Identifying the Absence of Effective Internal Controls

An Alternative Approach for Internal Control Audits

Werner, Michael ; Gehrke, Nick

Document Version
Accepted author manuscript

Published in:
The Journal of Information Systems

DOI:
10.2308/isys-52112

Publication date:
2019

License
Unspecified

Citation for published version (APA):
Werner, M., & Gehrke, N. (2019). Identifying the Absence of Effective Internal Controls: An Alternative Approach for Internal Control Audits. *The Journal of Information Systems*, *33*(2), 205-222. https://doi.org/10.2308/isys-52112

Identifying the Absence of Effective Internal Controls: An Alternative Approach for Internal Control Audits

Michael Werner, and Nick Gehrke

Journal article (Accepted manuscript*)

Please cite this article as:
Werner, M., & Gehrke, N. (2019). Identifying the Absence of Effective Internal Controls: An Alternative Approach for Internal Control Audits. *The Journal of Information Systems*, *33*(2), 205-222. https://doi.org/10.2308/isys-52112

DOI: https://doi.org/10.2308/isys-52112

Identifying the Absence of Effective Internal Controls – An Alternative Approach for Internal Control Audits

Abstract

Auditors face new challenges when auditing internal controls due to the increasing integration of information systems for transaction processing and the growing amount of data. Traditional manual control testing procedures become inefficient or require highly specialized and scarce technical knowledge. This study presents audit procedures that follow a new approach. Instead of manually testing internal controls, automated procedures search for the absence of those controls. Process mining techniques are combined with advanced statistical analysis where process mining serves as a data analysis technique to create process models from the recorded transaction data. These are searched for critical data constellations in combination with an exploratory factor analysis to identify systematic deficiencies in the internal control system. The manual and time-intensive inspection of individual controls is replaced by automated audit procedures that cover the totality of recorded transactions. The study follows a design science approach and uses case study data for illustration.

Keywords: auditing, internal control system, monitoring, compliance, data analysis, journal entry data, process mining, enterprise resource planning systems, exploratory factor analysis, indicators

I. Introduction

Organizations use information systems such as Enterprise Resource Planning (ERP) systems to support and automate the processing of business transactions. These systems are also an important data source for financial reporting. The financial statements and consequently all financially relevant transactions recorded in the source systems during the reporting period are subject to statutory audits[1]. An important part of current risk-based audits is the testing of internal controls. The International Standards on Auditing (ISA)[2] demand that:

> "The auditor shall obtain an understanding of internal control relevant to the audit (…)" (IFAC 2012, sec. 12).

This requires:

> "(…) an understanding of the information system, including the related business processes, relevant to financial reporting (…)" (IFAC 2012, sec. 18).

During an internal control audit the auditor has to assess the entity's control activities that are a central part of its internal control system (COSO 1992). Control activities include all policies and procedures established and executed to provide reasonable assurance to management that the entity's objectives are achieved.

---

[1] The requirements are specified in national laws such as, for example, the Securities Act and the Securities Exchange Act (United States Congress 2012a, 2012b) in the USA or the Handelsgesetzbuch (Deutscher Bundestag 2013) in Germany.

[2] Similar requirements can be found in national accounting standards such as AS 12 (PCAOB 2010) or IDW PS 261 (IDW 2009)

To test internal controls an auditor has to understand the audited entity's business processes, how information systems support the processing of transactions and if control activities exist to ensure that the business processes are operated according to management objectives.

Auditors face new challenges while carrying out such audits due to the increasing automation of transaction processing, growing amounts of processed data and necessary specialized knowledge about the different source systems. It becomes questionable if traditional internal control audit procedures are still sufficient in such environments. This study deals with the question of how internal control audits can be conducted without the actual manual, time- and expert knowledge-intensive testing of relevant internal controls. It introduces alternative audit procedures that do not intend to test individual internal controls but to identify the absence of those controls and therefore deficiencies in the audited entity's internal control system. These procedures consider the totality of recorded data and substitute the manual inspection of internal controls by automated data analysis techniques. They also avoid the need for specific knowledge related to testing of internal controls that are embedded as automated controls in the specific source systems. The differences between traditional audit procedures and the new procedures described in this study are illustrated in Figure 1. The left column conceptually shows how internal control audits are traditionally conducted. The right column illustrates the new procedures.

Figure 1 Comparison of traditional and indicator-based audit procedures

The auditing of internal controls is carried out in different phases: (I) Information collection, (II) design effectiveness testing of internal controls, and (III) operating effectiveness testing of internal controls (IDW 2002).

Following traditional audit procedures the auditor first gains an understanding of the relevant business processes, related internal controls and information systems in an organization through interviewing individuals from the audited entity.[3] If a process documentation is not already available it is prepared by the auditor using basic modelling tools such as Microsoft Visio or in the form of simple diagrams or narratives using word processing software such as Microsoft Word. The auditor then identifies internal controls related to the inspected business processes. The controls are commonly documented in a control matrix.[4] The identified internal controls are finally assessed whether they are designed appropriately (design effectiveness testing) and have actually been carried out effectively (operating effectiveness testing) to achieve the desired control objectives. The testing of operating effectiveness is mainly done by manually inspecting a representative sample of control evidence in the form of physical documents, such as approval signatures on purchase orders. Automated controls, such as user access controls, are tested manually by inspecting relevant configuration parameters in the source system.[5] The result of this testing is usually documented in some kind of report which serves as a guidance to direct further substantive audit activities.

---

[3] These procedures are also commonly referred to as performing a walkthrough.

[4] An example of a typical business process model for a procurement process and the corresponding control matrix are shown in Appendix B.

[5] Some auditors use software like ACL (Audit Command Language) or proprietary software (PwC 2009, 2013) to support audits. This type of software can help the auditors to extract transaction data and control configuration parameters from the source systems, but the assessment and interpretation of the extracted data is mostly done manually.

It can be assumed that the efficiency and ultimately the effectiveness of traditional audit procedures decrease with the increase of automation, transaction volume and system heterogeneity, for which the inspection of automated controls requires specialized technical knowledge for each individual source system. This type of knowledge is volatile and often scarce in real audit scenarios.

The new approach is illustrated on the right-hand side of Figure 1. Instead of using potentially unreliable information from inquiries or few selected source documents the suggested alternative approach exploits the totality of data that has actually been recorded in the source ERP system.[6] The auditors hence receive more reliable and complete information about the real business processes and related internal controls.

The recorded data are extracted from the source system and analyzed by using process mining as a specific data analysis technique. Process mining algorithms produce business process models automatically by analyzing a given input. The input for process mining algorithms are event logs. These are essentially simple tables that include records for each activity carried out in the source system.[7] Process mining makes it possible for the auditor to automatically discover and graphically represent a process model on a given event log. Besides being able to represent pro-

---

[6] For a discussion of the benefits associated to using this type of data please compare Jans, Alles, and Vasarhelyi (2013).

[7] Each executed activity is recorded as an event in the event log. Each event is represented as a separate row in the event log table. Events are grouped into cases where each case represents a unique execution (process instance) of a particular business process. Please compare Gehrke and Werner (2013) for an introduction of process mining and the structure of event logs. A complete specification for event logs is available in C. W. Günther and Verbeek (2012).

cess models graphically, process mining tools store these models digitally in a structured format which makes it possible to analyze these models quantitatively. The mined models are analyzed for particular data constellations which indicate a missing internal control or weakness in an existing one. Each particular critical data constellation is described and represented by an indicator. These indicators are tagged to individual process models if a critical data constellation is found. The mined and tagged models are subsequently subject to an exploratory factor analysis (EFA). The EFA abstracts from the level of individually observed indicators and condenses the gained information. The interpretation of the EFA results by identifying typical indicator constellations leads to deficiency profiles. A deficiency profile reveals systematic deficiencies (weak or missing internal controls) in the entity's internal control system related to a specific business process. These profiles support auditors throughout the audit process to direct substantive audit procedures to those processes that exhibit a critical deficiency profile.[8]

In summary process mining algorithms produce business process models automatically replacing traditional manual process modelling techniques via interviews, observation, inspection and re-performance. Indicator tagging algorithms identify weak or missing internal controls and substitute the manual testing of internal controls via inspection, observation or re-performance.

