-pay process is selected. The purchase-to-pay (P2P) process is a well-known process within organizations, and auditors are familiar with it. For process modeling language, BPMN (rather than Petri net) is chosen because of its simplicity and ease of understanding.

The P2P process design is highly dependent on a company's internal processes, and it can become complex when considering all details. Therefore, we design three simple models to represent a generic picture of a procurement process. The three models are based on real P2P event logs, and they are in line with the P2P models as proposed in Jans et al. (2014) and Sadiq et al. (2007). These models are created in such a way that they allow for discussions of all

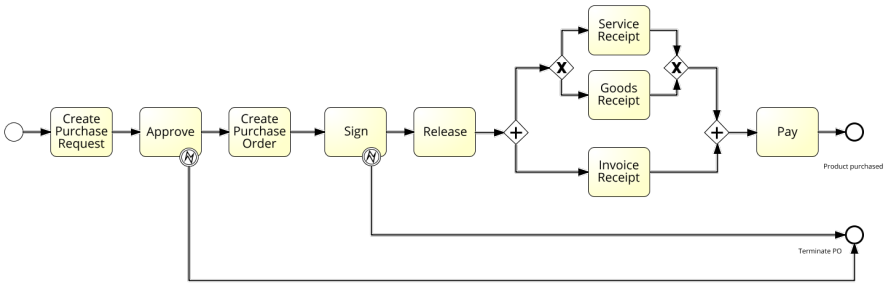

Figure 2: Normative process model of a procure-to-pay process, as used throughout discussions with auditors.

deviation categories that were identified in the literature. At the same time, they maintain an acceptable level of simplicity.

To design a model that both meets the requirements described above and aligned with the auditors' background, we requested input from experts. The models are sent to three auditors that are not included in the interview sample, along with some questions on the extent and complexity of the models. Based on the feedback from auditors, we create the model in Figure 2 as the normative process model in this study.

By using a procedural model as a normative process model, we have two advantages compared with starting with a set of predefined rules (as in Kuhn Jr and Sutton (2010); Alles et al. (2006a); Issa (2013); Li et al. (2015); Alles et al. (2006b, 2008)). First, it skips the time-consuming rules definition procedure prior to analyzing the transaction data. Second, although the rules are defined by a process expert panel, it is still possible that some potential risky behavior is not considered due to human error (False negatives). In comparing event logs with the corresponding procedural normative model, the number of false positive cases might be higher than in the case of comparing event log with a predefined set of rules. However, we use the procedural normative models to avoid the false negative cases, which are critical in auditing. In fraud detection,

it is a well-known fact that a false negative error is usually more costly than a false positive error (Phua et al., 2010).

3.2.2. Set of deviating process executions

Aside from the normative process model, we develop a set of process executions (or traces[8]) that are not according to the model. These types of executions are highlighted by conformance checking algorithms, and, therefore, represent the information we aim to investigate.

The set of deviating process executions is developed by the authors and is based on the categories from literature. It should be noted that not all possible combinations of (eight) activities and (six) deviation categories are included. To select which combinations of activities and deviations are included in the set of traces, we rely again on theoretical sampling (Eisenhardt, 1989). Two criteria were applied to include a combination into the set of traces. First, the trace must be unique and informative, allowing for generalization of the findings. Second, the examples must be realistic, rather than presenting abstract cases that are never encountered in real engagements. Therefore, a real-life event log of a P2P process is used as guidance to select realistic and representive process devaiations. The set of deviating traces that was used during the interviews can be found in Table 3.

As can be seen in most of the process executions traces in this set, the deviating trace can be interpreted in different ways. For instance, the deviation in trace number 13 can be interpreted in three ways: i) $Release$ is replaced by Pay (1 deviation), ii) $Release$ is skipped and Pay is inserted (two deviations), or iii) $Release$ is skipped, $< GoodsReceipt \rightarrow InvoiceReceipt >$ and Pay are swapped, and Pay is repeated (three deviations). To verify whether all deviation categories from literature are represented in the set of traces, we consider

[8]"Traces" or "sequences" are used interchangeably throughout this paper.

Trace no.	Deviating Purchase-to-pay trace
1	< CreatePR → ApprovePR → CreatePO → SignPO → Release → IR → SR → ModifyPO → ModifyPO → Pay >
2	$< CreatePR \rightarrow ApprovePR \rightarrow CreatePO \rightarrow SignPO \rightarrow ModifyPO \rightarrow Release \rightarrow GR \rightarrow IR \rightarrow Pay >$
3	$< CreatePR \rightarrow ApprovePR \rightarrow Pay \rightarrow CreatePO \rightarrow SignPO \rightarrow Release \rightarrow GR \rightarrow IR \rightarrow Pay >$
4	$< CreatePR \rightarrow ApprovePR \rightarrow CreatePO \rightarrow SignPO \rightarrow Release \rightarrow GR \rightarrow SR \rightarrow IR \rightarrow Pay \rightarrow Pay >$
5	$< CreatePR \rightarrow CreatePO \rightarrow SignPO \rightarrow Release \rightarrow GR \rightarrow IR >$
6	$< CreatePR \rightarrow ApprovePR \rightarrow GR \rightarrow IR \rightarrow Pay >$
7	$< CreatePR \rightarrow ApprovePR \rightarrow CreatePO \rightarrow SignPO \rightarrow GR \rightarrow IR \rightarrow Pay >$
8	$< CreatePR \rightarrow ApprovePR \rightarrow CreatePO \rightarrow SignPO \rightarrow CreatePO \rightarrow SignPO \rightarrow Release \rightarrow SR \rightarrow SR \rightarrow IR \rightarrow Pay \rightarrow Pay >$
9	$< CreatePR \rightarrow ApprovePR \rightarrow CreatePO \rightarrow SignPO \rightarrow CreatePO \rightarrow SignPO \rightarrow Release \rightarrow GR \rightarrow IR \rightarrow SR \rightarrow IR \rightarrow Pay \rightarrow Pay >$
10	$< CreatePR \rightarrow ApprovePR \rightarrow CreatePO \rightarrow SignPO \rightarrow Release \rightarrow GR \rightarrow IR \rightarrow Pay \rightarrow CreatePO \rightarrow SignPO \rightarrow Release \rightarrow SR \rightarrow IR \rightarrow Pay >$
11	$< CreatePR \rightarrow ApprovePR \rightarrow CreatePO \rightarrow SignPO \rightarrow Release \rightarrow GR \rightarrow SignPO \rightarrow IR \rightarrow SignPO \rightarrow Pay >$
12	$< CreatePR \rightarrow ModifyPO \rightarrow ModifyPO \rightarrow ModifyPO \rightarrow SignPO \rightarrow Release \rightarrow SignPO \rightarrow Release \rightarrow IR \rightarrow SR \rightarrow Pay >$
13	$< CreatePR \rightarrow ApprovePR \rightarrow CreatePO \rightarrow SignPO \rightarrow Pay \rightarrow GR \rightarrow IR \rightarrow Pay >$
14	$< CreatePR \rightarrow ApprovePR \rightarrow CreatePO \rightarrow ApprovePR \rightarrow Release \rightarrow SR \rightarrow IR \rightarrow Pay >$
15	$< CreatePR \rightarrow Pay \rightarrow ApprovePR \rightarrow CreatePO \rightarrow SignPO \rightarrow Release \rightarrow GR \rightarrow IR >$
16	$< CreatePR \rightarrow GoodsReceipt \rightarrow IR \rightarrow Pay \rightarrow ApprovePR \rightarrow CreatePO \rightarrow SignPO \rightarrow Release >$
17	$< CreatePR \rightarrow ApprovePR \rightarrow PO \rightarrow GoodsReceipt \rightarrow SignPO \rightarrow Release \rightarrow IR \rightarrow Pay >$
18	$< CreatePR \rightarrow ApprovePR \rightarrow IR \rightarrow Pay \rightarrow SignPO >$

