Building a valuable event log for process mining: an experimental exploration of a guided process

Peer-reviewed author version

JANS, Mieke; Soffer, Pnina & JOUCK, Toon (2019) Building a valuable event log for process mining: an experimental exploration of a guided process. In: Enterprise Information Systems, 13 (5), p. 601-630.

DOI: 10.1080/17517575.2019.1587788
Handle: http://hdl.handle.net/1942/28116

Building a valuable event log for process mining: An experimental exploration of a guided process

Mieke Jans[a], Pnina Soffer[b*], and Toon Jouck[a†]

[a]Hasselt University, Martelarenlaan 42, 3500 Hasselt, Belgium; [b]University of Haifa, Carmel Mount Rabin building, 31905 Haifa, Israel

ARTICLE HISTORY

Compiled February 7, 2019

Abstract

The digitalisation of business processes has led to new opportunities for companies. One such opportunity is to investigate process executions by using process mining, but it may require a complex preparation step of building an event log. This task is often perceived as a technical task, although business considerations should be involved due to possible business implications. We argue this is a complex task that requires proper guidance to be performed adequately. In this paper, we examine whether and how a guiding procedure supports the performance of this task. In an experimental study, we follow and compare the lines of thinking of novices that follow procedural guidance with an undirected control group. Our findings provide insights into the parts played by business and technical considerations in this task and suggest that procedural guidance positively impacts the process of building an event log and its outcome.

KEYWORDS

process mining; event log; procedural instructions;

1. Introduction

Process mining has received increasing attention in both academic research studies and industry. Process mining, as other data science areas, builds on the increasing availability of event logs that are generated by systems recording every operation that takes place. With these logs process mining algorithms can be used for process discovery, conformance checking, and model enhancement, thus providing valuable business insights (van der Aalst, 2011). However, a prerequisite for applying process mining algorithms is to have an event log whose structure

CONTACT Mieke Jans. Email: mieke.jans@uhasselt.be

*spnina@is.haifa.ac.il

†toon.jouck@uhasselt.be

is suitable for mining. For certain kinds of systems, most notably process aware information systems, such logs are readily available, corresponding to a predefined process specification. For other kinds of systems (namely, process unaware systems) logs can be generated from the data existing in the (typically relational) database. To build such an event log is a crucial step for gaining valuable process insights and is considered a complex task (see for example Lu et al. (2015)).

Transforming available data into an event log is usually perceived as a technical task to be executed by a data scientist (Howe et al., 2017). However, to enable the creation of business value through process mining, various business considerations should be taken into account even at the data preparation phase (Jans and Soffer, 2017). The decisions taken during this preparatory phase have a quality impact on further analysis. Where previous work focused on the technical challenges of creating an event log (e.g. Calvanese et al. (2015)) or on discovering an accurate graphical model that describes all document transactions (Lu et al., 2015), this paper emphasizes the challenge of building an event log that would be valuable *in a certain business context*. Namely, different event logs can be technically constructed out of the same database, each of them highlighting different aspects of a valuable analysis. Aside from taking all these aspects into consideration, there is also the matter of noise: irrelevant information that may be presented to the analyst, so a distinction is needed between relevant and irrelevant information. This makes the challenge of building a valuable event log even more demanding. This paper aims to provide guidance for novice process analysts in performing this complicated task. We further aim that this guidance would be based on an understanding of how it can affect the process of building an event log.

Building an event log entails a number of tasks which build upon each other. The question we address is whether guidance that explicitly relates to their procedural aspect supports analysts in better performing the task as compared to when analysts rely only on the relevant concepts and principles. Furthermore, we examine the detailed process of event log building to understand the difference made by a procedure-based guidance. A procedure that was used for this purpose consists of a sequence of actions that should lead to the task goal. It provides a way to organize information in order to get to the solution: an event log that is aligned with the project problem. By providing structure to the problem, the procedure is aimed to guide the analyst through the decisions that are taken during the event log building task. Our in-depth examination of the process reveals whether and how this structuring is accomplished.

To investigate the effect of procedural guidance on the event log building process, an experiment was conducted with novice process mining analysts. A treatment group was taught on the topic of event log building by means of a procedure, which later on served for guiding them when performing the task, and a control group was only taught about the concepts involved and their relationships. Subjects in both groups performed a log building task in pairs and their discussions were recorded as verbal protocols. Log building performance was measured and compared for the two groups, and content analysis of the verbal protocols yielded insights about the thinking process observed in each group.

The remainder of the paper is structured as follows: We start in sections 2 and 3 by providing the background and related work concerning the challenges of building a high-quality event log as well as cognitive aspects of problem solving, from which the theory-guided objectives of the experimental study are derived. Section 4 describes the procedure that was used as an instrument in the study. Section 5 describes the experiment design. Sections 6 and 7 present and discuss the findings. The paper is concluded in section 8.

2. Background

We first describe the challenges related to building an event log. Next, we turn to some theoretical background on problem solving.

The event log

Gaining process insights

Information systems store *event data*, i.e. historical data about how information was manipulated or interchanged. This storage provides unique opportunities to analyse the process of how information flows in an organisation, or to deduce insights on the supported business process. For example: are claims being responded to in an appropriate time frame? The field of process mining offers techniques to conduct these tasks, an overview is provided in van der Aalst (2011). A prerequisite is the availability of an *event log*, to tie events uniquely to process instances. While event logs are largely available in process-aware information systems, this structure is not inherent in all kinds of information systems, in which case the event log often needs to be *built* from the available data.

The event log structure

An event log contains events related to a single business process captured by an information system. It covers *what* has been done, *when* it was done, by *whom* it was done (can also

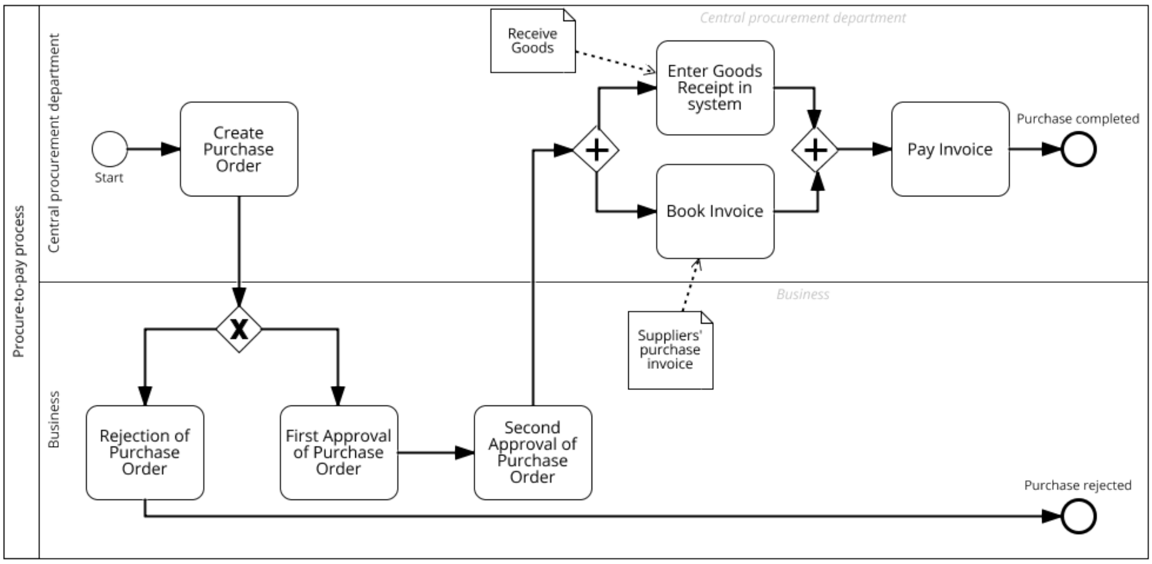


Figure 1.: Example process model of a procurement process

have additional descriptive data), in relation to *which* process instance, also known as a case. Therefore, we could state the following three requirements that are satisfied by an event log. Firstly, the event log contains a case identifier which relates multiple events to a specific case. Secondly, each event has a label that relates an event to a well-defined step or activity instance in the process, identified by an activity name. Thirdly, each event has a timestamp such that there exists a total order between the events related to one case. Additionally, an event log can contain case and event attributes other than the case identifier, activity name and event timestamp. Examples of such attributes are resources and costs.

To illustrate a possible event log structure for a classic procurement process, consider the process model in Figure 1. Such a standard process is often supported by an ERP system. The BPMN process model presents a simplified version of the implemented process in the system. The process starts with someone from the central procurement department creating a purchase order (PO). This PO is then inspected by the business department, which makes a decision (indicated by the cross - XOR gateway) to either approve the PO or reject it. If the PO is rejected, the process ends there. Otherwise, a second approval is needed, after which two activities are executed in parallel (indicated by the plus - AND gateway): the goods receipt is entered in the system and the invoice is booked. The process ends after the invoice is paid.

Each of the activities in the process is logged by the supporting information system in the database. A fragment of an example event log of the procurement process extracted from the database is shown in Table 1. The case that is followed throughout the process is a purchase

Table 1.: Example event log of procurement process

Case ID	Event ID	Timestamp	Activity	Resource
1	1	2018/08/16 10:16	create purchase order	Joyce
2	2	2018/08/16 13:07	create purchase order	Steve
2	3	2018/08/17 12:11	reject purchase order	Max
1	4	2018/08/17 14:13	1st approval PO	Jim
1	5	2018/08/17 17:25	2nd approval PO	Mike
1	6	2018/08/28 08:47	enter goods receipt	Dustin
1	7	2018/08/29 09:34	book invoice	Nancy
1	8	2018/09/16 11:46	pay invoice	Jonathan

order. Each purchase order is assigned a unique identifier shown in the 'Case ID' column. Each line in the log refers to an event which is assigned a unique identifier in the second column. For each event, a timestamp is logged together with the related activity and the resource that executed it, i.e. the information in the third, fourth and fifth columns. The first case in the fragment refers to a purchase order that leads to a completed purchase, while the second case concerns a rejected purchase order.

The event log architecture relates to the choices of the structure of the event log: which entity or document in the system is the case, which events are included, which attributes are needed to describe the events? This is discussed further in this Section and also in Section 4.2.