The presented research follows a design science research (DSR) approach. It contributes to the body of knowledge by introducing a new type of audit procedures that uses reliable data and considers the totality of all processed transactions through the combination of process mining, indicator tagging and exploratory factor analysis (EFA). The study itself is exploratory in nature. It therefore does not aim to evaluate the presented results quantitatively. Instead, its applicability

[8] A summary of the specific terminology used in this study is listed in Appendix A.

is demonstrated by using a case study. The case study refers to a real audit project which was set up to improve a company's internal operations. Although the case study relates to an internal audit project, the described audit procedures are applicable to the auditing of internal controls in general regardless if it is a part of an external or internal audit.

The next section provides an overview of contemporary literature that is relevant to this study. Section III discusses the research methodology. The new audit procedures are described in detail in section IV in combination with the results derived from the case study. The manuscript closes with a conclusion and outlook to future research in section V.

II. Background

Of particular interest for this study is research related to process mining and auditing. Process mining is a Business Intelligence technique for analyzing large data sets. It is primarily used for discovering processes by producing process models in the form of graphical representations. Other application areas are conformance checking and process enhancement (van der Aalst et al. 2012). Research on process mining has matured in the past decade with the development of powerful heuristic (Weijters, van der Aalst, and de Medeiros 2006), fuzzy (Günther and van der Aalst 2007) and genetic (de Medeiros 2006) mining algorithms. The Process Mining Manifesto (van der Aalst et al. 2012) provides a comprehensive summary of contemporary challenges in process mining. An overview of basic and advanced concepts on process mining can be found in (van der Aalst 2016).

Process mining has already been successfully applied in the context of internal audits (Jans, Alles, and Vasarhelyi 2014, 2013; Jans et al. 2011; Jans, Alles, and Vasarhelyi 2010; Jans et al. 2008) and financial audits (Werner 2016; Werner and Gehrke 2015; Gehrke and Müller-Wickop 2010) for the automated discovery and modelling of process models.

Other publications deal with the use of information technology in the context of auditing. Software that supports auditors is called computer aided auditing tools (CAATs). ISA 330 mentions that computer-assisted audit techniques (CAATs) may be used by auditors to obtain additional evidence (IFAC 2010, sec. A16). Braun and Davis refer to this type of software as "tools and techniques employed to audit computer applications and to tools and techniques that extract and analyze data from computer applications" (Braun and Davis 2003, 726). This kind of software has been used in the auditing practice for several years. Prominent representatives of this kind of software are ACL (Audit Command Language) or IDEA (Interactive Data Extraction and Analysis). These tools support certain audit procedures and provide functionality for data queries, sample extractions and statistical analysis or specific purposes like user access or segregation of duties analysis. However, this type of software is still rarely used for audit purposes (Bierstaker, Janvrin, and Lowe 2014). The Big Four audit firms have developed proprietary data analytic tools such as the Automated Controls Evaluator (PwC 2009, 2013) and very recently a new generation of data analytic tools called Halo (PwC 2017) and Clara (KPMG 2017). Although these aforementioned tools help auditors to carry out advanced data analytics none provides process mining functionality. Scientific publications about their scope of application and effectiveness are currently outstanding.

III. Methodology

A DSR approach (Hevner et al. 2004; March and Smith 1995) was chosen for this study due to the objective of creating artifacts that are relevant for the application domain. DSR consists of the phases analysis, design, evaluation and diffusion (Österle et al. 2010). This study focused on the design and evaluation phase of a particular research cycle that aimed at developing a new

audit method to conduct internal control audits by incorporating process mining and advanced statistical methods.[9]

The primary research methods for the design of the presented solutions were method engineering (Brinkkemper 1996) and prototyping. A method in this context consists of different parts (method fragments) that can be combined and reused (Harmsen, Brinkkemper, and Oei 1994). A new method can be engineered by combining existing method fragments in a new manner or by developing completely new method fragments. The method fragments described by Gehrke and Müller-Wickop (2010) served as input for the development of indicator-based audit procedures described in this study. Gehrke and Müller-Wickop (2010) introduced how data from recorded journal entries can be used to create an event log where recorded events are matched to cases. This technique served as a foundation for the implementation of a process mining algorithm that produces models for individual process executions.[10] The mining algorithm was combined with a

---

[9] According to Gregor's and Hevner's classification scheme (2013) the study's main DSR contributions are the new audit method, related constructs (indicators and deficiency profiles) as well as instantiations of the related designed artifacts (algorithm, prototype, tagged process models and deficiency profiles from the case study).

[10] A general purpose process mining algorithm, such as the Fuzzy Miner (Günther and van der Aalst 2007), usually creates a process model for a set of similar process executions (process instances). This means that a generic process mining tool creates a single model for a given event log. But in reality each individual execution can differ from the mined model which serves as an abstraction of the represented process executions. In the case of the process model example shown in Appendix B there might exist, for example, process executions without recorded activities for received goods because they relate to ordered services. An auditor is interested in each individual execution to be able to follow the audit trail and in particular in deviations from standard procedures. The algorithm presented by Gehrke and Müller-Wickop (2010) was chosen because it is able to produce a model (process instance model) for each individual execution of a business process.

novel indicator tagging method and exploratory factor analysis. This study relied on EFA due to the nature of the observed phenomenon. EFA is a statistical method that allows to discover hidden structures (factors) from observable phenomena (variables).[11] EFA was useful in this context as an exploratory statistical method because the observed population was comparatively large and the systematic deficiencies were a priori unknown. It served as a solution to discover the systematic deficiencies in an internal control system which were actually the cause for the observed critical data constellations. The results of the EFA were finally transformed into deficiency profiles by interpreting the constellations of indicators represented in each factor.

The different methods were implemented in a software prototype. Prototyping (Naumann and Jenkins 1982) served as a research method in this study to develop a software artifact which was used to evaluate the designed methods. The prototype was developed in several research cycles and embedded in the development of commercial audit software. It consisted of an extraction and a mining module. The source data from a SAP ERP systems served as input for the prototype that produced process models as an output.[12] These models were subsequently analyzed quantitatively.

The evaluation was carried out by referring to a case study project. It was tested if the developed methods and implemented prototype could successfully be applied in a real world scenario. A German publicly listed company operating in the manufacturing industry with production and

---

[11] As shown in Figure 1 indicator-based audit procedures start with the extraction of the relevant source data into an event log. This serves as an input for the process mining algorithm which produces different models. These are then searched for critical data constellations and tagged with indicators if such constellations are found. The results from the applications in practice have shown that the amount of mined models is too high for manual inspection.

[12] Please compare Table 1 for addition information about the provided source data.

distribution sites in several countries and a global sales volume of several billion euros provided the necessary data. The project was initiated by the company's internal audit function and carried out in cooperation with a small audit firm specialized on novel data analysis techniques. The participating company's aim was to improve their internal processes and to identify weaknesses in its internal operations.

The new analysis methods were developed by the authors of this study and implemented as algorithms in the software prototype in cooperation with the involved audit company. Critical data constellations and indicators were defined in a common exercise with the involvement of all project partners by relying on publically available information (ISACA 2015) and professional judgement. The software was executed and the results analyzed by the authors and members of the partner audit company. The software automatically extracted and analyzed the totality of all recorded transactions, created corresponding process models, numbered and tagged indicators to the mined models and finally created the input for the EFA. The EFA was carried out by the authors using SPSS. The deficiency profiles were interpreted in a common exercise done by members of the case study company, the audit company and the authors. The research project stopped with the identification of the deficiency profiles. The subsequent actions and audit activities undertaken by the case study company were not investigated.

IV. Indicator-based Audit Procedures

This section presents the study's main contribution by describing new audit procedures as an alternative to traditional internal control testing procedures. They are called indicator-based as indicators form an essential concept of the new procedures. An indicator represents and describes a critical data constellation in the source data which indicates a missing or weak internal control.

The section follows the structure as illustrated in Figure 1. Whereas a conceptual overview of the new procedures has been provided in the introduction the following subsections focus on detailed technical information of how the different steps for (a) data extraction and automated process discovery, (b) indicator tagging and (c) EFA with the identification of deficiency profiles can be carried out.