Table 3: Set of deviating process executions used in the interviews (a selection of this set is used per interview). The abbreviations of process activities that are used in this table (and later in Table 4) are *PR* (Purchase Request), *PO* (Purchase Order), *IR* (Invoice Receipt), *GR* (Goods Receipt), and *SR* (Service Receipt). However, in the interviews, the full name of activities is shown.

the interpretation that has the least combinations of deviations. In trace number 13, for example, this is the interpretation that *Release* is replaced by *Pay* (deviation category "replace two activities"). The classification of our example traces is provided in Table 4. This is used when developing the traces, but is not shared with the interviewees.

The distinction between a "repeat" and an "insert" example requires some attention. Since there is ambiguity in the literature on this aspect, it is important to note that in our design i) "insert" is either an insertion of an activity that is not part of the normative model, or it reflects the presence of an activity that is expected to take place later, ii) "repeat" reflects the presence of an activity that already took place during the process execution at the moment it was expected.

3.2.3. Pilot run

After developing the normative process model and the set of deviating traces, the instrument is used with a group of ten academic researchers in a pilot run, before conducting the actual interviews. This run is used to evaluate and refine the instrument and resulted for example in the accompanying introduction text when sending the model to the interviewees. It also helps to test whether the collected data from interviews will enable us to answer the research questions (Saunders, 2011). There were six researchers with a business and process mining background and four researchers with a computer science background who have reviewed the process instances and the process model to assure their clarity and completeness. No significant changes were made, based on their feedback. Only some small remarks on formulation were given.

Trace	Deviation pattern	Description of deviation in example
1	insert	$< ModifyPO \rightarrow ModifyPO >$ is inserted after $< SR >$
2	insert	$< ModifyPO >$ is inserted after $< SignPO >$
3	insert	$< Pay >$ is inserted after $< ApprovePR >$
4	insert, loop	$< SR >$ or $< GR >$ is inserted, $< Pay >$ is executed in loop
5	skip, skip	$< ApprovePR >$ is skipped, $< Pay >$ is skipped
6	skip	$< CreatePO \rightarrow SignPO \rightarrow Release >$ is skipped
7	skip	$< Release >$ is skipped
8	loop, loop, loop	$< CreatePO \rightarrow SignPO >$ is executed in loop, $< SR >$ is executed in loop, $< Pay >$ is executed in loop
9	loop, insert, repeat, loop	$< CreatePO \rightarrow SignPO >$ is executed in loop, $< SR >$ is inserted, $< IR >$ is repeated, $< Pay >$ is executed in loop
10	repeat, insert, repeat	$< CreatePO \rightarrow SignPO \rightarrow Release >$ is repeated, SR is inserted, $< IR \rightarrow Pay >$ is repeated
11	repeat	$< SignPO >$ is repeated (two times)
12	replace, loop	$< ApprovePR \rightarrow CreatePO >$ is replaced by $< ModifyPO \rightarrow ModifyPO \rightarrow ModifyPO >$, $< SignPO \rightarrow Release >$ is on loop
13	replace	$< Release >$ is replaced by $< Pay >$
14	replace	$< SignPO >$ is replaced by $< ApprovePR >$
15	swap	$< Pay >$ and $< ApprovePR \rightarrow CreatePO \rightarrow SignPO \rightarrow Release \rightarrow GR \rightarrow IR >$ are swapped
16	swap	$< ApprovePR \rightarrow CreatePO \rightarrow SignPO \rightarrow Release >$ and $< GR \rightarrow IR \rightarrow Pay >$ are swapped
17	swap	$< SignPO \rightarrow Release >$ and $< GR >$ are swapped
18	skip, skip, swap	$< CreatePO >$ is skipped, $< SignPO \rightarrow Release \rightarrow GR >$ is skipped, $< SignPO >$ and $< IR \rightarrow Pay >$ are swapped

Table 4: The deviating process executions of the interview instrument, along with their process deviation patterns from literature. This information was not disclosed to the participants.

4. Analyses and results

This section presents the analysis of the field research data and its results. Firstly, we describe the results of coding the interview data. Next, we discuss the audit deviation categories that emerge from the coding and relate these findings to the theoretical process deviation categories in literature.

4.1. Interview coding

In total 1,519 information terms were coded, among which 388 codes were explicitly related to the interpretation of the deviations from a control-flow perspective. Table 5 shows two examples of a quote from an interview, its information term (underlined), the first-order code that was retrieved from it, and the concept it is assigned. Some quotations have been modified slightly to improve clarity. More examples of the first-order codes are included in appendix.

In the transition phase between two cycles of coding, the codes are categorized to generate the underlying concepts. The first-order codes were categorized based on the common term that was used. For instance, all the information terms like "task X is missed," "missing task X is risky for the company, not for me as an auditor," "I miss task X here in this sequence," and "they just miss a few of the steps" are categorized under the concept *"Miss"*. Table 5 mentions the concepts of the two examples (*"Without"* and *"No X"*). The more elaborate table in appendix shows example quotes for all concepts that are assigned at least 10 times to information terms.