Relational database versus event log

The omnipresence of Enterprise Resource Planning (ERP) systems to support business processes, is often seen as a viable source of data for process mining projects. As a result, many event log building tasks involve the 'conversion' of a relational database into an event log. An ERP system organizes information in *documents* that are related to each other through one-to-many and many-to-many relations, and transactions relate to these documents. In a procure-to-pay process for example, *purchase requisitions, purchase orders, goods receipts*, and *invoices* are classical documents to organize information around. Transactions, like *'approve order'* and *'pay invoice'* relate to changes of these documents. Between the documents, different relations can hold. One purchase order for example might result in multiple good receipts. This is called *data divergence*, since one case (a purchase order) is related to multiple events of the same type (Lu et al., 2015). The opposite, *data convergence* holds when one event is related to multiple cases. Think for example of an approval of a purchase order that relates to three preceding purchase requisitions. The choice between the different solutions should be taken based on the business needs that drive process analysis.

The task of building an event log

Building an event log holds two major phases: deciding on the architecture of the log at first, and feeding the architectural frame with data next. The challenges of building the event log that are addressed here are related to the former part: deciding on the architecture (or the structure) of the log. This phase is a prerequisite for the next phase, of extracting data for populating the log, whose challenges are not in the scope of this paper. The log architecture design involves considering questions such as which events to select, tied to which entity as a process instance, and at which granularity level. This is comparable to the decisions an architect takes when drawing the structure of a new house or facility. Some decisions may stem from technical constraints only, but bear substantial consequences for the end user. The architect needs therefore to understand the consequences of the possible choices when making the decisions. We argue that a good event log builder, like a good architect, should be aware of the decisions that are taken and understand their consequences.

Theoretical background on problem solving

Due to the inherent mismatch between the structure of a relational database (multi-dimensional) and an event log (flat), different aspects need to be evaluated and positive consequences of design options should outweigh potential downsides. To involve business (or abstract) considerations in a technical task, is not intuitive for novices (Feitelson, 2015; Huang and Burns, 2000). Therefore, we argue that building an event log structure for a specific project goal requires adequate guidance.

We consider the creation of a log from a relational database as a problem-solving task. Furthermore, log creation entails a large number of operations, intended to address relevant concepts, integrate them and structure them in order to attain specific goals. It can hence be classified as a complex cognitive task, based on the number and nature of component skills involved and the complexity of the goal hierarchies of the problems that must be solved in the task domain (Paas and Van Merrinboer, 1994).

When human beings are solving problems, they form internal representations of the problem, to which solution procedures, stored in their brains, can be applied (Larkin and Simon, 1987). To this end, the information of the current problem is mapped by analogy to cognitive schemata that can be acquired by learning. We note that log creation entails two main sub-tasks that emphasize different types of information and may hence interfere with each other (Shaft and Vessey, 2006). One is the technical task of transforming a relational database into a log structure, emphasizing concepts related to these two structures; the other is the business task of supporting a specific analysis goal, focusing on the business concepts and their relations with the log structure.

Cognitive load theory (Sweller, 1988) is concerned with the manner in which cognitive resources are focused and used during such processes of problem solving. In particular, complex cognitive tasks, such as event log building, entail high *intrinsic* load and require strategies for alleviating it. Building on ample analysis of cognitive load sources, much research efforts have been devoted to the development of strategies for effective learning and problem solving. For example, according to Kirschner, Sweller, and Clark (2006), empirical evidence shows that when dealing with novel information, learners should be explicitly shown what to do and how to do it to achieve required tasks. Sweller (2011) stresses that high intrinsic cognitive load is caused by high interactivity of the elements to be processed. High element interactivity requires simultaneous processing of these elements, and thus entails high intrinsic cognitive load. Reducing the interactivity and the required simultaneous processing can therefore reduce the cognitive load associated with a task. Considering the log building task, it is, as described, a complex task, entailing a number of highly interacting elements or sub-tasks. However, if we break it down to a series of smaller sub-tasks which can be performed sequentially, where each sub-task addresses a limited number of elements, this may reduce the peaks of intrinsic cognitive load in this task, spreading the load along the different sub-tasks. Reducing the peaks implies reducing the chances of cognitive overload, which is known to hamper problem solving. We expect that an explicit procedure that would serve as guidance in this task would have such a effect.

3. Related work

The discrepancy between relational databases and event logs led to some related research efforts. Calvanese et al. (2015) present an ontology-driven approach to extract event log information from legacy relational databases, supporting domain knowledge.This approach provides a technical solution for the task, also addressing the data from a semantic perspective. However, business considerations that relate to the intended goal of analysis are not supported. Also González López de Murillas, van der Aalst, and Reijers (2015); González López de Murillas, Reijers, and van der Aalst (2016) provide an approach to technically flatten relational databases into an event log, solving the technical problem but leaving business goals and considerations out of scope.

de Murillas, Reijers, and van der Aalst (2018) present an approach that, rather than creating an event log from a relational database, relies on available transaction or 'redo' log and uses it directly for process mining. Such logs are typically available in relational databases and capture data operations, such as insert, update, or delete. However, as indicated by Tsoury, Soffer, and Reinhartz-Berger (2018), they are typically at a level of granularity which is not suitable for business analysis. In Jans and Soffer (2017), the quality impact of event log building decisions is discussed. One of these decisions is the selection of the process instance. In that same line, Lu et al. (2015) pinpoint the difficulty of mapping events to cases in an ERP system to the fact that these systems relate events to *documents*, and not to a process case.

To summarize, one could say that the challenges of one-to-many and many-to-many relations have been addressed mainly from a technical viewpoint and different solutions are possible (Raichelson, Soffer, and Verbeek, 2017). But the business needs that drive the process mining analysis are not addressed by the technical solutions that were proposed. Since our aim is to provide guidance in producing a valuable log for defined business needs, we focus on this aspect of the transformation of database information to an event log.

4. Procedural guidance

4.1. *Purpose of the study*

Our aim is to investigate possible support for novice process analysts in building an event log for a specific business goal. Following the above discussion, this could be accomplished by providing a procedure that should break the task down to a sequence of distinct steps, and guide the analysts through these steps. Our aim is to investigate whether and how such a procedure affects the log building process and its resulting product. As a consequence, we set two objectives for the study:

Objective 1: to investigate to what extent a procedural guidance affects the performance of the complex task of log construction, as compared to an undirected task performance. The study employs an experimental research method to conduct this investigation.

Objective 2: to analyse and understand the problem-solving process that is followed with and without a procedural guidance. In particular, we wish to explain the differences in task performance by how the process is carried out, and examine whether these differences equally apply to the technical aspects and to the business aspects of the task. The study employs a verbal protocol analysis to conduct this investigation.

4.2. *Building valuable event logs - A procedure*

As instrument for this study, a procedure was developed and used in the experiment. The procedure itself is not fully elaborated in the paper, given its size. It is however freely available on `http://researchgate.net`, doi `10.13140/RG.2.2.11343.69289`.

The procedure systematically addresses fundamental concepts that bridge relational databases and event logs, and breaks the task into a sequence of steps to be taken. Other procedures may be proposed, and differ from the one used here in granularity or in grouping sub-tasks into distinct steps, but the same building blocks and their relations are expected to be included. Therefore, we suggest that the results of our experiment may not be biased towards this specific procedure. Further, the procedure is a product of best practices and has been used in different industries (utility services, internet and telephony provider, and chemicals) and for different types of projects, related either to operational excellence or to financial statement auditing. These experiences support the suggested sequence of steps in the procedure. The following paragraph describes the objectives and design of the procedure. Next, the procedure is shortly summarized.

Objectives and design of the procedure

The procedure design seeks to meet the following objectives: it will

(1) be possible to be employed by process analysts with limited knowledge of process mining, i.e. familiar with the concept of process mining and the prerequisites of an event log, but not necessarily experienced in conducting a process mining analysis in person.

(2) increase the analysts understanding of the decisions and their consequences, related to the choice of process instance, activities and attributes.

(3) break the log building task into an easy-to-follow sequence of well-defined actions to be performed.

In order to support the construction of an event log that is valuable for a specific project, the designed procedure is based on three core principles. First, the procedure directs the user's attention to targeted sources of information. Second, it uses a realistic running example, based on a real ERP system database structure, to assure a close link to real-life process mining projects and support reasoning by analogy. Third, attention has been devoted to the organisational and managerial aspects of integrating knowledge of the domain and the IT experts and the process analyst. The aim of this principle is to assure that business considerations are taken into account during the log building process.

Procedure instrument

The procedure includes nine steps, briefly described here.

Step 1: Set primary business goal of the project The main goal of this step is to establish the business goal that is considered as a 'must have' by the project sponsor. This is often either to increase the efficiency of the process, or to assess its compliance with some predefined procedure or regulations. It is recommended to distinguish between must-have goals and nice-to-have goals to differentiate the real goal from additional desires. Also, in order to have the goals understood in the same way by all parties, articulating key questions that need to be answered during the project is important.

Step 2. Identify key process cornerstones The key cornerstones are the key activities in the process, according to the project stakeholders. At this stage, these cornerstones are not yet verified as possible (logged) activities in the event log, hence the differentiating terminology to avoid confusion. This step involves determining the boundaries of the process under investigation and the core activities that are of interest to the project stakeholders.

Step 3. Identify key tables For a log to serve a specific purpose, only a subset of the available tables is relevant. The third step supports the selection of these relevant tables. Note that a typical database includes a large number of tables. However, the intended log should relate only to those that are relevant for the one specific process that should be studied. The key process cornerstones that are identified in step 2 are used as guidance for the table selection.

Step 4. Identify relationships between tables Starting from the identified tables, the database structure sets relationships between tables, typically in the form of foreign keys. A main issue in this step is to articulate the cardinality of the relationships, which can be one-to-one, one-to-many, or many-to-many. The cardinality of relationships has a major role in the following steps, and is often related to the granularity level of the entities whose data is stored in the tables.

Step 5. Select the process instance document In order for an event log to reveal valuable business insights, it is recommended to select a document that is related to the process as process instance (the case in the log), even if information from one document is stored in multiple tables. Starting from the agreed upon goal, the key cornerstones, and the underlying database structure, this procedure step requires a decision on which document that could be. Only in the next step, a follow-up decision is made on which particular aspect of that document (and hence, which table) will be followed. Specific considerations relate to how the primary analysis goal would be affected by the possible selection. For example, for an efficiency-related analysis, an important capability would be to identify cases that had started and never completed. To support such analysis, the process instance should be possible to trace the beginning of the process, thus it would be recommended to select a document which starts the process (called 'start document', e.g. purchase order) as process instance. In contrast, a compliance-oriented analysis may aim to reveal cases which have not been started normally. Hence, selecting a document created towards the end of the process (called 'end document', e.g. invoice) may better support this kind of analysis.