Data Extraction and Automated Process Discovery

Before mining processes it is necessary to extract relevant data from the source systems. This study relies on a process mining algorithm which accepts journal entry data as input. The Financial Process Mining (FPM) algorithm exploits the structure of recorded journal entries (Gehrke and Müller-Wickop 2010). Whenever a financially relevant activity is recorded in an ERP system this creates entries in the financial accounts. The entries created by one activity clear open entry items created by another activity that belong to the same process instance. This relationship is illustrated in Figure 2.

Figure 2 Example of a simple process model

It shows a simple example of a procurement process that consists of three activities. Each rectangle represents a separate activity in the process. The circles denote the start and end of the process. The involved financial accounts and posted entry items are shown at the top. The arrows between the rectangles denote the control flow and structure of the process. The arrows between the rectangles and the accounts illustrate the relationship between entry items in the different accounts and the activities that created the corresponding entry items. The example process starts with the recording of received goods *(A)*. The corresponding activity creates a journal entry

which is recorded as an event in the source system. The journal entry consists of two entry items, a debit posting on the *Raw Materials* account *(a1)* and a credit posting on the clearing account *Goods Received / Invoices Received (GR/IR) (a2)*. The next activity *(B)* clears the open entry *(a2)* posted by activity *(A)* with a debit posting on the *GR/IR* account *(b1)* and a corresponding credit posting on the *Trade Payables* account *(b2)*. The process ends with the payment of the received invoice *(C)* that creates a debit posting on the *Trade Payables (c1)*. This clears the open item on that account *(b2)* posted by activity *(B)* and a credit posting on the *Bank Account (c2)*.

The relationship between the accounting entries makes it possible to discover the original logical structure of the underlying business process. The FPM algorithm starts with an arbitrarily activity to mine a process model. Referring to the example shown in Figure 2 the algorithm starts, for example, with the journal entry created by activity *(B)*. It searches if items posted by activity *(B)* were cleared by items posted by other activities. The algorithm identifies that *(b2)* was cleared by *(c1)*, and therefore infers that *(B)* happened before *(C)*. The search is repeated for all newly identified events. The algorithm analyses if there are any items posted by *(C)* that were cleared by items posted by other activities. Activity *(C)* just created one further posting on the *Bank Account (c2)* which was not cleared by any other item. The algorithm therefore stops at this point with the identified activities $\{(B), (C)\}$ and inferred control flow $\{(B) \rightarrow (C)\}$. It then proceeds with a backward search. All identified activities are analyzed whether they cleared entry items of activities that have not already been identified. In the previous example the set of identified activities is $\{(B), (C)\}$. The algorithm analyses the entry items belonging to these activities and identifies that item *(b1)* cleared item *(a2)* posted by activity *(A)* where $(A) \notin \{(B), (C)\}$. Activity *(A)* is then added to the set of identified activities $\{(B), (C), (A)\}$ with the control flow now being $\{(A) \rightarrow (B) \rightarrow (C)\}$. For the example shown in Figure 2 the algorithm terminates at this

point. In general all activities and created journal entry items are searched in repeated forward and backward searches until all activities and journal entry items are found that belong to the same execution of the business process. The search is repeated for all recorded activities until all events in the event log are matched to cases.[13]

The FPM creates models, which are semantically identical to the type of models shown in Figure 2, for each identified execution of a particular business process. Table 1 summarizes the case study data which served as input for the FPM algorithm. The mining algorithm discovered 25,051 process executions (process instances) for this data set with a corresponding model for each individual execution.

Table 1 Case study dataset overview

Analyzing Mined Models and Indicators

Once the models are mined these are searched for data constellations that indicate whether a deficiency in internal control is likely. Each explicitly defined critical data constellation is represented and described by an indicator. If a critical data constellation is discovered in a mined model the model is tagged with the corresponding indicator. An example of a simple but critical data constellation is if a transaction has been executed with administrator access rights. Administrative access rights should be restricted for the maintenance of a system and not for processing business transactions because such access rights allow the respective user to override other con-

[13] A more extensive description of the algorithm is available in (Gehrke and Müller-Wickop 2010). The original FPM algorithm was extended for the purpose of this study to be able to mine activities and sub-processes that do not affect financial accounts directly.

trols. A mined model can be analyzed accordingly. For the example shown in Figure 2 each of the recorded activities {*(A), (B), (C)*} can be analyzed to find out who executed the respective activities and if this user had administrative access rights. In a SAP ERP system this would be those activities carried out by a user with SAP_ALL access rights, for example.

The existence of indicators in mined models is a sign for a deficiency in the audited entity's internal control system. Other examples for indicators are: invoices were paid too late leading to discount losses, all activities belonging to a single process instance were executed by a single user, or an invoice was posted prior to the corresponding purchase order.

A challenge in real world projects is the identification of data constellations which meet the criteria of being an indicator, because data constellations which indicate a deficiency in internal control can differ from organization to organization. Some auditing firms maintain proprietary rule sets for identifying specific deficiencies in internal control. But even without such proprietary knowledge it is possible to identify and create a list of applicable indicators. The Information Systems Audit and Controls Association (ISACA) has released a document listing critical segregation of duty (SoD) conflicts (ISACA 2015). Such publically available documents can be used to identify which data constellations are critical from an internal control perspective. The ISACA guideline, for example, identifies the activities '*maintain bank master data'* and '*process accounts payable payments'* as a critical combination. The corresponding indicator descriptor could be: '*Bank master data are maintained and accounts payable payments processed by the same user'*.

As part of the case study project 100 indicators were identified. A limit was set to a maximum of 100 indicators in order to focus on those indicators that were considered most important.[14] The participating auditing firm and the internal auditors of the case study company agreed on the final set of indicators that were selected based on their professional experience and the characteristics of the audited company. Table 2 shows an extract of 19 selected indicators that are discussed in this study.[15]

Table 2 Exemplary indicators used in the study

The indicators listed in Table 2 were linked to a process instance model if the described data constellations were found. They were linked to a complete process instance or individual activities within the instance, depending on the type of an identified data constellation. The following paragraphs describe how the linkage was established:

- Indicator #1 (Transaction is executed with administrator access rights): the indicator was tagged to those activities in a process instance model processed by any user with administrative user rights.
- Indicator #2 (Complete process instance is executed by a single user): the indicator was tagged to a process instance model if all identified activities had been processed by the same user.
- Indicators #3 to #19 (Critical data constellations due to missing segregation of duties): the respective indicator was linked to the pair of activities in the process instance model

[14] The list and description of all indicators considered by the involved auditors is available at www.zapliance.com.

[15] Except for the indicators 1 and 2, all indicators refer to risks resulting from a lack of segregation of duties.

if the activities were in conflict with each other according to the segregation of duty rules described by the respective indicator.

The results for the case study data are illustrated in Figure 3 and Figure 4. Indicators were tagged to 9,218 (36.80%) out of 25,051 mined process instance models. For the interpretation of both diagrams it should be taken into account that the multiple occurrence of one indicator in a single process instance model was not captured separately (either the related data constellation was identified at least once or not at all). Figure 3 shows the distribution of indicators over the number of tagged models. The ordinate is displayed in logarithmic scale. The diagram shows that two indicators (#13 and #17) were tagged very rarely in less than 0.05 percent of the tagged models. Seven indicators (#5, #6, #7, #10, #14, #15 and #16) were tagged rarely (between 0.1 and 1.0 percent). Eight indicators (#2, #3, #8, #9, #11, #12, #18 and #19) were tagged frequently (between 1.1 and 10.0 percent). Indicators #1 and #4 were tagged very frequently with 78.5 and 25.5 percent respectively. This means that 78.5 percent of the overall 9,218 tagged models (28.9 percent of all mined models) showed the critical data constellation that at least one transaction was executed with administrative access rights. This was an alarming high number proposing a poor access right structure for privileged user accounts. The high occurrence of indicator #4 also suggested a fundamental segregation of duties problem in the procurement process and the processing of incoming invoices.