In the second-order coding, all terms and concepts that are found in the first-order analysis are systematically integrated into a set of core categories, called themes [9] Five themes that are used when auditors describe process deviations

[9] Conceptual coding is used, which "functions like an umbrella that covers... all other codes

Quote from interview, including the information terms	First-order code	Concept
(Interviewee 11) When there are payments without an invoice for a service, that's typically a risk.	Pay without an IR	Without
(Interviewee 4) If you have an invoice but no purchase order, then first of all, the question is whether they really ordered it.	No purchase order	No X

Table 5: Example quotes, their first-order codes, and concepts

are identified: *Missing, Reordering, Duplicating, Inserting*, and *Substituting*. Three categories comprise multiple concepts, and they represent a relatively large number of them. They are *Missing, Reordering, and Duplicating*. The other two themes, *Inserting* and *Substituting*, represent only one or two concepts each. The details of these themes are covered in the following paragraphs.

Tables 6 to 10 show how all concepts are categorized into the five themes. The tables contain the concepts, the frequency of these concepts over all interviews, and the themes. As can be seen in the tables, the categories that describe deviations that relate to *Missing*, *Reordering*, and *Duplicating* cover several related concepts that auditors use. Table 6 shows that the most first-order codes (150) reside under the category *Missing*, and just two concepts, "Without" (41) and "No" (31), account for nearly half of them. Table 7 shows a similar distribution for the category *Reordering* (98), led by "After" (29) and "Before" (25). The category *Duplicating*, in Table 8, covers 88 codes, with the

and categories." The conceptual categories should be able to describe all the subsets(Saldaña, 2016).

Concept	Frequency	Theme (Total frequency)
Without	41	Missing (150)
No	31	
Not/ never + verb	21	
Miss	15	
Not + verbs: be, have, happen, exist, execute	14	
Straight / Immediately / Automatic	5	
Skip	5	
Forgotten	4	
Lack of	4	
Bypass	4	
A and C (B is missed)	3	
Go (Start) with X	1	
Gap in the flow	1	
Loose invoice	1	

Table 6: Deviation concepts that aggregate into *"Missing"* as a deviation category

most frequent concepts, "Twice" (17), "N-times" (13), "Two-three-four" (13), and "Duplication" (11) accounting for 54 of those codes. Those three categories reflect a clear split from the fourth category, *Inserting* (13) and *Substituting*, which covers only one concept (see tables 9 and 10).[10] These lower concepts and frequencies might indicate that the two themes are less important to auditors or that these categories have a very clear structure in the auditor's internal (domain knowledge) representation. Perhaps the latter is the case for *Inserting*. We investigate this further in the next subsection.

[10] Although *Substituting* covers only one code, we did not aggregate it with another theme because of the content-wise parallel with the deviation category "replace" that exists in the literature.

4.2. Audit process deviation categories

Based on the analysis of the terms that auditors use to describe deviations, we identified five process deviation categories. However, only three categories seem to be used broadly: *Missing*, *Reordering*, and *Duplicating*. These three themes were also mentioned widely by auditors in the first part of the interviews, when they were answering the general question on what kinds of deviations they could think of in the context of P2P process executions. The category *Inserting* was expressed fewer times. The last category, *Substituting*, was only mentioned once.

To compare these deviation categories with their potential (theoretical) counterparts in literature, we analyze the data that are gathered during the second part of the interview. In that part, auditors are confronted with deviating traces, and they are asked to describe those deviations (in case they find any deviation in the trace). Since the example traces are also classified according to the categories in literature (see table 1), we are able to relate the concepts that auditors use to the categories from literature. Table 11 provides an overview of these connections. The process deviation categories from literature are shown in the columns and the categories that the auditors use are shown in the rows.

Skip, Repeat, and Loop

The frequency table shows some clear patterns. We start with discussing the categories from literature that were grouped under exactly one category of

Concept	Frequency	Theme (Total frequency)
After	29	Reordering (98)
Before	25	
Afterwards	8	
Advanced	6	
Pre-	4	
Up front/In front	4	
Too late/ Later	4	
Prior	3	
Postpone	2	
Timing issue	2	
Too soon	1	
Skipping the queue	1	
Not timely X	1	
Not yet X	1	
Somebody can X whenever he want	1	
X in the middle of process	1	
X happens without any order	1	
X is not in the right place	1	
Y in 2016 and X in 2017	1	
X is a step behind	1	
X with a back date	1	

Table 7: Deviation concepts that aggregate into *"Reordering"* as a deviation category

Concept	Frequency	Theme (Total frequency)
Twice	17	Duplicating (88)
N-times	13	
Two-three-four	13	
Duplication	11	
Double	8	
Re-	7	
Multiple/several	6	
Second	4	
Again	4	
Over-	2	
Repeat	1	
Split in different lines	1	
A lot of X and you get only one Y	1	

Table 8: Deviation concepts that aggregate into *"Duplicating"* as a deviation category

Concept	Frequency	Theme (Total frequency)
Verb X	10	Inserting (13)
Verb + activity X	1	
Expect X or Y, here you have both	2	

Table 9: Deviation concepts that aggregate into *"Inserting"* as a deviation category

Concept	Frequency	Theme (Total frequency)
Instead of X, you have Y	1	Substituting (1)

Table 10: Deviation concept that aggregates into *"Substituting"* as a deviation category

		Deviation category in literature						
		insert	skip	replace	swap	repeat	loop	total
Deviation category used by auditors	missing	-	24	14	9	-	-	47
	reordering	5	-	7	19	-	-	31
	duplicating	5	-	5	-	22	33	65
	inserting	13	-	-	-	-	-	13
	substituting	-	-	1	-	-	-	1

Table 11: Matrix with frequencies of deviation categories used by auditors when interpreting deviations that are categorized according to previous literature

the auditors: "skip," "repeat," and "loop." All the "skip" deviation examples that are shown to auditors, are described only with concepts (such as*without, no, miss, skip,*) that are coded as the "*Missing*" category. None of the 24 interpretations of auditors is classified under another category. We can state that the "skip" deviation category from literature is used by auditors in the same way as it is interpreted in process literature.

Almost the same situation holds for the examples of "repeat" (22 examples) and "loop" (23) deviations. All the concepts that auditors used to describe these examples are grouped under one particular category (*"Duplicating"*). None of the examples is categorized in another theme. However, reading the table in the opposite direction reveals that *"Duplicating"* is an umbrella category from auditors for more than one deviation category from literature. It seems that auditors bundle the deviation categories "repeat" and "loop," and to a certain extent also "insert" and "replace," into this category. Auditors apparently do not make a distinction between repeating an activity "somewhere in the process execution" and repeating an activity "right after it has been executed," which would be called a (self-)loop in process literature.