Step 6. Select the process instance level Once a process instance document is selected, another major decision relates to the granularity level of this instance. In particular, some documents can be represented by two or more database tables. A typical scenario is when a high-level ("parent") document has several detailed entries in a related table ("child"). As an example, consider a purchase order which has several order lines, each for a different item. A main issue considered for this decision is that process activities may too relate to different granularity levels (e.g., sending an order relates to the parent, while receiving an item relates to the child). Selecting one over the other might lead to either artificial duplicates of parent-level activities in the log (in case the child was selected as the instance level) or multiple repetition of child activities in untraceable self-loops (in case the parent was selected as the instance level). This step of the procedure considers alternative solutions and possible aggregation functions to support the business goal of the analysis.

Step 7. Select activities After selecting the process instance, the activities can be selected. Note that activities are not explicitly stored in databases, but are deduced from timestamp information. Candidate activities would therefore be database entries that have timestamp attributes. The actual selection can be guided by two concerns: select only relevant activities, considering the process instance and the goal of analysis, and maintain an appropriate granularity level. An additional consideration at this step is whether or not to consider variants of activities as distinct activities. As an example, assume a delivery document has a "delivery type" attribute, whose values can be "express" or "regular". It may be beneficial to define two different activities, for an express delivery and for a regular one, so they can appear separately in the resulting analysis, and emphasize the existence of different variants of the process. These variants would not be apparent otherwise.

Step 8. Select attributes From among the available attributes that exist in the database, this step identifies the ones that can be relevant for the analysis, guided by the business goal.

Step 9. Relate attributes to activities Following the XES standard (Verbeek et al., 2011), attributes can be related to the case or to an event (case and event attributes). This step determines to what each of the selected attributes would relate, considering the stability of its value across the process. In other words, attributes whose value may be determined by different activities across the process (e.g., executing user) are related to activities (events in the log), while those whose value remains stable across the process (e.g., supplier) are related to the case. In this step, some attributes may be identified to aggregate in order to have more meaning. Other attributes, of which different values represent different activities, could be considered to incorporate as additional activity. These are called attribute-dependent activities and might not have been considered as activity before (when merely looking at timestamp attributes).

5. Experiment

The first objective of the experiment was to examine the log building performance with and without a procedural guidance. Following the theoretical discussion, we hypothesized that a procedural support would lead to a better performance than an undirected process. The second objective of the study was to gain an understanding of the log building process when such support is given and how differences in the process are related to the resulting performance. To reach these objectives, we conducted an experiment and a verbal protocol analysis of the output of the experiment. In the following sections, more details are provided about the experiment design, the subjects, the assignment, the execution of the experiment, and the measurement of affecting and affected factors.

Experimental strategy

We describe the experiment following the framework suggested by Gemino and Wand (2004) in the field of conceptual modeling. Although our task is not about conceptual modeling, we consider the dimensions of the framework as relevant for our study. The first dimension is the affecting variables: the factors that can influence the outcome of the process. In our experiment, the support that is given (with procedural guidance or without guidance) is the manipulated affecting variable. The second dimension is the one of affected (outcome) variables; these are the dependent, observable outcomes that are the primary interest of the experiment. Within this dimension, two categories are proposed: (1) the focus of observation to which the affected variables relate (*process* and/or *product*), and (2) the criteria for comparison (*effectiveness* and/or *efficiency*) manifested by the variables. For Objective 1 of our study, log building performance relates to both effectiveness and efficiency of the log building process. These can be evaluated at a *black-box* view or at a *white-box* one. Using the (common) black-box view, the effectiveness of the process is assessed by focusing on the product (its outcome - the event log), whose quality is a comparison criterion as an approximate measure of the effectiveness of its creation process. The efficiency of the log building process is evaluated at a back-box view based on its time. We further propose "white-box" measures of effectiveness and efficiency that complement the black-box measures and allow explaining differences in the performance. These measures rely on the findings related to Objective 2, whose focus of observation is the process, with the intention to qualitatively study it and characterize its steps and considerations with the different levels of guidance.

Experiment Design

The experiment was set up as follows. Students were taught in two groups on building an event log: a procedural-oriented (treatment) group and a control group. In both groups they were then provided with a use case, and their assignment was to build an event log for a given business goal of analysis. In pairs, the students discussed the case in order to complete the assignment together. The pairs were formed within a single group ('procedural' or 'control'). We opted to have discussions in pairs of students for two reasons. The first reason is that working in teams with intensive interaction is a realistic setting that most likely will also occur when executing this task in a professional setting. The second reason is more pragmatic in that it allows us to capture the thinking process of the students. Given the resemblance between discussing this task as part of an assignment and discussing a similar task with a colleague in a professional setting, we assume this design option may not bias our findings (as might be the case for more artificial settings, such as thinking aloud). At a black-box view, the output of the assignment was dual: a written answer and a recording of the pair discussion. The written 'solution' was graded and quoted as part of their course grade (15% - after levelling out possible group differences). The assignment grade was used as an approximate measure for the *effectiveness* of the process: how valuable is the event log that was created. The time needed to fulfill the assignment was an approximate measure for the *process efficiency*.

The coded recordings on the other hand reflect the *problem-solving process*. The recordings were employed as input for a verbal protocol analysis, revealing the thinking process of the subjects as they address the task. Using pair discussions for a verbal analysis has the benefit of following the thinking process in a natural setting, rather than the artificial setting and bias that may be associated with think-aloud (Russo, Johnson, and Stephens, 1989).

In addition, two questionnaires were filled out by each participant: one before starting the assignment and the other after completing it. The questionnaires were used to measure the (potential) affecting factors before and after performing the task (see below on p. 15).

Subjects and instruction

Twenty-two master students of the program *Business and information systems engineering* at Hasselt University in Belgium were taught on the topic of building an event log. The class was part of the course *Business process analytics*, where the most prominent concepts and techniques of process mining are taught. The twenty-two students were split in two equivalent groups, to receive their 'event log building' class at the same day. The structure of the class was equal for the two groups, showing the following similarities:

- A single instructor gave both classes.
- Each class started by a discussion of the XES standard. Basic knowledge on this standard is a prerequisite for building a valuable event log. The students read some material as preparation for class, and 15 minutes were devoted in both classes to discuss potential questions.
- In both classes, the same exercise served for experiencing the challenges of turning a relational database into an event log. The exercise contained the tables from a relational database and the students were asked to build an event log from it, as if it is a straightforward transformation.
- In both classes, the students had 30 minutes to work on this exercise independently, experiencing the different design decisions they needed to take.
- Next, 75 minutes were used to discuss underlying issues of building an event log, tying the design decisions and the related concepts with the consequences.
- The same presentation slides on underlying concepts were used as teaching material.

The crucial difference between the two groups was that the procedural group got the content presented as a sequence of actions to execute, while in the control group the content was presented as separate building blocks without an operational guidance on how they should be used together.

Assignment

The overall assignment was to provide a structure of an event log (process instance, activities, attributes) for this case, along with the advantages and disadvantages they identify in their own solution. The exact formulation of the assignment, along with the presented case, can be found in appendix.

The case was designed to resemble a real-life setting as much as possible. It contained two main parts: a textual and graphical process description, and an overview of database tables. The textual and graphical process description was given in a format similar to the type of information that one would receive in a face-to-face meeting with the process stakeholders, including the main goal of the analysis. This includes: spoken language text, a semi-structured process description, a mixture of 'nice to have' and 'must have' expressions describing the analysis goal, additions of noise, and a process graph in layman notation, open for ambiguous interpretation. The overview of database tables is more structured, provided in the format of a data dictionary.

The noise that was added to the process description entailed four categories: (1) activities that were clearly out of scope according to the process analysis aim, (2) activities on a too low level of granularity for the information system, (3) names of responsible persons, functions, and roles, and (4) activities of which no timestamps are captured.

Experiment execution

The study was conducted in December 2015 at Hasselt University, Belgium. The session on the assignment and the test was organized as follows: the students were located in two classrooms (one for the procedure-group, and one for the control group). In each classroom, sufficient space was available for the teams (five and six teams) to sit at a comfortable distance from each other, not to interfere with other discussions. Before the discussions started, the students received the first questionnaire individually. Instructors mentioned that no grades would be assigned to the answers on the questionnaires (neither the prior, nor the follow-up questionnaire), only to the written event log solution that each team would create. Once all questionnaires were handed in, the assignments were handed out, audio recordings were started, and time was recorded. No time limit was set. Pairs of students that finished the assignment stopped their recording, handed in their solution, and received the follow-up questionnaire, which they again filled out individually.

Measuring affecting factors

The primary affecting factor of interest is the procedural guidance, but other factors that might have an influence on the task were also measured. The perceived ease of understanding (PEOU) of both the assignment and the log-building process was measured by means of questions whose answers were on a scale from one to seven. Also, self-efficacy and the environment during the assignment were measured by means of scale variables (see appendix for the exact questions). The environment refers to the setting of the assignment, whether the students found they had enough time and whether the assignment was hard to execute or not. With regards to the PEOU of the assignment, the self-efficacy, and the environment, no significant differences were expected between the two groups. Only the PEOU of the log-building process was evidently expected to differ between the two groups. We did not measure 'past experience' since we assume that the variability between the participants is minimal and that the random assignment increases the chance that the groups have on average equal experience.

Although the affecting factors were measured quantitatively, statistical tests to compare group means would not be possible, given the small sample size. This small size is attributable to the focus on the problem-solving process, studied qualitatively by means of recorded and coded discussions. This is explained further in the next section.

Two other potentially affecting factors were used to balance subjects over the two groups: academic achievement and sex. It was expected that a higher academic achievement resulted in a higher task performance. It might also have impacted the problem-solving process, although we do not build any theory on the direction or intensity of this impact, since this is not the topic of this study. The other factor we controlled for was sex. Although we had no expectations that sex would impact the results, due to reported evidence of gender-differences in verbal communication (Hyde and Linn, 1988), it was safe to control for.