Figure 3 Distribution of indicators tagged to process instance models

Figure 4 displays the distribution of indicators tagged per model over the number of tagged models. It shows that for 72.7 percent of the tagged models only one type of indicator was tagged to each model. In 20.8 percent of the cases a combination of at least two tagged indicators were

observed. A combination of more than two indicators was only identified rarely in the remaining 6.4 percent of tagged models. The highest number of different indicators tagged to the same process instance model was five. The number of different indicators per model dropped exponentially. This means that the likelihood of different deficiencies in internal control strongly decreased with each additionally observed deficiency represented by a tagged indicator. These observations were reasonable. Most of the different indicators related to different business processes (e.g. sales process, procurement process etc.) and it was not expected to observe all of them in a single model. The decreasing number of observed multiple indicator combinations in the mined models meant that one deficiency did not attract the occurrence of further deficiencies. If a control weakness did exist or if a control was missing this was already sufficient for a compliance violation to occur.

Figure 4 Distribution of indicators per model over the number of tagged models

Exploratory Factor Analysis for Identifying Deficiency Profiles

The analysis results showed that a large amount of models were tagged with indicators. Each tagged model indicated a potential compliance violation. From a practical point of view the number of suspicious models were far too high to inspect each individually. It was therefore necessary to provide analysis results on a higher abstraction level. This was achieved by clustering instances tagged with similar indicators through statistical analysis.

A set of indicators which refers to a significant amount of process instances is called a *deficiency profile* in this study. These profiles illustrate the inherent deficiencies in internal control for a particular process within an organization. A deficiency profile is meant to express which

systematic combinations of indicators, that actually indicate missing or weak internal controls, are hidden in the company's processes.

Once indicators had been linked to process instance models these models could serve as the input for statistical analyses. This study relied on EFA to identify deficiency profiles. The variables in our case were the observed indicators, the unobserved factors were the deficiency profiles. Each indicator as defined in this study is a theoretical concept. The same indicator could be tagged to many instance models and activities belonging to these models. In order to be able to differentiate the individual occurrences of an indicator among a set of process instance models and activities each occurrence of an indicator was called an *indicator instance*. It had a unique ID and was attached to a single instance model or an activity within such a model.

Table 3 shows three different process instance models. The numbers shown within each activity symbol were unique and referred to the recorded activities from the event log. The round symbols represented instances of the indicators A, B and C. The indicator instances were attached to the activities. The numbers below the indicator instance symbols are the unique indicator instance IDs. Several different indicator instances were tagged to a single process instance. The two right columns in Table 3 illustrate the different indicator combinations and their respective instance IDs that were observed in the different process instances.

Table 3 Processes instance models and identified indicator combinations

The identified indicator combinations were listed for each mined process instance in a cross tabulation as shown in Table 4. If an indicator was tagged to a process instance, this was recorded as a *'1'* in the respective field for that particular indicator and process instance. If an indicator was not tagged to the particular process instance, the data entry was *'0'*. Table 4 just refers to the

example instances shown in Table 3. For the complete case study data the cross tabulation produced by the software prototype included 19 columns, one for each indicator, and 9,218 rows, one for each mined and tagged process instance model.

Table 4 Derived data matrix as input for the EFA

The cross tabulation with the structure shown in Table 4 served as the input for the EFA. EFA takes advantage of the correlation of the observed variables (here the indicators) to draw conclusions on underlying, latent (not directly observable) variables that are called factors (Backhaus et al. 2016). The columns represent the observed variables and the rows the individual observations. If, as shown in the example in Table 4, the indicators A and B appear frequently (but not necessarily together), the EFA reveals a factor that contains strong proportions of indicator A and B. The extracted factors are orthogonal to each other in a geometrical sense. This means that they are clearly separated from each other. Consequently well-defined deficiency profiles were detected by using this method.

The statistical calculations were performed by the authors of this study using SPSS. Factors were extracted by using principal component analysis (PCA). Factor interpretation was facilitated by rotation using the varimax method (Kaiser 1958). The result was a rotated factor matrix. Related literature presents different criteria for determining the number of factors to be extracted. This study relied on the traditional Kaiser-Guttman criterion (Guttman 1954; Kaiser and Dickman 1959). Factors were extracted until the eigenvalue of the factor fell below one and hence the factor did not explain the variance of the data record to a sufficient extent anymore.

Table 5 Factor contribution for explaining the observed variance

Table 5 shows the contribution of the factors for explaining the variance of the observed data. 8 factors were extracted for 19 initially observed variables. Factor 9 only explained 5.18% of the variance, which is less than 100/19, and consequently all factors following factor 9 on were discarded. The remaining 8 factors explained approximately 65.5% of the total variance. Approximately 34.5% of the variance remained unexplained. As a result of the EFA the initially observed 19 dimensions (each dimension representing one variable or in this context one indicator) were reduced to 8 dimensions.

Table 6 Extracted factors

Table 6 shows the extracted factors and their loading for each individual variable. The different factors were now interpreted column by column by the authors and the auditors from the project partner companies. In Table 6 those variables are highlighted with different colors and patterns that were especially important for the respective factor. The interpretation of factors was facilitated by referring to different value ranges. The value ranges for interpreting the factor loading of the case study data are shown in Table 7. Each value range was assigned a name for the particular constellation. The thresholds for the value ranges were not scientifically derivable and consequently subject to professional judgement. In the case study project the thresholds were identified as a result of a consensus finding process between the involved researchers and the auditors from the project partner companies.

Table 7 Threshold values for the interpretation of factors and deficiency profiles

The values in Table 6 are colored according to the applicable value ranges shown in Table 7. The occurrences of the constellations for each factor are summarized in Table 8.

Table 8 Deficiency profiles

The names of the constellations in Table 7 refer to the characteristics of the indicators within a factor. The constellation names intend to express the role of significant indicators for a particular deficiency profile. The constellation *'strongly dominates'* indicated that a factor was dominated by a single indicator. This meant that the indicator was highly likely to be the primary reason for the existence of the identified deficiency profile and hence the underlying deficiency in internal control. From an audit perspective it implied that if it was possible to remediate the underlying control weakness or to implement a corresponding missing control for that particular indicator the related systematic deficiency in internal control would be remediated. The constellation *'weakly dominates'* meant that an indicator was important for a factor but did not strongly dominate it. The constellations *'collaboratively strongly dominate'* and *'collaboratively weakly dominate'* related to the loading of multiple indicators. *'collaboratively strongly dominate'* meant that two or more strong indicators dominated a factor. The indicators worked together, which means that they appeared in combination and potentially reinforced or required each other. *'collaboratively weakly dominate'* referred to the constellation when important but not dominating factors appeared in combination. The constellation *'antipode'* characterized the situation where an indicator was frequently absent whereas other significant ones were present.

The meaning of the different indicator constellations and the interpretation of an individual factor are illustrated for the different observed constellations by referring to factors 1 to 3 as examples:

- **Factor 1** was classified as being *'strongly dominated'* by indicator #17. This meant that indicator #17 was dominant in this factor ('*A sales invoice is changed and payments are changed for it by the same user*', loading 0.876). The activities that were related to this

indicator allowed a user to change a sales invoice and to change the received payment for it. If these activities are executed by the same user such a user can decrease the sales amount on the invoice and the amount of the received payment. This circumvents other internal controls such as a prior approval of the invoice and can be used to divert payments afterwards.

Due to the further indicator constellation of *'collaboratively weakly dominates'*, this phenomenon sometimes coincided with indicator #16 *('An accounts receivable subsidiary ledger account is adjusted using payment runs and this is then concealed with general ledger entries'*, loading 0.688). For example, an outgoing payment for a customer can be posted to an accounts receivable account in the sales ledger and in a second step the appropriate open item of the customer is credited against expenses with a general ledger entry.

Factor 1 continued to load as *'collaboratively weakly dominated'* by indicator #13 *('The customer master data is changed and an unauthorized invoice entered by the same user',* loading 0.713). This allows a user to change the customer master data including bank details and to record unauthorized invoices. The combination of the different indicators meant that fraudulent activities for diverting assets in the form of cash could have been executed and concealed completely by a single user.