Missing, Reordering, and Duplicating

Continuing the analysis from the deviation categories as used by the auditors, following observations can be made. The *"Missing"* category hosts several deviation categories from literature: "skip" (24), "replace" (14), and "swap" (9). As already pointed out, the "skip" examples all find their way to this category. On top of that, some examples of activities that are "replaced" by other activities are interpreted as *"Missing"*. Further, some of the "swap" examples are somewhat surprisingly categorized as "missing" by auditors. For example, the "swap" examples in traces number 16 and 17 are interpreted by auditors as missing an approval. They describe these deviations as the sequence of $< GoodsReceipt, InvoiceReceipt, Pay >$ or the activity $< GoodsReceipt >$ was done *"without approval"* or when *"no approval"* was in place.

A comparable pattern is found for the *"Reordering"* category. This category also hosts examples of several categories from literature: "insert" (5), "replace" (7), and "swap" (19). From a theoretical, content-wise point of view, swapping activities matches most with reordering them. This is partly confirmed by our data: 19 "swap" examples are indeed classified as *"Reordering"*. However, nine other "swap" examples are classified as *"Missing"*. Hence there is no one-on-one translation between the auditor's *Reordering* and the literature's "swap" category. The auditors' *"Reordering"* classifications also include "insert" and "replace" examples and not all "swap" examples are perceived as "Reordering."

Also, for "*Duplicating*," a link with deviation categories from literature is present, but it is not a sharp-cut translation. Aside from all "repeat" and "loop" examples that are classified as "*Duplicating*," also some "insert" and "replace" examples are classified under the same deviation category. As mentioned before, "insert" examples can refer to the execution of an activity earlier than expected. If that same activity is also executed at the expected moment, the process execution exhibits twice the same activity, hence the link to du-

plication. For instance, in example number 3, based on the deviation patterns from literature, *Pay* is inserted after *ApprovePurchaseOrder*, since it is executed in its proper time as well. However, in the auditors' interpretation, it is seen as a combination of *"Reordering"* and *"Duplicating"*. The same line of thought exists with the "replace" examples. A "replace" can result in having the same activity twice. In those situations, an interpretation as "*Duplicating*" is logical. These observations could indicate that auditors are more focused on the consequences (one activity is executed twice, hence duplicated), and not on the underlying mechanism (an activity is inserted in the process execution, or is replaced by another).

Inserting

In contrast to the first three categories, the category of "*Inserting*" category was backed up by only a few underlying concepts. This might indicate that this category is either not frequently used, or that it represents a well-understood concept to auditors. When inspecting the usage of the *"Inserting"* category by the auditors, we observe that when this category was used, this was indeed during a discussion of "insert" examples. However, the used concepts are only weakly connected to the word "*Inserting*," as shown in Table 9. On top of that, only some "insert" examples reside under this category, since the majority was classified as *"Reordering"* and/or *"Duplicating"*. Given this ambiguous situation, not only here but also in literature, we explore this category more in depth.

Six example traces contained an "insert" deviation (examples 1, 2, 3, 4, 9 and 10). In two examples *Modify Purchase Order* is inserted somewhere in the process, while it was not in the set of activities allowed by the model. In one example, *Pay* is inserted somewhere in the middle of the process, and in other examples, *Service Receipt* or *Goods Receipt* is inserted, while only one of both was expected according to the model.

In examples 1 and 2, the insertion of $ModifyPurchaseOrder$ in the traces

did not raise any concern or curiosity to the auditors, as long as they can see an approval after it. For instance, Interviewee 7 describes the second example as follows: "The purchase order is modified, it should be re-approved as well, before being released. So, at this point, I would say that this is a non-authorized modification." So, although the $ModifyPurchaseOrder$ was not in the set of activities in the model, it was acceptable for the interviewees. Their main concern was whether there is a control on this inserted task. These two examples are in all interviews interpreted as "Inserting" by the auditors. So, it seems that when the "insert" refers to a new activity, auditors also categorize this as an "*Inserting*" deviation.

When the inserted activity is part of the set of allowed activities in the model, we notice that auditors do not use a specific term for the "insert" examples. The auditors describe this sort of deviations as "the modification of the Purchase Order was performed after approval" and "this Purchase Order is overwritten" in interview 7, or "you can still change the PO, and this is dangerous of course" in interview 9. To include these observations in our data, we used the concept "Verb + activity X" and "Verb X" for these information terms, which we grouped into "*Inserting*."

For the examples in which either *Service Receipt* or *Goods Receipt* is allowed by the model and both are present in the traces (traces 4 and 10), auditors do consider them deviations. For instance, a participant in the first interview said, "this is a bit strange, because normally you have Goods Receipt or you have Service Receipt." In the second interview it was mentioned that "in most cases you'd expect a GR or SR, here we have both." So, although they do not use a specific information term, it is obvious that they consider such traces as deviations and they look for justifications. We conclude that although there are only a few terms used when interpreting the "insert" deviations, auditors consider it as a type of deviation that needs to be explored further and can be interpreted in different ways, depending on the context of the "insert." As a

result, this deviation category from literature cannot be transposed uniquely to the *"Inserting"* deviation category that is sometimes used by auditors.

4.3. Process deviation categories that are not adopted by auditors

Some of the process deviation categories that are described in literature, are not used by auditors or are merged into bigger deviation categories. They are "replace" and "loop."

Replace

From all process deviation categories from literature, "replace" does not seem to be adopted by auditors. The majority of "replace" examples are described by auditors in a less abstract level: they are categorized under *"Missing"* in combination with *"Missing"* with *"Reordering"* or *"Duplicating"*. For instance, Interviewee 7 describes the deviating trace 13 as "there is [...] an approval missing." Or, Interviewee 8 describes the deviation in trace 14 as "the approval of the PO, that's missing," indicating the importance of absence of activities.

The other terms that auditors use when describing the other deviations are either linked to *"Reordering"* (7), or *"Duplicating"* (5) (see Table 11). For instance, Interviewee 1 considers the deviation in trace 13 as an "advanced payment," and Interviewee 7 describes it as "paying prior to the goods and invoice receipt." Both of these information terms are concepts of the reordering theme. Interviewee 2 describes the deviation in the same example as a "second payment" and interview 7 uses the term "duplicate payment." These terms belong to the duplication theme.

Over the course of all the interviews, we heard the term "instead" only once. This occurred when the third interviewee was describing example 14. He said "Signature is missing, so instead of the signature, you have again the approval." We considered this term as an interpretation of *"Substituting"*, which would have been the most straight-forward transformation of the "replace" examples.

However, even in this quotation, he described his interpretation by mentioning *"Missing"* and *"Duplicating"* categories ("Signature is missing" and "you have again the approval"). No other term similar to or related to *"Substituting"* was used by any other interviewee.