Measuring affected factors

The affected factors, or dependent variables, relate to the performance of building a valuable event log (Objective 1), first at a black-box view and then at a white-box view. The detailed white-box measures are based on the deep qualitative analysis of the log creation as a problem-solving process (Objective 2). We describe these accordingly.

Measuring the performance of event log building as a black-box

The black-box performance measures of building the event log relate to its effectiveness and efficiency. **Effectiveness** of the process relates to the quality of its product, namely, the solution the students handed in on the assignment. This was captured by an assignment grade. To reach an objective measure, the grades of the solution were based on a solution key, following a predefined grading scheme and scale (see Appendix). The solution key contained specific elements that needed to be addressed, like the goal and key questions, and expressed expected levels of linking decisions and consequences, like for example mentioning the timestamp field when it comes to activity selection. Further, we required arguments for the taken choices: students needed to justify why they took a certain decision. This is essential because the decisions cannot be classified binarily as right or wrong. Rather, we expect the decisions to be 'aware of the consequences', namely, made with the possible consequences taken into consideration. Consequently, the justification was needed to assess the quality of their reasoning behind the choice. To mitigate the risk that the students might not explain their justification correctly on the written assignment, the recorded discussions were also taken as input for the grade. It is important to note that the grading only related to the end result, and not to the process of how the log was built. The evaluation focused entirely on the connection of the design decisions and the understanding of the consequences. Consequently, we prevent a potential bias resulting from the procedural support on the evaluation procedure.

To check whether the assigned grades were fair, a second researcher evaluated four randomly-selected assignments. Then, Cohen's Kappa coefficient (Cohen, 1960) was calculated to measure the agreement between the graders. The coefficient is a ratio that expresses the possible agreement beyond chance that was actually observed. A maximum value of 1 means that graders are in complete agreement, a value of 0 or less means that the only agreement of the graders is due to chance. To take the ordinal character of the grades into account, the weighted version of Cohen's Kappa coefficient (Cohen, 1968) was used . The weighted Cohen's Kappa coefficient of our scores is 0.818, indicating a high level of agreement between the two graders.
While the assignment grade was used as a black-box measure of effectiveness, we were also interested in the efficiency of the log building process. We used the time spent on the assignment as a black-box measure of its **efficiency**.

Analysing the problem-solving process

For analysing the problem-solving process, the discussions of the students were recorded and coded using a 34-category scheme. This scheme was based on the procedure that was used for guiding the procedural group. This scheme was elaborated during the coding of the first two discussions to allow for a finer-grained distinction of some concepts. No further adaptations were necessary. The coding categories concerned three dimensions: procedure steps, concepts addressed, and process reflections. These dimensions were later on used to explore the differences of the problem-solving processes in both groups.

The first dimension, the **procedure steps** should be seen as sub-tasks of building an event log. Segments of the discussion text were coded based on their relation to steps of the guidance procedure (for both groups). The following steps were coded:

- State goal (P1)
- Identify process cornerstones (P2)
- Identify key tables (P3)
- Identify relationships between tables (P4)
- Select process instance document (P5)
- Select process instance level (P6)
- Select activities (P7)

- Select attributes (P8)
- Consider attributes to incorporate in activities (P9)

When building a valuable event log in a conscious manner, sub-tasks corresponding to all steps of the scheme (termed steps hereafter for brevity) should be addressed. By coding the different steps in the verbal protocols, a trace of the different steps that the students discussed was captured. Abstraction was made from the wording they used, since the control group was not expected to use exactly the same wording as mentioned in the procedure. If a procedure step was readdressed later in the discussion, this was tracked again. That way, we gained insights into how structured their reasoning was: whether they iterated a lot, or whether a straight forward approach was applied.

The second dimension that was coded was the **concepts** that were discussed when addressing the procedure steps (as coded). Within one procedure step, concepts were only coded on their presence, not on their sequence. As opposed to the procedure steps, identified by their goal, it only takes one sentence to address a concept. Attempting to trace the sequence of concepts might lead to bias in the coding. For example, consider the following situations: 1) a lengthy discussion on a certain concept A is interrupted by a single sentence on concept B, only to address B properly after finishing the discussion on A, and 2) a discussion on concept A is interrupted by a lengthy discussion on concept B, switches back to A and ends with B again. Both cases would be coded as A-B-A-B, while there is clearly a difference in the line of thinking. Therefore, coding was limited to the level of presence (*Which concepts have been discussed in each procedure step*), leaving out possible repetitions and a sequence analysis. The following concepts (see explanations in the procedure description) have been coded:

- Project Goal (C1)
- Nice to have project goal (C2)
- Must have project goal (C3)
- Efficiency goal (C4)
- Compliance goal (C5)
- Key questions (to articulate goal) (C6)
- Process instance (C7)
- Start document (C8)
- End document (C9)
- Cardinality of tables (C10)
- Timestamp (C11)
- Activity (C12)
- Cornerstone (C13)
- Attribute (C14)
- Case attribute (C15)
- Event attribute (C16)

- Self-loops (C17)
- Artificial multiplication (C18)
- Attribute aggregation (C19)
- Document (C20)
- Attribute-dependent activity (C21)

The listed concepts have all been addressed in class, in both groups. As a baseline, we created a reference heatmap depicting the ideal combination of concepts and procedure steps in case the procedure would be applied by an expert. The concepts that *must* be discussed are assigned a weight of 1 (mean presence of concept per procedure step). The concepts that *could* be discussed, but are not mandatory for a thorough understanding, are assigned a weight of 0.5. This *ideal* heatmap is presented in Figure 5, next to the actual heatmaps of the two groups as part of the findings. Insights gained from both the procedure steps and concepts discussed, are treated as information on the problem-solving process, the subject of investigation of our second objective.

The third dimension that is coded, relates to **process reflections**. Similar to other studies (Soffer, Wand, and Kaner, 2015), we kept track of expressions that reflect on the process they were experiencing. The following categories were coded:

- Explicit difficulty expressions (ME1)
- Evaluating alternatives (ME2)
- Revisiting text (ME3)
- Strategic planning of the work (ME4)

6. Findings

Descriptive statistics

Twenty-two students participated in the experiment, making 11 teams of two students each. However, after analyzing the recorded discussions, it was recognized that one team did not master the concepts of process mining and databases at the desired level. As this team was not a good proxy for practitioners who would be assigned this task in real life, it was discarded from further analysis. The descriptive statistics summarize the remaining 20 subjects, representing 10 teams: five teams in the procedure group, and five teams in the control group. The groups were formed to be balanced in terms of academic achievements of the students and their sex (see Table 2).

With respect to the affecting variables *perceived ease of understanding* of both the assignment and the procedure, *self-efficacy*, and the *environment of the assignment task*, no discriminating effects were found. See appendix for more details.

Table 2.: Distribution of sex and academic achievement over groups

Group	Sex		Academic achievement
	Female (#)	Male (#)	Mean (st.dev.)
Procedural	3	7	67.625 (10.514)
Control	3	7	64.25 (7.63)

Event log building performance (black-box measurement)

Objective 1 of the study was to examine to what extent a procedural guidance supports an effective and efficient log construction process. At a black-box view, this would be manifested as differences in the quality of the event log, as well as its construction time, with procedural guidance and without it. To measure the quality of the final product, the written solutions were examined, along with the discussions. At a black-box view, the grades on the assignments served as an approximate measure of the effectiveness of building an event log. The time spent on this assignment an approximate measure of its efficiency.

A frequency table of the achieved grades is presented in Table 3. The procedural group teams have performed better than those of the control group. The procedural group's grades are mostly B and C grades, with one A grade, whereas the control group did not gain any A or B grades, only C and D grades, with an emphasis on D. This suggests that the procedural support increases the effectiveness of building a valuable event log, which is clearly a direction for further research.

Table 3.: Frequency of assignment grades

Grade	Procedural	Control
A	1	0
B	2	0
C	2	2
D	0	3

With regards to the efficiency of event log building, the duration of completing the task was measured. The mean (std.) duration, expressed in minutes, was 135 (39.287) and 134.4 (26.539) for the procedural and control group respectively. This implies that there is no observed difference in the average efficiency when using a black-box assessment. We return to this later on when analyzing the problem-solving process and provide finer-grained (white box) efficiency indicators.

To conclude on the impact of procedural support on the performance of building an event log, our findings, analyzed at a black-box view, suggest that a higher quality might be reached when providing procedural guidance to support the analysts in the task. No differences are observed in terms of efficiency.

Problem-solving process

Having observed a better log building effectiveness of the procedural group, our next objective was to gain an understanding of the problem-solving processes followed by the groups and the differences between them. To gain such understanding, we turn to a verbal protocol analysis. The ten assignment discussions were captured in 1347 minutes (approximately 22 hours) of recording. Three categories were coded in the recorded text: the process steps, the discussed concepts, and the explicit mentioning of process reflections. In the following paragraphs, insights on these three levels will be discussed.

Process analysis

Process step frequency The first dimension that was coded, is the dimension of the process steps that were discussed in all teams. In Table 4, the total frequency of addressing each process step is reported for the procedural and the control group. Apparently, the control group performed several repetitions of some process steps (selecting activities and attributes). However, they often skipped identifying process cornerstones and the related key tables. Since the procedure addresses the fundamental concepts of bridging relational databases and event logs through domain knowledge, addressing these steps is essential.

Table 4.: Frequency of process steps executed

Process Step	Procedural (5 pairs)	Control (5 pairs)
State goal	5	7
Identify process cornerstones	6	1
Identify key tables	4	0
Identify relationships between tables	5	5
Select process instance document	6	8
Select process instance level	4	9
Select activities	6	16
Select attributes	6	13
Consider attributes to incorporate in activities	0	2

Process maps To have a better view on the identified iterations, an event log was built out of the coded recordings. Selecting only the coded procedure steps, ordered chronologically and linked to the discussion team (as process instances), two distinct event logs were built: one for the procedural group and one for the control group. Figures 2 and 3 show the related process maps, when visualizing all the paths of the two logs using the Fluxicon Disco tool[1].