- **Factor 2** was interesting because it contained a strongly negative variable ('*antipode*'). Indicator #1 *'A transaction is executed with administrator access rights*' negatively dominated this factor (loading -0.855). This meant if other indicators were present (in this case the indicator #4 *'An incoming invoice is created and payment for it initiated by the same user'* and indicator # 3 *'An accounts payable subsidiary ledger account is adjusted*

using the transaction to record an incoming invoice and then concealed via general ledger entries by the same user'), indicator #1 was frequently absent, i.e. behaved in a somewhat repelling way. In this particular constellation it could be assumed that users with excessive access rights were not systematically responsible for the described deficiencies in this factor.

- **Factor 3** was characterized by two indicators that *'collaboratively strongly dominate'* it. This meant that two different indicators accumulated in this factor. This was the case for the two indicators #10 *('A sales document is created and a billing document generated for it by the same user'*, loading 0.875) and #18 *('The shipment of goods is concealed by maintaining a fictitious sales document'*, loading 0.896). Hence, for factor #3 the data constellations of *'Process sales order'* and *'Prepare outgoing invoice'* on the one hand, and *'Process shipment'* and *'Process sales order'* on the other, were detected in combination frequently. This deficiency profile indicated deficiencies in internal control in the sales process because of accumulated segregation of duties violations as the three critical sales activities *'process sales order', 'process shipment'* and *'prepare outgoing invoice'* were frequently not segregated.

In summary Table 6 and Table 8 describe the quantitative results generated by exposing the previously described indicator-based audit procedures to the case study data. The case study demonstrates how factors were identified and interpreted to develop deficiency profiles for an exemplary set of indicators. The identification of critical data constellations and defining indicators by the researchers and auditors took several hours. The same was the case for organizing and executing the automated data extraction from the source system. The necessary computation time for the different steps of mining models, tagging indicators, and EFA each just took several seconds

on a modern laptop computer. The interpretation of deficiency profiles by the researchers and auditors was done during a workshop and several subsequent discussions.

The question now arises how an auditor can make use of the gained information and how it can be integrated into the overall audit process. As shown on the left side of Figure 1 traditional audit procedures usually result in findings or recommendations in the form of a report that provides information about the effectiveness of the audited entity's internal control system. It is commonly used to determine the nature, timing and extent of subsequent substantive audit procedures. If the traditional audit procedures are substituted by indicator-based audit procedures the output is a set of deficiency profiles that unravel systematic deficiencies in internal control related to an entities' business processes. Each deficiency profile relates to a set of process instances. If the set of those instances is too large for inspecting all individual instances a sample can be selected for targeted tests of details.[16] The link between deficiency profiles via tagged indicators to the mined processes instances serves as a target criterion for the sampling. The selected process instances should relate strongly to the respective examined factor. The sampled process instances can be inspected to determine if the observed deficiency in internal control has actually led to a compliance violation, and if so, what the effect of such a violation is. The findings of these substantive audit procedures can be combined with the overall interpretation of the identified deficiency profiles and the percentage of process instances which did not show any deficiencies. Traditional process audit procedures do not provide such information. This type of

[16] This sampling differs in scope from traditional statistical or non-statistical sampling. Traditional sampling draws samples from the whole population of transactions. Here, the whole population is first analyzed completely and samples are just drawn from the set of identified suspicious transactions which is a significant smaller subset of the overall population.

information is novel and helps the auditor to direct the audit effort to those business transactions which exhibit a high risk for error or manipulation due to existing deficiencies in internal control.

The case study company was not aware of the identified systematic deficiencies and used the produced audit results as a starting point for further investigation into those business processes where significant weaknesses were identified. The actual identification and remediation of identified weak or missing controls was not part of the research project which was set up as an exploratory study to investigate the applicability of the presented approach.

Although this study did not aim for providing empirical evidence on potential audit efficiency gains it showed that the effort to set up the new audit procedures is comparatively low requiring just a couple of hours. In addition, once indicators have been identified the same set can be used for subsequent analyses and other companies that exhibit similar characteristics. The key benefits are the consideration of the totality of all processed transactions, the automated analysis of the source data and the aggregation of the provided information which makes it accessible to the auditors. It can be assumed that this combination greatly enhances not just the effectiveness of internal control audits but also their efficiency.

V. Conclusion and Outlook

The auditing of internal controls is important for organizations to ensure that business operations comply with internal and external requirements. Auditors face new challenges as business transactions are processed by increasingly diverse information systems and the amount of recorded financial data grows extremely. The data recorded by these systems do not only form the basis for internal and external financial reporting but they also provide a valuable source for audit purposes. This study introduces a novel approach for auditing internal controls which exploits the internal data that are available in the companies' ERP systems. The core idea of the new ap-

proach is not to test internal controls directly but to instead scan the totality of recorded transactions for data constellations which indicate weak or missing internal controls. Such data constellations are represented by indicators. They disclose if an internal control was ineffective or non-existent for a particular process execution. The existence of an indicator means that the chance of compliance violations in such a process is high.

In order to identify these specific data constellation it is first necessary to analyze the source data by using process mining techniques. These produce models of the underlying process executions. In a second step the models are assessed whether the indicators relate to the mined models. If this is the case they are linked to each other. The combination of mined models and linked indicators serves as the input for an exploratory factor analysis. The EFA makes it possible to abstract from the identification of indicators for particular process executions into deficiency profiles which provide an overview of the systematic deficiencies in the audited entity's internal control system. The profiles can be used by an auditor to create targeted samples on processes that exhibit a critical deficiency profile. By applying substantive audit procedures to suspicious process executions the auditor can assess if compliance violations have actually occurred and what their impact is.

In contrast to traditional internal control audit procedures the analysis of the source data itself takes places in a completely automated manner. Process mining algorithms produce business process models automatically making traditional manual process modelling techniques via interviews, observation, inspection and re-performance obsolete. Indicator tagging algorithms identify weak or missing internal controls which substitutes the manual testing of controls via inspection, observation or re-performance. Major benefits of the proposed new audit procedures are that they do not require a priori knowledge of the implemented internal controls and cover the

totality of recorded transactions. The type of data used is very difficult for the users of the source system to manipulate and therefore a valuable resource of information which is not accessible to the auditor by using traditional audit procedures. The deficiency profiles are identified by using reliable statistical methods and provide relevant information to the auditor in a very compact form.

Although the data analysis is carried out automatically by software programs certain human interaction and professional judgement is still necessary. This is particularly the case for the identification of relevant indicators, the interpretation of deficiency profiles and potentially for any subsequent substantive audit procedures. Many different indicator definitions exist in practice. Currently a scientific discussion about the completeness and the significance of indicators is missing and has to be assessed by professional judgement. Research in this area is likely to be highly relevant.

The new audit procedures were applied in a real world audit project to assess whether they were applicable in practice. The case study results showed that the new procedures were successful in discovering deficiencies in internal control that had not been identified by traditional audit procedures. The analyzed data set was derived from a SAP ERP system of a manufacturing company. The chosen mining algorithm took advantage of the inherent structure of journal entries which is, in principle, independent from individual ERP implementations. However, without the actual application to data sets from other types of ERP systems it remains unclear if the suggested procedures actually work for such systems as well. The presented findings derive from a single ERP system. Especially larger organizations use multiple information systems to process business transactions. Research on how the presented approach can work across several source systems is currently outstanding.

References

Aalst, Wil M. P. van der. 2016. *Process Mining: Data Science in Action*. Berlin, Heidelberg: Springer Berlin Heidelberg.

Aalst, Wil M. P. van der, A. Andriansyah, Alves K. de Medeiros, F. Arcieri, T. Baier, T. Blickle, J. C. Bose, P. van den Brand, R. Brandtjen, and J. Buijs. 2012. "Process Mining Manifesto." In *BPM 2011 Workshops Proceedings*, 169–194.

Backhaus, Klaus, Bernd Erichson, Wulff Plinke, and Rolf Weiber. 2016. *Multivariate Analysemethoden: Eine Anwendungsorientierte Einführung*. 14., Überarbeitete und aktualisierte Auflage. Lehrbuch. Berlin□;Heidelberg: Springer Gabler.

Bierstaker, James, Diane Janvrin, and D. Jordan Lowe. 2014. "What Factors Influence Auditors' Use of Computer-Assisted Audit Techniques." *Advances in Accounting* 30 (1): 67–74.