As a result, we consider the "replace" pattern from literature as too abstract to use in an auditing context. The analysis suggests it is better to map this deviation pattern into a combination of deviation categories like *"Missing"* with *"Reordering"* or *"Missing"* with *"Duplicating."*

Loop

Table 11 shows the categories that are used to describe "loop" examples. As can be seen, all the 33 concepts used by auditors to describe these examples are categorized as *"Duplicating"*. However, as mentioned before, this category also hosts all "repeat" examples. In other words: auditors do not use a dedicated category for deviations that relate to looping an activity.

5. Discussion

5.1. Process deviation categories in auditing

As mentioned in the results section, we have identified three process deviation categories that auditors use when confronted with deviating process executions. These are partly in line with the theoretical process deviation patterns. These are *"Missing"*, *"Reordering"*, and *"Duplicating"*, coinciding with the three categories that were also used in the auditing study of Chiu and Jans (2019), but not with the categories that were used in the process management and mining research fields. To a lesser extent, the category *"Inserting"* showed up.

Three theoretical categories are unambiguously classified in one of the 'auditing categories': "skip" deviations are all grouped under "*Missing*," and "repeat"

and "loop" examples are all grouped under "*Duplicating*." With regard to the latter category: auditors do not seem to make a distinction between repeating an activity "somewhere in the process execution" (a "repeat") and "right after it has been executed" (a "loop"). Based on the cognitive fit theory, this implies that when auditors are presented information on deviating process executions in the format of "repeats" and "loops," this information would not be optimally aligned with how their domain knowledge is mentally represented, namely with the concept of "*Duplicating*." As a consequence, the task performance (correctly assessing the risk of the deviating pattern(s)) would decline.

To continue, "*Duplicating*" captures both "repeat" and "loop" patterns, but it serves an even broader purpose. Particular examples of "insert" also find their way to "*Duplicating*." Recalling our design choice to make a distinction between "repeat" and "insert," this finding relates to the split in literature where García-Bañuelos et al. (2015) and Weber et al. (2008) distinguish these patterns from each other while Adriansyah et al. (2010) do not. The latter classify all these deviations as "insert." We followed the former studies to classify them as "insert" when the deviation concerns an activity that is not expected at all, or when an additional execution of an activity takes place *before* the moment it is expected. When this happens *after* the expected moment, it is a "repeat" example. It seems that the auditors do not take all these nuances or underlying mechanisms into account. Rather, they focus on the end result: an activity has been executed multiple times; regardless of whether this was in a loop or not, or whether this was "too soon" or not. This could also explain the link between "replace" and "*Duplicating*." There is no clear line in the way "replace" examples are categorized, but this could be interpreted as another example of focusing on the result and not at the cause. When the replacement resulted in multiple executions of the same activity, it is classified as "*Duplicating*."

So, although there are three theoretical deviation categories from literature that are fully subsumed into categories that auditors use, there are no one-

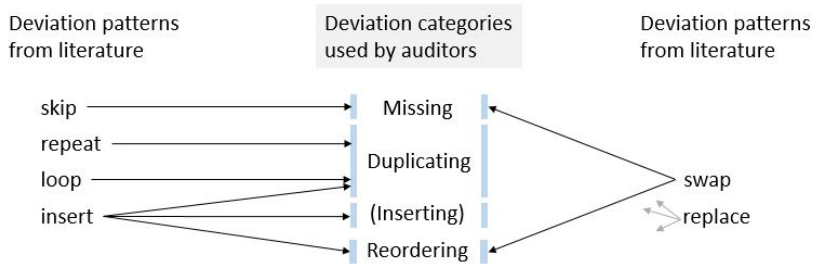

Figure 3: Relation between process deviation categories in literature and categories used by auditors.

on-one transpositions. Like "*Duplicating*," "*Missing*" covers multiple patterns. "Skip" examples are all included in the "*Missing*" category, but "swap" and "replace" examples also are included there. Figure 3 presents the links we encounter in our analysis.

The relationship between the "insert" pattern as it is used in the literature and the "*Inserting*" category used by the auditors is peculiar. When interpreting examples of "insert," auditors sometimes link it to "*Inserting*," although not in similar formulations. The concepts inside this category are more general. So, although only "insert" examples triggered the use of the "*Inserting*" category, these deviations were mostly described as "*Reordering*" and/or "*Duplicating*." Again, a focus on the end result seems to prevail. It might indicate that auditors' mental representation of their domain knowledge is directed more towards facts than towards possible root causes or information at a process level that is too detailed. Again, as with the deviation patterns "repeat" and "loop", our findings suggest not to adopt the "insert" pattern from the process literature blindly, as this might not be in line with the ways that auditors look at these deviations. Instead, it seems better to use "*Duplicating*" and/or "*Reordering*."

According to our results, the "replace" pattern seems too abstract for auditors to use. The auditors interpret these deviations at a lower level, as a

combination of "*Missing*" and "*Reordering*" an activity or "*Missing*" and "*Duplicating*." A comparable situation holds for "swap," which is mostly described as "*Missing*" an activity on one hand, and "*Reordering*" on the other hand.

In sum, there is not a single deviation pattern, as it is presented in the business process management and process mining literature, that seems suited for presenting deviating process executions to auditors. Even when there are conceptual links, auditors' categories are generally broader, or they are at a different level. The implications of this mismatch can be great. If auditing algorithms use the known deviation categories of process management literature, this would structure information in a way that is not aligned with the user schemas of the auditor. According to the Cognitive Fit paradigm, this lesser fit results in lower performance. As previous research confirm this theory in accounting settings, a negative impact on both effectiveness and efficiency of the auditor's risk assessment and control testing activities is expected when the deviation categories of process management literature would be used in audit tools.

Our study does not investigate why auditors seem to use less nuances when discussing process deviations than scholars in the field of process management and mining. It would be very valuable, though, to better understand the underlying reasons. One possible explanation for this finding might be that auditors are not 'raised' in processes. As assurance is provided over financial reporting, which could be seen as the last activity(ies) of a business process, the traditional focus of an auditor is not an end-to-end focus. Also, testing controls boils down to zooming in on process fragments, making again abstraction of the process dimension. This absence of a process-oriented view on organizations during engagements, might be one explanation why auditors for example, rather see a "*Missing*" and "*Reordering*" activity than a "swap". The assertion that auditors exhibit a diminished process focus and the conjectured explanation are unsupported by the findings of this study, thus warranting further investigation in future research.