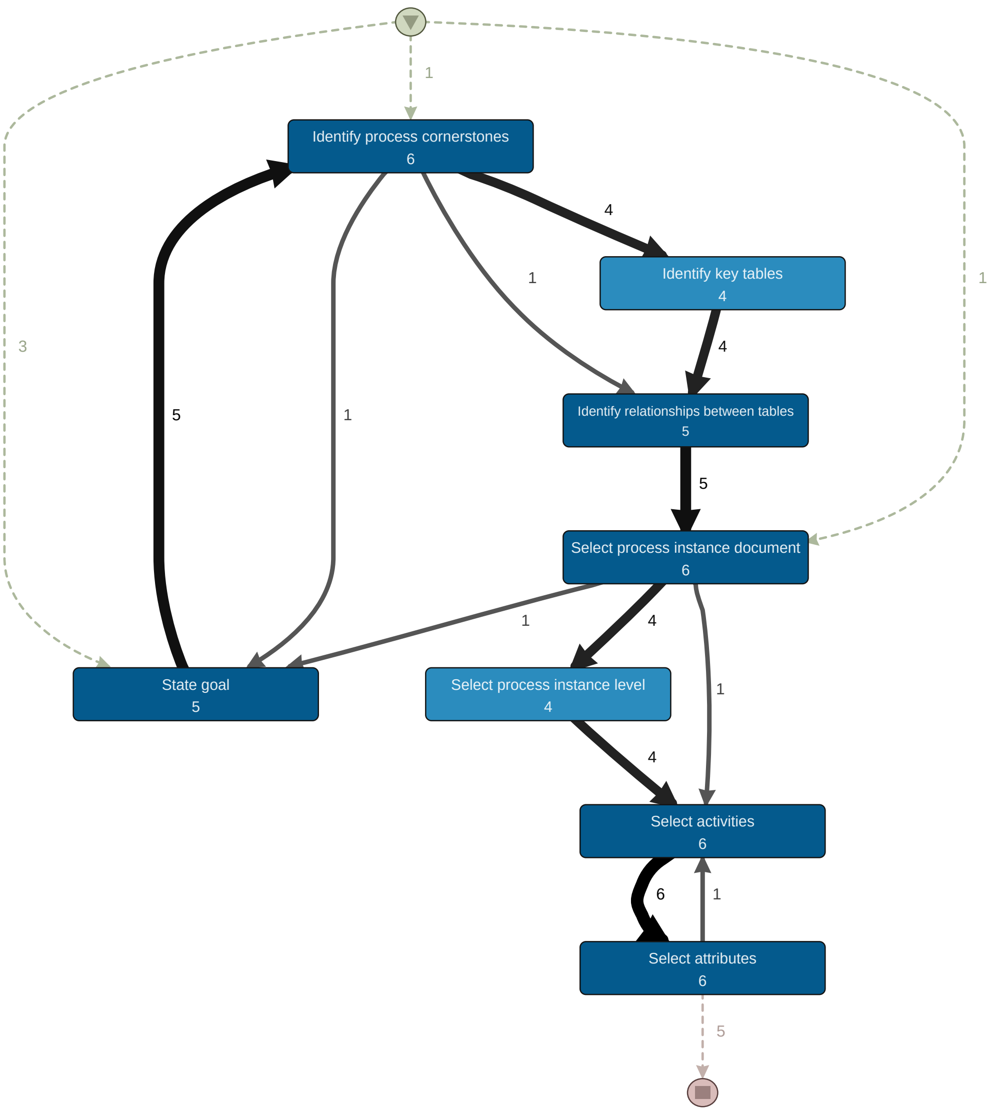


Figure 2.: Flow of process steps during discussion in the teams of the procedural group

Based on the process maps, the previous observations are confirmed visually. First, the procedural group shows a more streamlined flow than the control group, as expected. Second, indicated by the color intensity of activity boxes, the frequency of procedure steps is more uni-

[1] www.fluxicon.com/disco

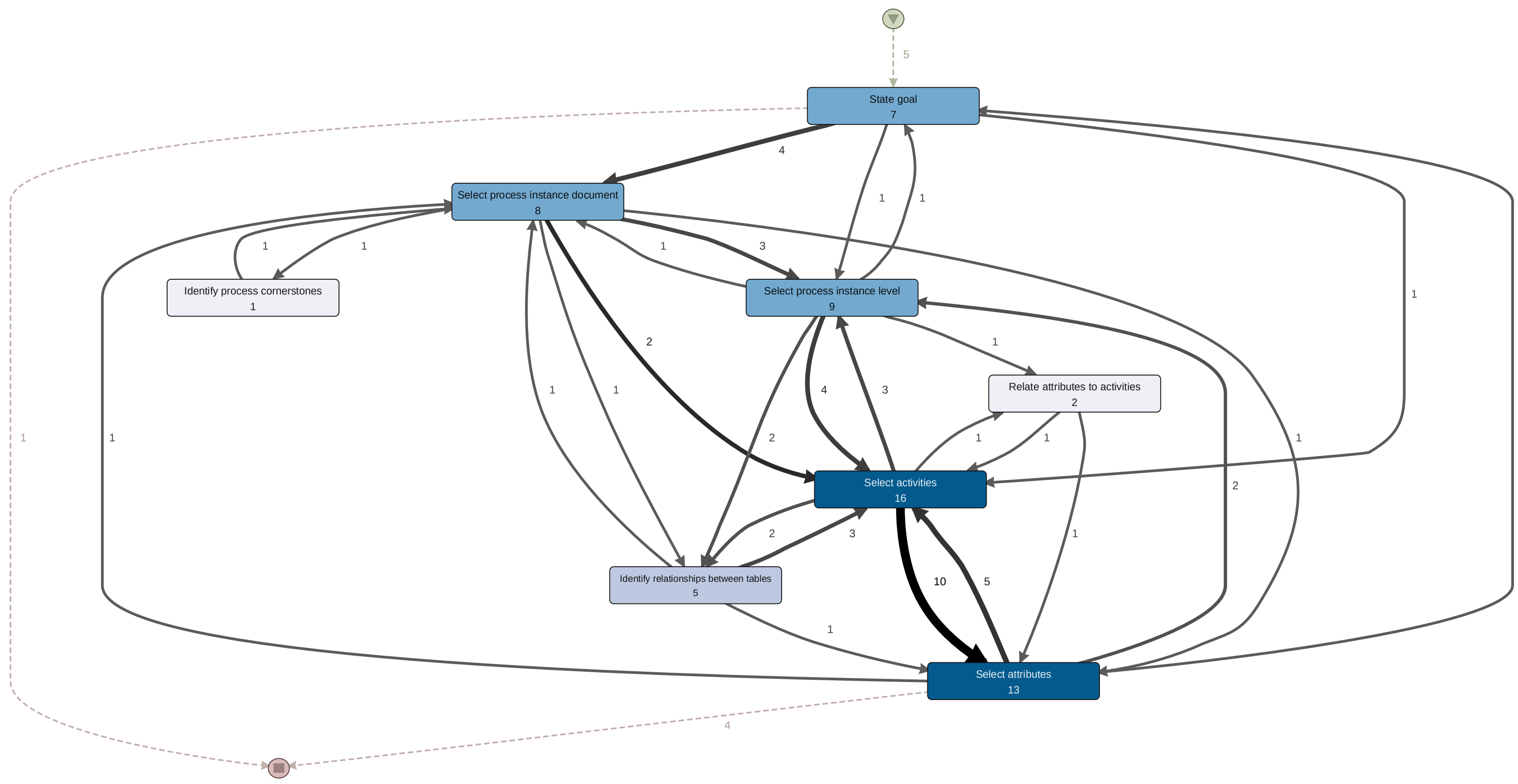


Figure 3.: Flow of process steps during discussion in the teams of the control group

form in the procedural group. In the control group a clear emphasis is noticeable on two steps that were revisited several times by each team (likely addressing highly interactive elements), whereas other steps were addressed less frequently. These observations suggest a more systematic process in the procedural group. While this may seem natural and expected, the details of the steps that were taken by each group are of interest. In particular, more insights into the concepts that are addressed when iterating process steps were needed.

Process traces Where the process maps suggest differences on a global process level, analysing full process executions provides insights on a finer-grained level. To this end, we analysed the individual process executions. A process execution of a team is represented by the sequence of activities throughout the process, called a trace. In Figure 4, the traces of the team discussions are visually represented using a line graph that connects the activities in order of occurrence. The upper five traces are the traces of the procedural teams. The lower five traces are the ones of the control teams.

The process traces of the control group are, in general, longer and less uniform than those of the procedural group. Where the procedural group fulfilled the assignment in a rather straightforward manner, the control group seemed to iterate back and forth to complete the same assignment. As an example, consider team H in the control group, showing the longest process trace. This team continued switching from selecting process instance level, to selecting activities and attributes and back. Also, the team was one of the two teams who actually considered to store attributes as activities (relate attributes to activities), which at first sight might be considered an appropriate action. However, this is a step that only makes sense when all other decisions are made and this team still iterated back to the selection of process instance level, activities, and attributes afterwards.