Braun, Robert L., and Harold E. Davis. 2003. "Computer-Assisted Audit Tools and Techniques: Analysis and Perspectives." *Managerial Auditing Journal* 18 (9): 725–731.

Brinkkemper, Sjaak. 1996. "Method Engineering: Engineering of Information Systems Development Methods and Tools." *Information and Software Technology* 38 (4): 275–280.

Chuprunov, Maxim. 2012. *Handbuch SAP-Revision: internes Kontrollsystem und GRC*. Bonn: Galileo Press.

COSO. 1992. "Internal Control - Integrated Framework.Pdf."

Deutscher Bundestag. 2013. *Handelsgesetzbuch*.

fluxicon. 2016. "Process Mining and Process Analysis - Fluxicon." www.fluxicon.com.

Gehrke, Nick. 2010. "The ERP AuditLab - A Prototypical Framework for Evaluating Enterprise Resource Planning System Assurance." In *Proceedings of the 43th Hawaii International Conference on System Sciences*, 1–9. Kauai: IEEE.

Gehrke, Nick, and Niels Müller-Wickop. 2010. "Basic Principles of Financial Process Mining A Journey through Financial Data in Accounting Information Systems." In *Proceedings of the 16th Americas Conference on Information Systems*. Lima, Peru.

Gehrke, Nick, and Michael Werner. 2013. "Process Mining." *Wisu - Das Wirtschaftsstudium*, no. 7: 934–43.

Gregor, Shirley, and Alan R. Hevner. 2013. "Positioning and Presenting Design Science Research for Maximum Impact." *MIS Quarterly* 37 (2): 337–55.

Günther, C., and Wil M.P. van der Aalst. 2007. "Fuzzy Mining – Adaptive Process Simplification Based on Multi-Perspective Metrics." *Business Process Management*, 328–343.

Günther, Christian W., and Eric HMW Verbeek. 2012. "XES Standard Definition." Eindhoven: Eindhoven University of Technology.

Guttman, Louis. 1954. "Some Necessary Conditions for Common-Factor Analysis." *Psychometrika* 19 (2): 149–61.

Harmsen, Anton Frank, J. N. Brinkkemper, and H. Oei. 1994. *Situational Method Engineering for Information System Project Approaches*. University of Twente, Department of Computer Science.

Hevner, A.R., S.T. March, J. Park, and S. Ram. 2004. "Design Science in Information Systems Research." *MIS Quarterly* 28 (1): 75–105.

IDW. 2002. *IDW AuS 330 The Audit of Financial Statements in an Information Technology Environment*.

———. 2009. *IDW PS 261 Feststellung Und Beurteilung von Fehlerrisiken Und Reaktionen Des Abschlussprüfers Auf Die Beurteilten Fehlerrisiken*.

IFAC. 2009. *ISA 520 Analytical Procedures*.

———. 2010a. *ISA 330 The Auditor's Responses to Assessed Risks*.

———. 2010b. *ISA 330 The Auditor's Responses to Assessed Risks*.

———. 2012. *ISA 315 (Revised), Identifying and Assessing the Risks of Material Misstatement through Understanding the Entity and Its Environment*.

ISACA. 2015. "Best Practices to Resolve Segregation of Duties Conflicts in Any ERP Environment."

Jans, Mieke, M. Alles, and M. Vasarhelyi. 2010. "Process Mining of Event Logs in Auditing: Opportunities and Challenges." Working paper. Hasselt University. Belgium.

Jans, Mieke, Michael G. Alles, and Miklos A. Vasarhelyi. 2014. "A Field Study on the Use of Process Mining of Event Logs as an Analytical Procedure in Auditing." *The Accounting Review* 89 (5): 1751–73.

Jans, Mieke, Michael Alles, and Miklos Vasarhelyi. 2013. "The Case for Process Mining in Auditing: Sources of Value Added and Areas of Application." *International Journal of Accounting Information Systems* 14 (1): 1–20.

Jans, Mieke, N. Lybaert, K. Vanhoof, and J.M. Van Der Werf. 2008. "Business Process Mining for Internal Fraud Risk Reduction: Results of a Case Study." In *Proceedings of the International Research Symposium on Accounting Information Systems*. Paris.

Jans, Mieke, Jan Martijn van der Werf, Nadine Lybaert, and Koen Vanhoof. 2011. "A Business Process Mining Application for Internal Transaction Fraud Mitigation." *Expert Systems with Applications* 38 (10): 13351–59.

Kaiser, Henry F. 1958. "The Varimax Criterion for Analytic Rotation in Factor Analysis." *Psychometrika* 23 (3): 187–200.

Kaiser, Henry F., and K. Dickman. 1959. “Analytic Determination of Common Factors.” *American Psychologist*, no. 14: 425–38.

KPMG. 2017. “KPMG Clara.” *KPMG*. https://home.kpmg.com/xx/en/home/services/audit/kpmg-clara.html.

March, Salvatore T, and Gerald F Smith. 1995. “Design and Natural Science Research on Information Technology.” *Decis. Support Syst.* 15 (4): 251–266.

Medeiros, A. K. Alves de. 2006. “Genetic Process Mining.” Eindhoven: Eindhoven University of Technology.

Naumann, Justus D., and A. Milton Jenkins. 1982. “Prototyping: The New Paradigm for Systems Development.” *MIS Quarterly*, 29–44.

Österle, H., J. Becker, U. Frank, T. Hess, D. Karagiannis, H. Krcmar, P. Loos, P. Mertens, A. Oberweis, and E.J. Sinz. 2010. “Memorandum on Design-Oriented Information Systems Research.” *European Journal of Information Systems* 20 (1): 7–10.

PCAOB. 2010. *Auditing Standard No. 12 Identifying and Assessing Risks of Material Misstatement*.

PwC. 2009. “ERP Risk & Controls.”

———. 2013. “Untangling SAP Security.”

———. 2017. “PwC’s Halo for Journals.” http://halo.pwc.com.

United States Congress. 2012a. *Securities Act of 1933*.

———. 2012b. *Securities Exchange Act of 1934*.

Weijters, A., Wil M. P. van der Aalst, and A. K. A. de Medeiros. 2006. “Process Mining with the Heuristics Miner-Algorithm.” *Technische Universiteit Eindhoven, Tech. Rep. WP* 166.

Werner, Michael. 2016. “Business Process Analysis Automation for Financial Audits.” Von-Melle-Park 3, 20146 Hamburg: Universität Hamburg.

Werner, Michael, and Nick Gehrke. 2015. “Multilevel Process Mining for Financial Audits.” *IEEE Transactions on Services Computing* 8 (6): 820–32.

Wood, Jason, William C. Brown, and Harry Howe. 2014. *IT Auditing and Application Controls for Small and Mid-Sized Businesses*. Wiley Corporate F & A Series. Hoboken, New Jersey: John Wiley and Sons, Inc.

Appendix A Terms and Definitions

Business process	A set of related activities that are carried out to achieve a specific business goal.
Business process activity	A single business activity which belongs to a specific business process.
Business process model	A model in a sense of a graphical representation of a specific business process. A process model abstracts from the individual executions of a business process and represents the behavior of a set of process executions.
Case	The recorded execution of a single process instance. A case is represented in the event log as a set of related events that carry the same case ID.
Critical data constellation	Specific observable data constellation which indicates a missing or ineffective internal control (deficiency in internal control).
Deficiency in internal control	A control is designed, implemented or operated in such a way that it is unable to prevent, or detect and correct, misstatements in the financial statements on a timely basis; or a control necessary to prevent, or detect and correct, misstatements in the financial statements on a timely basis is missing.
Deficiency profile	A set of indicators which refers to a significant amount of process instances. A deficiency profile expresses which systematic combinations of missing or weak internal control (represented by indicators) are hidden in the audited entity's business processes.
Event	The recorded execution of a single business process activity which is stored as an entry in the event log.
Event log	A set of recorded events usually in the form of a simple table. Each event is represented in the event log as a separate row.
Process instance	A single execution of a business process. It is recorded as a case in the event log.
Process instance model	A model in a sense of a graphical representation of a specific business process instance.
Process mining	Business Intelligence technique for analyzing large data sets. It takes an event log as input and creates process or process instance models as output.
Indicator	Represents and describes a critical data constellation.
Indicator constellation	Describes the role of one or more indicators which is or are present in a specific factor.
Indicator instance	Instantiation of an indicator that has a unique ID and is attached to a single instance model or an activity within such a model.