5.2. Internal versus external auditors

As explained in Section 3.1, the auditor sample consisted of both internal and external auditors (five internal and nine external auditors). After analysing the interview transcriptions for all interviews, we have analyzed the concepts from both internal and external auditor groups separately once more. We executed this additional analysis to investigate whether the interpretations of internal and external auditors are different from each other as their objectives differ. Generally speaking, the main objectives of the internal auditor is to examine organizations' business practices and risks, including testing the effectiveness and efficiency of internal controls. On the other hand, the objective of external auditors is to assure the financial statements are free of material misstatements.

Checking the concepts and themes that were uncovered from the transcriptions per subgroup, we found no notable difference between the interpretation of two groups. In the first interview with an internal auditor, four different concepts (i.e., Missing, Reordering, Insertion, Duplication) were coded. The following internal auditor interviews revealed no additional concepts. The five internal auditors interpreted the deviations as follows:

- All loops and repetitions are interpreted as "Duplicating"
- All skips are interpreted as "Missing"
- Swap is interpreted as "missing and reordering" by one internal auditor, while interpreted as "Missing" by two other auditors.
- Insert was interpreted as "Insert" by two auditors, once as "Duplication" by another auditor and once as "Reordering and Insert"

As mentioned before, "Substitution" is only mentioned once during the 62 discussions (by an external auditor). "Replace", the last theoretical category, was never recognized, neither by external, nor by the internal auditors.

5.3. Opportunities in handling deviations

Knowing that auditors primarily use these three main categories when dealing with process deviations, analytical procedures can respond accordingly. Instead of presenting process traces in full–like in Jans et al. (2014), leaving the auditor to distill the deviations from it–, the deviations can be described per category, knowing that these are meaningful to the auditor. So instead of presenting 1.000 process variants to be inspected one by one, this information can be presented as 300 times an activity was "*Missing*", 200 times they were "*Reordered*" and 200 times there was a "*Duplication.*" Compressing the process executions in this way, will presumably require a lower cognitive effort of the auditor. A tool can be made to drill down and see, within each category, which activities were missing, reordered, or duplicated. This way, the auditor can classify the deviations as exception or as anomaly.

Although compressing process executions into deviation categories might be relatively more efficient than checking all process variants, it remains a manual task that is likely to be economically unfeasible. Therefore, an active learning approach like presented in the 'Catalyst approach' of Jans and Hosseinpour (2019) is a possible way forward. Our findings add to this framework by making one piece of the puzzle more concrete: how best to present the information on deviations to auditors, the oracle. The development of a classifier that uses the auditor as an oracle and learns from fuzzy input, is still an open research challenge. A key aspect to such an approach is, however, to present the oracle with the most informative input to make the learning as efficient as possible. The insights of our study contribute to this. Additionally, 'audit-proof' conformance checking algorithms that provide the output in these three categories, is an open research problem that might be addressed by the process mining research field. Other studies could also delve into the potential return on investment for firms that adopt this approach. It would be interesting to investigate what the cost implications are, which requirements must be met ex ante, what the saved time is, and what the effects on the audit quality are.

5.4. Limitations

Our findings are generic in the sense that the discovered categories are process-independent. Although these categories emerged while discussing deviations on a procurement process, the categories could also serve to classify deviations to other processes. In every process, activities can be missing, reordered or repeated. However, our findings may be restricted by some limitations of our study. To start with, we investigate process deviations from a sequence point of view, as a first step in exploring this new topic. However, some important aspects such as the data that are needed for checking quantities and prices or the segregation of duties, are not included in the control-flow perspective. As expected, auditors mentioned during the interviews that they require more information like values, suppliers, and related persons to classify these process executions as deviations. For instance, Interviewee 8 needed more information on purchase order data and said, "There [might be] no matching between what actually has been delivered and what has been ordered." Interviewee 10, however, needed information from the resource segregation of duties perspective and said, "Modify Purchase Order, then I would have an extra check in place to see if it's not the same person for this type of thing, and that the person with the right authority actually approved this change." Still, to implement a fully automated business process analysis to support the auditor, it is imperative to understand deviations in terms of control-flow. Including additional data on the process instance, the involved resources and the time dimension could be part of future research.

Moreover, the instrument used during the interviews might have its shortcomings. Although developed as realistic as possible and test-run by auditors, the simplicity of our normative model, both in the number of possible process executions and the number of activities, might have limited our findings. The same applies to our set of deviating process execution examples. The absence of a stronger "*Inserting*" category may be due to this simplified setting. Future research could expand this instrument with more activities and more deviating

traces to investigate whether "*Inserting*" is prevalent in more complex business processes.

Lastly, our sample's mixture of internal and external auditors creates confounded data. Although additional analyses indicated no significant difference between the two groups of auditors in our sample, future research on this aspect might completely rule out potential biases. A larger sample size could provide more robust results. Also, a more qualitative approach, delving into the different nuances of internal and external auditors would offer valuable additional insights.

6. Conclusion

Starting from the importance of understanding processes and their related internal controls, this paper investigates how auditors interpret process deviations. The interpretation is an important phase in identifying potential control deficiencies.

Whereas process mining techniques seem to be wel-suited to uncover transactions that deviate from the expected process, thousands of deviations will surface through a process mining analysis. To process all these alarms and classifying them as either a harmless exception or an anomaly to investigate further, the 'Catalyst approach' was presented by Jans and Hosseinpour (2019). This approach suggests to combine artificial intelligence with the domain expertise of the auditor through an active-learning algorithm. Whether it would be this particular approach, or another, similar approach: it is vital to present the information in a meaningful way to the expert.

Building on Cognitive Fit theory, it is important that the presented deviations are aligned with the way auditors process information on deviating process executions. That is, if the presentation of the problem does not fit the

way domain knowledge is structured internally in the expert's mind, the task performance decreases (Vessey and Galletta, 1991). So, before moving forward with processing this alarm flood, whether it is by a machine learning algorithm or not, we need a better understanding of how information on identified deviations is interpreted by auditors. Only then can this information be presented in a way that supports the auditor. Our study is a first step in this direction by uncovering deviation categories that auditors use, revealing underlying knowledge structures.

In the process mining and business process management literature we identified six patterns to describe mismatches between a (real) process execution and the modeled process execution: skip, insert, replace, swap, repeat, and having a loop on an activity. These patterns were introduced without any post-validation. Starting from these theoretical deviation categories, we investigated through a field study which categories (if any) are applied by auditors when they are presented with information on process deviations. We found that some of the categories that auditors use relate to the theoretical deviation patterns in process mining literature. However, they are used with a different interpretation. Some theoretical patterns are grouped together into a single category by the auditors. Other patterns are scattered over different categories, depending on the example the auditors are shown. Comparing our results with the one study in accounting that mentions deviation categories in a process mining context, we find supporting evidence for the categorization of Chiu and Jans (2019). This is not surprising since the authors of that study are familiar with the auditing context. Still, they are no auditors, nor would the opinion of a single set of authors suffice to validate these categories.