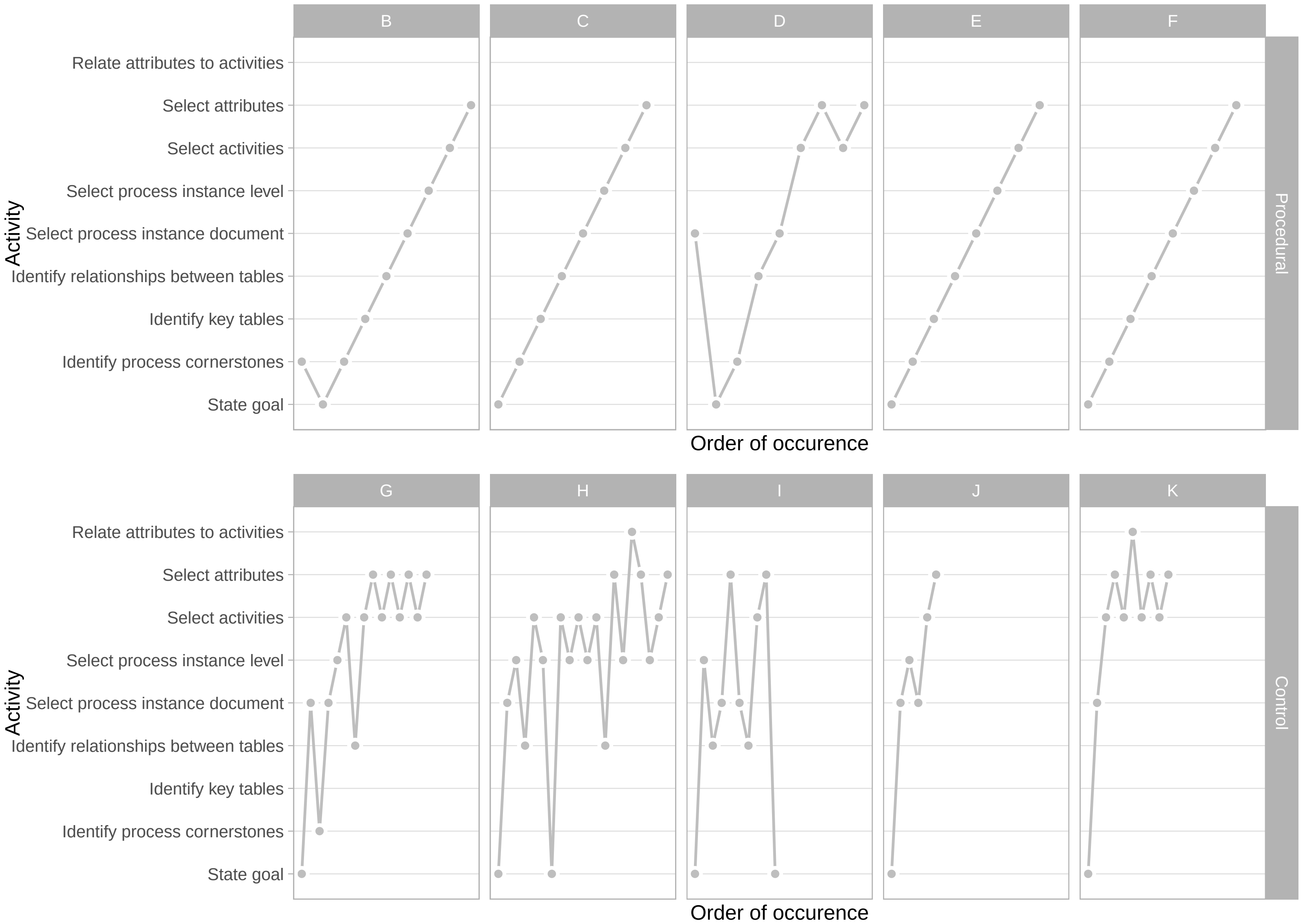


Figure 4.: Line chart diagram of event log building process (Created in bupaR, an R-package for process mining (Janssenswillen et al., 2019))

Process flows To systematically analyze the differences in the process traces, we follow the flow as suggested by the procedure, as these capture some real dependencies that exist between the different steps. If the procedure flows are followed, the source activity includes preparatory actions that set the ground for the target activity (the following step), so it can be performed in an appropriate manner. Of course, it is also possible to interrupt a flow with another step, without hampering these prerequisites of the target activity. Still, we will organize our flow analysis around the flows of the procedure. We take this approach to compare the process behaviour in both groups in a systematic way. Without claiming that the procedural flow is the norm, it allows us to identify the differences between groups in a consistent manner.

State goal → Identify process cornerstones

Subjects in both groups stated goals of the project. The procedural teams all stated their analysis goal quite in the beginning of the process. Four teams started with this, and one team tried to start the identification of cornerstones, but then turned to state the goal, probably experiencing the prerequisite character of stating the goal for the next step. The control teams all started their process with stating the goal of the analysis, with the difference that two teams returned to this step later in the process.

While in the procedural group identifying cornerstones was almost always preceded by stating the goal (five out of six times directly), the control group did not pay much attention to identifying cornerstones at all. There was a single execution of cornerstone identification, which was also preceded by stating the goal (although not directly).

Identify process cornerstones → Identify key tables

For the procedural group, this flow was followed by all the four teams who were engaged in identifying key tables. There were no repetitions in the activity of identifying key tables, which seems to indicate that all prerequisites were fulfilled to execute this step (at least in the experience of the students). In the control group, identification of key tables was not present at all.

Identify key tables → Identify relationships between tables

In the procedural group, all teams identified relationships between tables. Four of the five teams did that (right) after identifying the key tables. The fifth team never identified key tables and started this step after stating the goal and identifying the cornerstones. Within the control group, only three teams attended the step of identifying relationships between tables. This step was not preceded by an identification of the key tables, but done in the context of other activities (select process instance document, level, and activities). The structure of the flows seem to indicate that, later in the building process, the students experienced the required output of this step in order to proceed with the other steps. Also, two of the teams needed to revisit this step, possibly due to a lack of a proper basis at first, suggesting a lower understanding of the necessity of this step.

Identify relationships between tables → Select process instance document

To thoughtfully select a suitable process instance (document), a clear notion of the information structure in the available data is needed. Therefore, understanding the table relationships is a required preparation for selecting the process instance document. The teams in both groups performed the steps of selecting the process instance document. However, while in the procedural group all teams, in the end, selected a process instance document after having identified the relationships between the tables, this was not the case at all within the control group (see also figure 4). Only one team (team I) performed the preparatory step somewhere before making this crucial decision. In the other cases, process instance selection was performed at the beginning of the process, immediately after stating the goal, without an appropriate basis.

Select process instance document → Select process instance level

Selecting the process instance level can only be meaningful after the process instance document has been selected. Where four teams in the procedural group selected a process instance level after selecting the instance document and never iterated this step, only two teams showed this clear pattern in the control group (team G and J). Team I attempted to select the level before the document had been selected; team K never conducted this step, and while team H selected a process instance level after selecting a document, they revisited this step several times, after engaging again and again in the relationships between the tables and in selecting activities.

Select process instance level → Select activities

Activity selection was done by all teams in both groups. This activity only makes sense when having decided which process instance to follow. Guided by the procedure, the teams of the procedural group follow this suggested approach. There were almost no repetitions in this step, suggesting that the students had the impression they had all the information they needed to take this decision. The control group addressed the activity selection more often (16 times in total). In four of the five teams, these executions were preceded by selecting the process instance level. In the remaining fifth team, the process instance was selected only at document level (just like one team in the procedural group only selected the process instance at document level). Since they had to iterate substantially more and repeat this step several times, it seems that the students did not feel sufficiently well prepared.

Select activities → Select attributes

In both groups, there appears to be a interdependency between the selection of activities and the selection of attributes. In the procedural group, this direct link was followed for all the instances of attribute selection. In the control group this was followed most of the times, and only once was selecting attributes attempted before the activity selection was even considered. Yet, the main pattern observed for the control group was a repeating back and forth movement between activity and attribute selection due to the high interaction of these elements, which were hard for the students to address separately.

Select attributes → Relate attributes to activities

The last step of relating attributes to activities was only executed by two teams of the control group. We will turn to the analysis of concepts to investigate whether the concepts that are related to this step might have been taken into account during other steps.

Concept analysis

The second category that was coded in the text was the discussed concepts. The concepts were coded per process step, and duplicates within a single process step were discarded. Per process step, a table was generated to report on the frequencies of all discussed concepts during this particular step, both in the procedural and in the control group (see appendix)[2]. For visualizing the differences between the groups, we used a reference "ideal" heatmap that ties concepts to procedure steps (Figure 5, the darker color indicate stronger ties).

The mean frequency of the times a concept was discussed within one process step was calculated for the procedural and the control group and depicted in Figure 6, comparably to the reference heatmap. Two observations can be made:

a) the procedural group addressed more concepts than the control group (more coloured boxes). In average, the teams in the procedural group addressed a total of 31.8 concepts, compared to an average of 25.4 concepts for the control group.
b) the control group showed less consistency in discussing certain concepts during certain process steps (a wider variety in colour intensity).

These observations suggest more profound and more streamlined discussions in the procedural group. Comparing the actual heatmaps depicted in Figure 6 (a) and (b) with the reference

[2]Beware that the frequency is the sum of a 0/1 indicator whether a concept was discussed during a particular process step or not.