Table 9 Terms and definitions

Appendix B Business Process Model and Control Matrix Examples

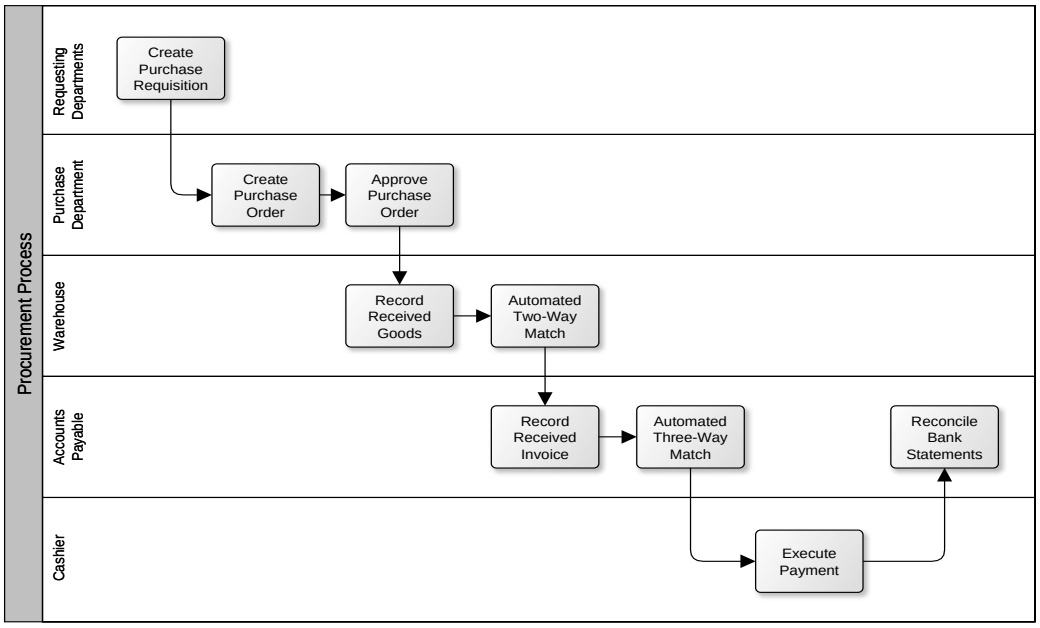


Figure 5 Example of a typical procurement process using Business Process Model and Notation (BPMN)

Control	Name	Description	Type	Control Objective
1	Purchase order approval	All purchase orders over $10,000 have to be approved by the purchase department manager	Manual	Validity, accuracy, restricted access
2	Two-Way-Match	The quantity and quality of received goods is checked against the purchase order	Automated	Validity, accuracy, completeness
3	Three-Way-Match	The quantity, quality and price of received goods is checked against the purchase order and receiving report	Automated	Validity, accuracy, completeness