Our study shows that merely starting from the output of current process mining algorithms without adapting to the auditing context would hamper the exploitation of these techniques to their full potential. We hope to mitigate the risk that these techniques are declared "not suitable", solely because the

output is not (yet?) presented in the correct way. Additionally, and perhaps more important, we call for prudence in incorporating deviation categories from the business process research field in an audit context. This might negatively impact the effectiveness and efficiency of auditors' risk assessment.

With this work, we have opened up the road for future research on this topic. Having more insight into the categories that help create connections to auditors' mental knowledge representation allows for further research into process mining algorithms in an auditing context. So, their potential can be leveraged to move towards continuous auditing and full population testing. Future research could further deepen our understanding of why auditors rely on these categories on one hand and illuminate how the algorithms and auditors deal with all this information on the other hand.

References

Adriansyah, A., van Dongen, B. F., and van der Aalst, W. M. (2011). Conformance checking using cost-based fitness analysis. In *Enterprise Distributed Object Computing Conference (EDOC), 2011 15th IEEE International*, pages 55–64. IEEE Computer Society.

Adriansyah, A., van Dongen, B. F., and Zannone, N. (2010). Controlling break-the-glass through alignment. *Social Computing (SocialCom), IEEE International Conference on Privacy, Security, Risk and Trust.*

Alles, M., Brennan, G., Kogan, A., and Vasarhelyi, M. A. (2006a). Continuous monitoring of business process controls: A pilot implementation of a continuous auditing system at siemens. *International Journal of Accounting Information Systems*, 7(2):137–161.

Alles, M. G., Kogan, A., and Vasarhelyi, M. A. (2008). Putting continuous auditing theory into practice: Lessons from two pilot implementations. *Journal of Information Systems*, 22(2):195–214.

Alles, M. G., Tostes, F., Vasarhelyi, M. A., and Riccio, E. L. (2006b). Continuous auditing: the usa experience and considerations for its implementation in brazil. *JISTEM-Journal of Information Systems and Technology Management*, 3(2):211–224.

Bierstaker, J. L., Bedard, J. C., and Biggs, S. F. (1999). The role of problem representation shifts in auditor decision processes in analytical procedures. *Auditing: A Journal of Practice & Theory*, 18(1):18–36.

Bierstaker, J. L., Hunton, J. E., and Thibodeau, J. C. (2009). Do client-prepared internal control documentation and business process flowcharts help or hinder an auditor's ability to identify missing controls? *Auditing: A Journal of Practice & Theory*, 28(1):79–94.

Bonner, S. E., Libby, R., and Nelson, M. W. (1997). Audit category knowledge as a precondition to learning from experience. *Accounting, Organizations and Society*, 22(5):387–410.

Boritz, J. E., Borthick, A. F., and Presslee, A. (2012). The effect of business process representation type on assessment of business and control risks: Diagrams versus narratives. *Issues in Accounting Education*, 27(4):895–915.

Braun, R. L. and Davis, H. E. (2003). Computer-assisted audit tools and techniques: Analysis and perspectives. *Managerial Auditing Journal*, 18(9):725–731.

Carmona, J., van Dongen, B., and Weidlich, M. (2022). Conformance checking: foundations, milestones and challenges. In *Process Mining Handbook*, pages 155–190. Springer.

Chan, D. Y. and Vasarhelyi, M. A. (2011). Innovation and practice of continuous auditing. *International Journal of Accounting Information Systems*, 12(2):152–160.

Chiu, T. and Jans, M. (2019). Process mining of event logs: A case study evaluating internal control effectiveness. *Accounting Horizons*, 33(3):141–156.

Corbin, J. M. and Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology*, 13(1):3–21.

Davern, M., Shaft, T., and Te'eni, D. (2012). Cognition matters: Enduring questions in cognitive is research. *Journal of the Association for Information Systems; Atlanta*, 13:273–314.

De Weerdt, J., De Backer, M., Vanthienen, J., and Baesens, B. (2012). A multi-dimensional quality assessment of state-of-the-art process discovery algorithms using real-life event logs. *Inf. Syst.*, 37(7):654–676.

Debreceny, R., Gray, G. L., Tham, W.-L., Goh, K.-Y., and Tang, P.-L. (2003). The development of embedded audit modules to support continuous monitoring in the electronic commerce environment. *International Journal of Auditing*, 7(2):169–185.

Debreceny, R. and Rahman, A. (2005). Firm-specific determinants of continuous corporate disclosures. *The International Journal of Accounting*, 40(3):249–278.

Dull, R. B., Tegarden, D. P., and Schleifer, L. L. (2006). Actve: A proposal for an automated continuous transaction verification environment. *Journal of Emerging Technologies in Accounting*, 3(1):81–96.

Dunn, C. and Grabski, S. (2001). An investigation of localization as an element of cognitive fit in accounting model representations. *Decision Sciences*, 32(1):55–94.

Dunn, C. L. and Gerard, G. J. (2001). Auditor efficiency and effectiveness with diagrammatic and linguistic conceptual model representations. *International Journal of Accounting Information Systems*, 2(4):223–248.

Dunn, C. L., Gerard, G. J., and Grabski, S. V. (2017). The combined effects of user schemas and degree of cognitive fit on data retrieval performance. *International Journal of Accounting Information Systems*, 26:46 – 67.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of management review*, 14(4):532–550.

Elliott, R. K. (2002). Twenty-first century assurance. *Auditing: A Journal of Practice & Theory*, 21(1):139–146.

Fahland, D. and van der Aalst, W. M. (2015). Model repair - aligning process models to reality. *Inf. Syst.*, 47(C):220–243.

Flowerday, S. and Von Solms, R. (2005). Continuous auditing: verifying information integrity and providing assurances for financial reports. *Computer Fraud & Security*, 2005(7):12–16.

García-Bañuelos, L., van Beest, N. R., Dumas, M., and La Rosa, M. (2015). Complete and interpretable conformance checking of business processes. *BPM center*, pages 1–23.

Gioia, D. A., Corley, K. G., and Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research: Notes on the gioia methodology. *Organizational Research Methods*, 16(1):15–31.

Glaser, B. G. and Strauss, A. L. (1968). *The discovery of grounded theory: strategies for qualitative research*. Weidenfeld and Nicolson, New Brunswick, 4. paperback printing edition.

Groomer, S. M. and Murthy, U. S. (1989). Continuous auditing of database applications: An embedded audit module approach. *Journal of Information Systems*, 3(2):53–69.