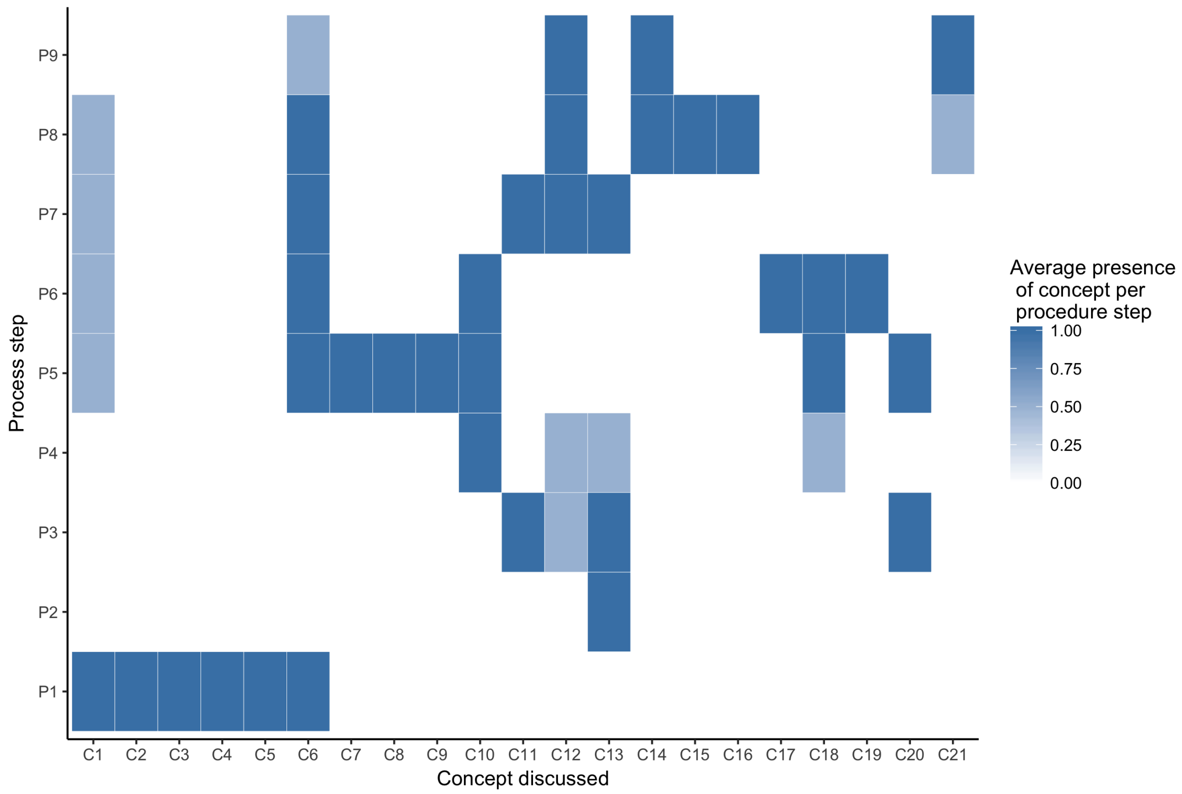


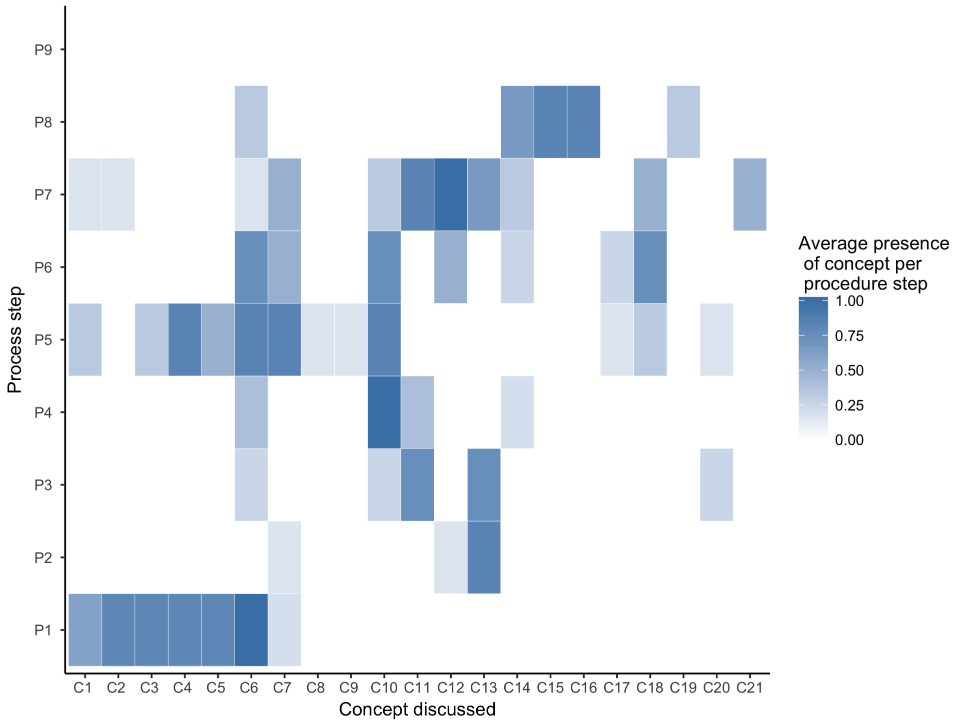
Figure 5.: Relation 'procedure steps - discussed concepts' in an optimal problem-solving process

one in Figure 5, it can be seen that for the procedural group (a), the heatmap has more shaded areas in common with Figure 5 than the control group (b).

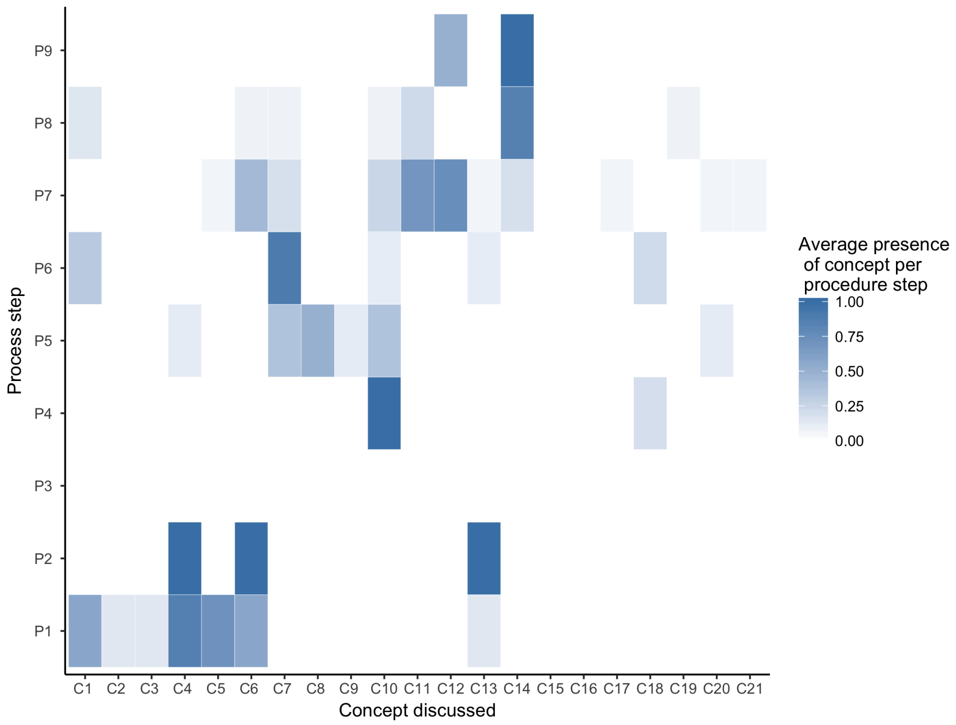
Another question that can be asked is whether all relevant concepts are addressed at all. For the procedural group, none of the concepts is completely missing: even though step 9 was not performed by any team, its relevant concepts were discussed. For the control group, case attributes (C15) and event attributes (C16) were not discussed at all (even by the teams who performed step 9), and additional concepts (C2, C3, C9, C17, C19, C20 and C21) were hardly discussed. This indicates poor effectiveness of the steps, failing to consider all the relevant concepts.

Effectiveness and efficiency: a white-box assessment

We now turn to consolidate the different aspects of the above process analysis into finer-grained indicators of effectiveness and efficiency of the problem-solving process. At a white-box view, based on the detailed analysis of the process, we consider the effectiveness of the process as the extent to which all the relevant concepts were attended to and all the required steps carried out. Considering that the procedure has nine steps, where a total of 21 concepts need to be addressed, we assess process effectiveness using two measures: (1) *Concept completeness* - the



(a) Procedural group



(b) Control group

Figure 6.: Average number of concepts discussed

number of distinct concepts discussed in the process divided by 21, and (2) *Step completeness* - the number of distinct steps performed in the process divided by 9. In our study the average concept completeness of the procedural and the control group was 0.73 and 0.57 respectively, and the average step completeness of the procedural and the control group was 0.82 and 0.67 respectively. This shows that the effectiveness of the process taken by the procedural group was higher than that of the control group, entailing more steps and addressing more concepts. It is not surprising that the resulting event logs of the procedural group were of higher quality.

Turning to the efficiency of the problem-solving process, we consider an efficient process as one where step repetitions and iterations are avoided. We therefore suggest a white-box indicator of *Process efficiency*: the number of distinct steps divided by the total number of steps in the process. Our results show that on average the procedural and control groups scored 0.90 and 0.53 on process efficiency respectively. In other words, the process taken by the procedural group was more efficient than that of the control group.

This result is not consistent with the black-box assessment of the process efficiency, measured by its time, for which no difference was observed between the groups. Yet, this inconsistency can be explained considering the differences in the process effectiveness, where the white box assessment was also reflected in the quality of the resulting log. The procedural group addressed more distinct steps and more concepts through a more efficient process. The resulting total time was similar to the total time of the inefficient process of the control group, where less concepts and less distinct steps were addressed, but were repeated and iterated through.

Process reflections

As a third dimension in analyzing the problem-solving process, explicit expressions of process reflections were coded in the verbal analysis. Table 5 reports the total frequencies of these expressions. The control group explicitly mentioned the difficulty of the assignment almost twice as much as the procedural group, and evaluated alternatives far more often (four times against one). These explicit process reflections confirm the observation of a less efficient problem-solving process; or at least it seems that the students experience it that way. The procedural group revisited the assignment considerably more (five times against one), possibly to ”orient” themselves in a mental process map. The subjects in the procedural group also gave considerably more thought to how to handle the assignment in a strategic manner. These findings suggest a more pronounced focus on the task in the procedural group, where the control group seemed to struggle with how to solve the task.

Table 5.: Frequency of process reflection categories

Process reflection category	Frequency Procedural	Frequency Control
Explicit difficulty expressions	7	13
Evaluating alternatives	1	4
Revisiting text	5	1
Strategic planning of the work	10	2

7. Discussion

Performance of log building task, linked with the problem-solving process

The findings of the study can be summarized and interpreted as follows: the procedural group in general addressed the required steps and the relevant concepts in an orderly and planned manner, focusing on the task at hand, and with a relatively high quality of the resulting product (event log). For the control group, we observe that the attention paid to the different tasks was not uniform. The more tangible tasks, whose output is concrete (selecting process instance document and level, selecting activities and attributes), received much more attention than more abstract tasks, like identifying process cornerstones and key tables, which were largely skipped. This kind of behaviour has been often observed as typical for novices (Feitelson, 2015; Huang and Burns, 2000). Since in our case the more abstract steps should set a firm basis for the "concrete" steps, the result for this group was a highly iterative thinking process, where many steps were repeatedly revisited. Also, more attention was given to discuss 'how' to solve the task than to solving the task.

In terms of the cognitive processes involved, two observations can be made. First, the procedural guidance, which broke the complex task into a sequence of distinct steps, spread the resulting intrinsic cognitive load along this sequence and reduced peaks that may cause overload. In contrast, without such support, the control group struggled with the high interactivity of the elements and iterated among the related decisions. Second, considering the mixture of the technical and business issues in the task, it can roughly be indicated that the steps that were emphasized by the control group were the steps related to the technical problem, while the business problem was only addressed by stating the goal. The (business-relevant) steps that were largely skipped by the control group were identifying process cornerstones and identifying key tables. As a result, it is likely that mental representations that should support problem-solving were incomplete, as can be seen in the partial list of concepts that were addressed by the control group. For this group, 'jumping' to the more concrete and technical tasks at the expense of developing a complete understanding and a mental representation of the business problem, resulted in poor effectiveness and efficiency of the problem-solving process afterwards. This was manifested in (a) many iterations over specific steps, mainly activity and attribute selection, (b) an incomplete set of concepts and sub-tasks (steps) addressed, (c) more difficulty expressed, and (d) lower quality of the final product. We note that the quality of the final product is largely related to its business value. Having paid low attention to the business aspects of the task, the members of the control group obtained logs of lower quality, considering the business problem at hand.