Hosseinpour, M. and Jans, M. (2016). Categorizing identified deviations for auditing. In Ceravolo, P., Guetl, C., and Rinderle-Ma, S., editors, *Proceedings of the 6th International Symposium on Data-driven Process Discovery and Analysis (SIMPDA 2016)*, number 1757 in CEUR Workshop Proceedings, pages 125–129.

IAASB (2019). International standard on auditing 315 (revised 2019) - identifying and assessing the risks of material misstatement.

Issa, H. (2013). *Exceptional exceptions*. PhD thesis, Rutgers University-Graduate School-Newark.

Jans, M., Alles, M. G., and Vasarhelyi, M. A. (2013). The case for process mining in auditing: Sources of value added and areas of application. *International Journal of Accounting Information Systems 14*, 14:1–20.

Jans, M., Alles, M. G., and Vasarhelyi, M. A. (2014). A field study on the use of process mining of event logs as an analytical procedure in auditing. *The Accounting Review*, 89(5):1751–1773.

Jans, M. and Hosseinpour, M. (2019). How active learning and process mining can act as Continuous Auditing catalyst. *International Journal of Accounting Information Systems*, 32:44–58.

Kelton, A. S., Pennington, R. R., and Tuttle, B. M. (2010). The effects of information presentation format on judgment and decision making: A review of the information systems research. *Journal of Information Systems*, 24(2):79–105.

Kim, Y. and Vasarhelyi, M. A. (2012). A model to detect potentially fraudulent/abnormal wires of an insurance company: An unsupervised rule-based approach. *Journal of Emerging Technologies in Accounting*, 9(1):95–110.

Kogan, A., Sudit, E. F., and Vasarhelyi, M. A. (1999). Continuous online auditing: A program of research. *Journal of Information Systems*, 13(2):87–103.

Kopp, L. S. and O'Donnell, E. (2005). The influence of a business-process focus on category knowledge and internal control evaluation. *Accounting, Organizations and Society*, 30(5):423–434.

Kuenkaikaew, S. and Vasarhelyi, M. A. (2013). The predictive audit framework. *The International Journal of Digital Accounting Research*, 13:37–71.

Kuhn Jr, J. R. and Sutton, S. G. (2010). Continuous auditing in erp system environments: The current state and future directions. *Journal of Information Systems*, 24(1):91–112.

Li, P., Chan, D. Y., and Kogan, A. (2015). Exception prioritization in the continuous auditing environment: A framework and experimental evaluation. *Journal of Information Systems*, 30(2):135–157.

Libby, R. and Luft, J. (1993). Determinants of judgment performance in accounting settings: Ability, knowledge, motivation, and environment. *Accounting, organizations and society*, 18(5):425–450.

Malsch, B. and Salterio, S. E. (2015). "doing good field research": Assessing the quality of audit field research. *Auditing: A Journal of Practice & Theory*, 35(1):1–22.

Mendling, J., Neumann, G., and van der Aalst, W. M. P. (2007). Understanding the occurrence of errors in process models based on metrics. In *On the Move to Meaningful Internet Systems 2007: CoopIS, DOA, ODBASE, GADA, and IS, OTM Confederated International Conferences CoopIS,DOA, ODBASE, GADA, and IS*, pages 113–130.

Nelson, M. W., Libby, R., and Bonner, S. E. (1995). Knowledge structure and the estimation of conditional probabilities in audit planning. *Accounting Review*, pages 27–47.

O'Donnell, E. and Schultz Jr, J. J. (2003). The influence of business-process-focused audit support software on analytical procedures judgments. *Auditing: A Journal of Practice \& Theory*, 22(2):265–279.

PCAOB (2010). Auditing standard no. 12 - identifying and assessing risks of material misstatement.

Perols, J. L. and Murthy, U. S. (2012). Information fusion in continuous assurance. *Journal of Information Systems*, 26(2):35–52.

Phua, C., Lee, V., Smith, K., and Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. *arXiv preprint arXiv:1009.6119*.

Ritchi, H., Jans, M. J., Mendling, J., and Reijers, H. (2020). The influence of business process representation on performance of different task types. *Journal of Information Systems*, 34(1):167–194.

Russell, N., van der Aalst, W. M., and ter Hofstede, A. H. (2006). Exception handling patterns in process-aware information systems. *BPM Center Report BPM*, pages 288–302.

Sadiq, S., Governatori, G., and Namiri, K. (2007). Modeling control objectives for business process compliance. In *International conference on business process management*, pages 149–164. Springer.

Saldaña, J. (2016). *The coding manual for qualitative researchers*. SAGE London, 3 edition.

Saunders, M. N. (2011). *Research methods for business students, 5/e*. Pearson Education India.

Seale, C. (1999). Quality in qualitative research. *Qualitative inquiry*, 5(4):465–478.

Shaft, T. M. and Vessey, I. (2006). The role of cognitive fit in the relationship between software comprehension and modification. *MIS Quarterly*, 30(1):29–55. Publisher: Management Information Systems Research Center, University of Minnesota.

Strauss, A. L. (1987). *Qualitative analysis for social scientists*. Cambridge University Press.

Stringer, E. T. (2014). *Action research*. Thousand Oaks, California : SAGE, 4 edition.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2):257–285.

Thiprungsri, S. and Vasarhelyi, M. A. (2011). Cluster analysis for anomaly detection in accounting data: An audit approach. *International Journal of Digital Accounting Research*, 11.

van der Aalst, W. M. P. (2016). *Process mining: Data science in action*. Springer, Heidelberg.

Vasarhelyi, M. A. and Halper, F. B. (1991). The continuous audit of online systems. *Auditing: A Journal of Practice & Theory*, 10(1):110–125.

Vasarhelyi, M. A., Kogan, A., and Alles, M. G. (2002). Would continuous auditing have prevented the enron mess? *CPA JOURNAL*, 72(7):80–80.

Vessey, I. and Galletta, D. (1991). Cognitive fit: An empirical study of information acquisition. *Information Systems Research*, 2(1):63–84.

Weber, B., Reichert, M., and Rinderle-Ma, S. (2008). Change patterns and change support features–enhancing flexibility in process-aware information systems. *Data & knowledge engineering*, 66(3):438–466.

Werner, M., Wiese, M., and Maas, A. (2021). Embedding process mining into financial statement audits. *International Journal of Accounting Information Systems*, 41:100514.

Willig, C. (2013). *Introducing qualitative research in psychology*. McGraw-Hill Education (UK).

Wilson, J. (2014). *Essentials of business research: A guide to doing your research project*. Sage.