With regard to the efficiency of building the event log, when measured by the duration of handling the assignment, no differences were present between the two groups. Yet, detailed analysis of the process showed that time is not a sufficient indicator of efficiency in this case, as it does not reflect the amount of steps and concepts that were considered.

Limitations

Several limitation of the study should be discussed. First, the small scale of the study, which does not allow statistical analysis, might limit the generalizability of its findings. Nevertheless, the focus of the study is on a qualitative analysis of the log building process based on the verbal protocol, which yielded a detailed inside view of log building along three dimensions (process, concepts, and process reflections). With the qualitative dimensions that were studied, clear differences between the groups were apparent. We note that for studies that use verbal protocols, smaller sample sizes are common and acceptable (Nielsen, 1994). As compared to think-aloud studies, our study enjoyed the benefit of tracing the thoughts in a natural manner through the pair discussions.

Second, a question may be raised about the comparability of a guided and an undirected group for a quantitative measurement. We took measures to make the groups as equally knowledgeable of the subject as possible. The amount of time dedicated to training was equal for both groups. While for the procedural group this time was mostly dedicated to presenting the procedure, in the control group it was dedicated to introducing the concepts and their relations, as well as the required decisions, their relevant considerations and potential implications.

Third, the procedure used in the study is not the only procedure that can be used for log building, nor is it necessarily the best one. Rather, it is a practice-oriented procedure whose flow can be argued for logically. For the purpose of this study, other procedures could serve as well. Note that dependencies among activities should be kept by every appropriate procedure. However, alternative procedures may use different granularity levels, different compositions of subtasks into steps, and additional steps that may relate to specific contexts. In our analysis of the discussion text we categorized textual segments by the steps of our base procedure. Had we used a different procedure that keeps the dependencies among subtasks, the result would expectedly be similar. In summary, we consider our procedure as a (representative) procedure that can be followed. We do not aim to evaluate it against other log building approaches, but to use it as a basis for guiding the process and for analyzing the discussion text.

Last, a question may be raised about how realistic the setting of the study is, and how applicable such procedure is, considering the magnitude of ERP databases, which may include thousands of tables. (Brehm, Heinzl, and Markus, 2001) We note that although the database structure that is used in the assignment might seem a simplified academic example, this is not the case. Indeed, ERP databases are characterized by their large number of tables. Screening all these tables for relevant information is not feasible. For that specific reason, the procedure provides guidance in how to target the relevant tables for one specific process, starting from the business need, assisted by an IT expert who knows where to find which information. As a result, the full ERP database remains largely out of scope. We also note that the procedure assumes the accessibility of the database and does not include data extraction-related operations. This is a realistic assumption in any in-house managed system, and may be possible also under different conditions (e.g., cloud-based systems used as a service), although not necessarily. As mentioned, the procedure has been used in practice for projects in various organizations.

8. Conclusion

This paper addresses the complex task of building an event log. Similar data preparation tasks are increasingly being taught in rapidly emerging data science programs and used in industry. As data analysis becomes prevalent in all aspects of business and industry, the outcome of these tasks may bear substantial business implications.

The main contribution of the paper is in revealing the problem-solving process involved in the log creation task. The findings of the study highlight the various sub-tasks and their dependencies, and make the need for procedural support for such task evident. Procedural guidance support a systematic, step-by-step consideration of all the relevant concepts, leading to a more effective and efficient process. An interesting observation concerns the difference between the lower-level "technical" issues and the more abstract, business-related ones. While the technical issues were addressed by all students, the more abstract business issues were neglected when not specifically instructed as part of a procedure. This observation is consistent with existing literature on novices. Yet, it has practical implications, indicating a need to emphasize the business considerations in any procedure that may be used for guiding this task.

Implied from these findings is that other tasks of this kind can also benefit from a procedural guidance. Consequently, procedures need to be developed for these tasks, both for practice and for educational purposes. Future research can develop and evaluate such procedures. Such studies can then be accumulated to form a body of empirical research concerning data science and related education and promote the business value of data analysis endeavors.

References

Brehm, L., A. Heinzl, and M.L. Markus. 2001. "Tailoring ERP systems: a spectrum of choices and their implications." Maui, Jan. IEEE.

Calvanese, Diego, Marco Montali, Alifah Syamsiyah, and Wil M. P. van der Aalst. 2015. "Ontology-Driven Extraction of Event Logs from Relational Databases." In *Business Process Management Workshops*, Lecture Notes in Business Information Processing, Aug., 140–153. Springer, Cham.

Cohen, Jacob. 1960. "A coefficient of agreement for nominal scales." *Educational and psychological measurement* 20 (1): 37–46.

Cohen, Jacob. 1968. "Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit." *Psychological bulletin* 70 (4): 213.

de Murillas, Eduardo, Hajo Reijers, and Wil van der Aalst. 2018. "Connecting databases with process mining: a meta model and toolset." *Software and Systems Modeling* .

Feitelson, Dror G. 2015. "Using Students as Experimental Subjects in Software Engineering ResearchA Review and Discussion of the Evidence." *arXiv preprint arXiv:1512.08409* .

Gemino, Andrew, and Yair Wand. 2004. "A framework for empirical evaluation of conceptual modeling techniques." *Requirements Engineering* 9 (4): 248–260.

González López de Murillas, E., Hajo A. Reijers, and Wil M. P. van der Aalst. 2016. "Connecting Databases with Process Mining: A Meta Model and Toolset." In *Enterprise, Business-Process and Information Systems Modeling: 17th International Conference, BPMDS 2016, 21st International Conference, EMMSAD 2016, Held at CAiSE 2016, Ljubljana, Slovenia, June 13-14,2016 , Proceedings*, edited by Rainer Schmidt, Wided Gudria, Ilia Bider, and

Srgio Guerreiro, 231–249. Cham: Springer International Publishing.

González López de Murillas, Eduardo, Wil M. P. van der Aalst, and Hajo A. Reijers. 2015. "Process Mining on Databases: Unearthing Historical Data from Redo Logs." In *Business Process Management: 13th International Conference, BPM 2015, Innsbruck, Austria, August 31 – September 3, 2015, Proceedings*, edited by Reza Hamid Motahari-Nezhad, Jan Recker, and Matthias Weidlich, 367–385. Cham: Springer International Publishing.

Howe, Bill, Michael Franklin, Laura Haas, Tim Kraska, and Jeffrey Ullman. 2017. "Data Science Education: We're Missing the Boat, Again." In *Data Engineering (ICDE), 2017 IEEE 33rd International Conference on*, 1473–1474. IEEE.

Huang, I.-Lin, and James R. Burns. 2000. "A cognitive comparison of modeling behaviors between novice and expert information analysts." *AMCIS 2000 Proceedings* 45.

Hyde, J. S., and M. C. Linn. 1988. "Gender differences in verbal ability: A meta-analysis." *Psychological bulletin* 104 (1): 53.

Jans, Mieke, and Pnina Soffer. 2017. "From Relational Database to Event Log: Decisions with Quality Impact." In *Business Process Management Workshops: First International Workshop on Quality Data for Process Analytics (QD-PA), Barcelona, Spain, September 11, 2017. forthcoming*, edited by Moe Thandar Wynn, Marco Comuzzi, Minseok Song, and Lijie Wen, Berlin, Heidelberg. Springer.

Janssenswillen, Gert, Benoît Depaire, Marijke Swennen, Mieke Jans, and Koen Vanhoof. 2019. "bupaR: Enabling reproducible business process analysis." *Knowledge-Based Systems* 163: 927–930.

Kirschner, Paul A., John Sweller, and Richard E. Clark. 2006. "Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching." *Educational psychologist* 41 (2): 75–86.

Larkin, Jill H., and Herbert A. Simon. 1987. "Why a Diagram is (Sometimes) Worth Ten Thousand Words." *Cognitive Science* 11 (1): 65–100.

Lu, X., M. Q. L. Nagelkerke, D. van de Wiel, and D. Fahland. 2015. "Discovering interacting artifacts from ERP systems (extended version)." *BPM Reports* 1508.

Nielsen, Jakob. 1994. "Estimating the number of subjects needed for a thinking aloud test." *International journal of human-computer studies* 41 (3): 385–397.

Paas, Fred G.W.C., and Jeroen J.G. Van Merrinboer. 1994. "Instructional control of cognitive load in the training of complex cognitive tasks." *Educational psychology review* 6 (4): 351–371.

Raichelson, Lihi, Pnina Soffer, and Eric Verbeek. 2017. "Merging event logs: Combining granularity levels for process flow analysis." *Information Systems* 71: 211–227.

Russo, J. E., E. J. Johnson, and D. L. Stephens. 1989. "The validity of verbal protocols." *Memory & Cognition* 17 (1): 759–769.

Shaft, Teresa M., and Iris Vessey. 2006. "The role of cognitive fit in the relationship between software comprehension and modification." *Mis Quarterly* 29–55.

Soffer, P., Y. Wand, and M. Kaner. 2015. "Conceptualizing Routing Decisions in Business Processes: Theoretical Analysis and Empirical Testing." *Journal of the Association for Information Systems* 16 (5): 345–393.

Sweller, J. 2011. "Cognitive load theory." *Psychology of learning and motivation* 55: 37–76.

Sweller, John. 1988. "Cognitive Load During Problem Solving: Effects on Learning." *Cognitive Science* 12 (2): 257–285.

Tsoury, Arava, Pnina Soffer, and Iris Reinhartz-Berger. 2018. ""A Conceptual Framework for Supporting Deep Exploration of Business Process Behavior"." In *International Conference on Conceptual Modeling, ER'18*, .

van der Aalst, W. M. P. 2011. *Process mining: Discovery, conformance and enhancement of business processes.* Heidelberg: Springer.

Verbeek, H. M. W., Joos C. A. M. Buijs, Boudewijn F. Van Dongen, and Wil M. P. Van Der Aalst. 2011. "Xes, xesame, and prom 6." 60–75. Springer.