Table 10 Corresponding control matrix for the business process shown in Figure 5

Figures

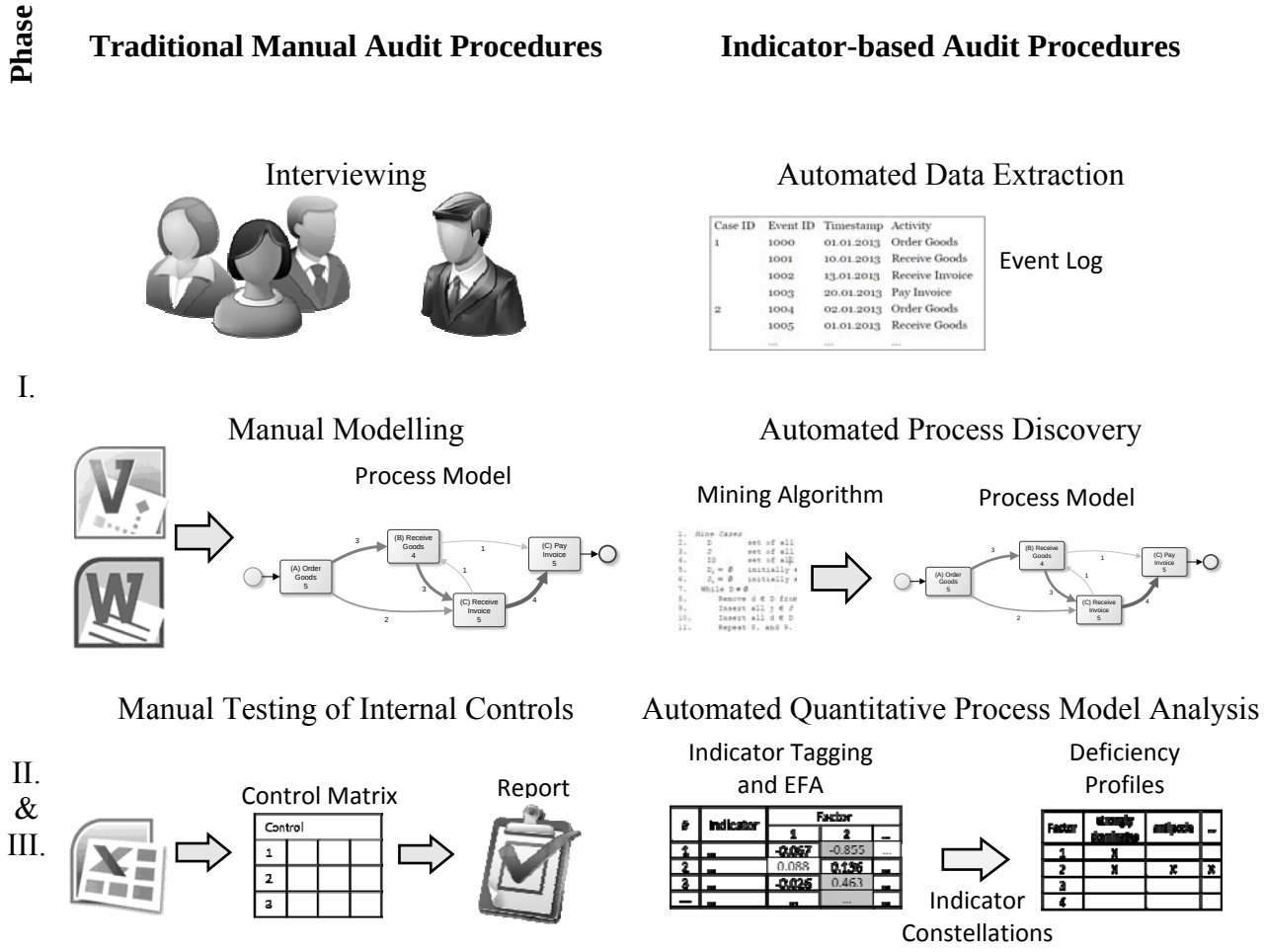


Figure 1 Comparison of traditional and indicator-based audit procedures

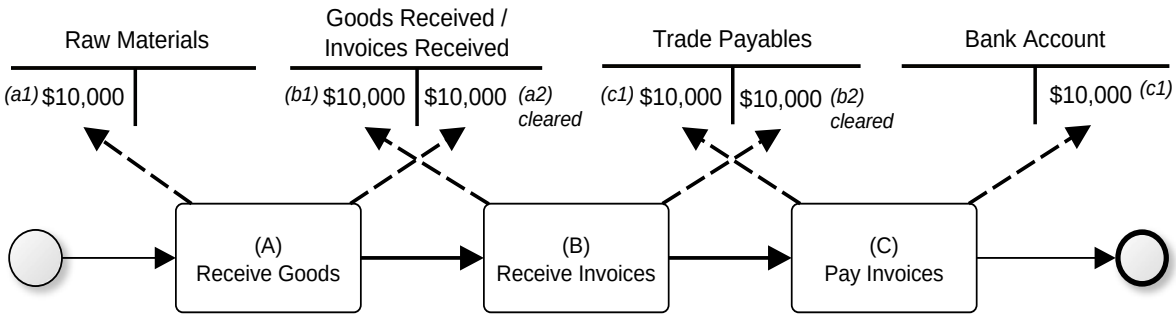


Figure 2 Example of a simple business process model

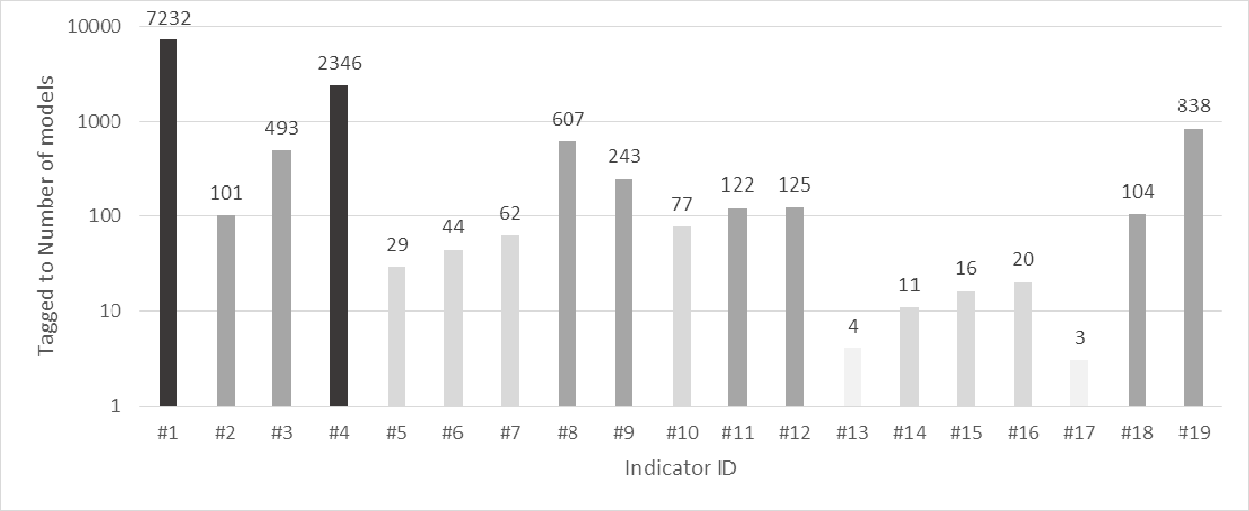


Figure 3 Distribution of indicators tagged to business process instance models

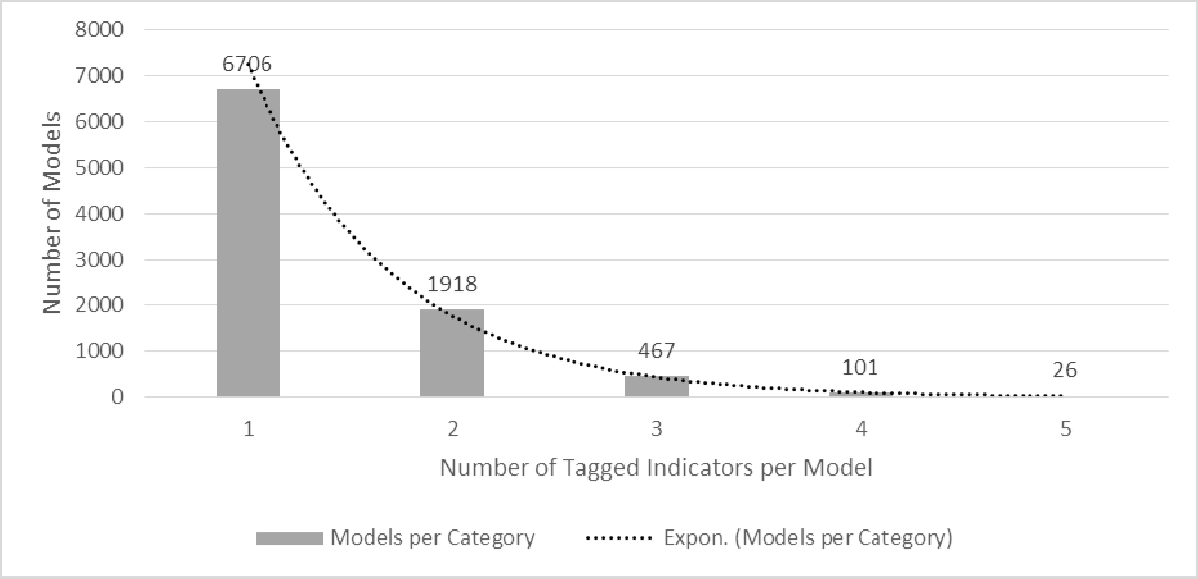


Figure 4 Distribution of indicators per model over the number of tagged models

Tables

Table Name	Table Content	Data Records
BKPF	Accounting document header data	31,283
BSEG	Accounting document segment data	127,159
EKKO	Purchasing document header data	1,579
EKPO	Purchasing document item data	5,895
EBAN	Purchase requisition data	2,304
VBRK	Billing document header data	291
LIKP	Sales document delivery header data	334
VBAK	Sales document header data	224
CDHDR	Change document header data	44,687
Sum		**213,756**

Time period: One year
Industry: Manufacturing
Mined process instances: 25,051

Table 1 Case study dataset overview

#	Indicator
1	A transaction is executed with administrator access rights
2	The complete process instance is executed by a single user
3	An accounts payable subsidiary ledger account is adjusted using the transaction to record an incoming invoice and then concealed via general ledger entries by the same user
4	An incoming invoice is created and payment for it initiated by the same user
5	The customer master data is changed and received cash modified by the same user
6	A sales document is entered and released by the same user
7	A purchase order is maintained and the delivery of services is recorded by the same user
8	An accounts receivable subsidiary ledger account is adjusted using the transaction to record payments and then concealed via general ledger entries by the same user
9	A supplier is created and disbursements directed to the supplier by the same user
10	A sales document is created and a billing document generated for it by the same user
11	An unauthorized item is purchased and hidden by recording just partial deliveries
12	A supplier is set up and payments initiated to this supplier by the same user
13	The customer master data is changed and an unauthorized invoice entered by the same user
14	Expenses are settled from an unauthorized order
15	The customer master data is changed and payments posted to that customer by the same user
16	An accounts receivable subsidiary ledger account is adjusted using payment runs and this is then concealed with general ledger entries
17	A sales invoice is changed and payments are changed for it by the same user
18	The shipment of goods is concealed by maintaining a fictitious sales document
19	An item (not complete purchase requisition) is requested and a purchase order is created from that requisition

Table 2 Exemplary indicators used in the study

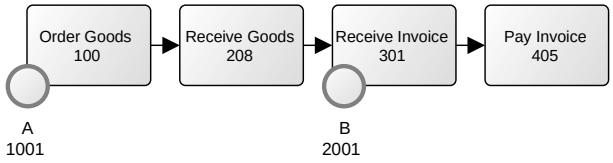
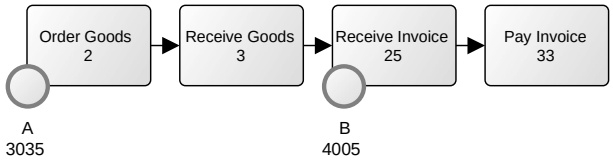
Process instance	Process instances with attached indicator instances	Indicator combinations	Indicator instances
1		A,B	1001, 2001
2		C	3045
3		A,B	3035, 4005

Table 3 Processes instance models and identified indicator combinations

Combination ID (Data Row Number for EFA)	Indicator A	Indicator B	Indicator C
1	1	1	0
2	0	0	1
3	1	1	0

Table 4 Derived data matrix as input for the EFA

Factors	Initial intrinsic values			Sums of squared factor scores for extraction			Rotated sum of the squared scores		
	Total	% of the variance	Cumulated %	Total	% of the variance	Cumulated %	Total	% of the variance	Cumulated %
1	2.234	11.757	11.757	2,234	11.757	11.757	1.866	9.821	9.821
2	2.021	10.639	22.396	2.021	10.639	22.396	1.825	9.606	19.426
3	1.760	9.264	31.661	1.760	9.264	31.661	1.721	9.060	28.486
4	1.526	8.030	39.691	1.526	8.030	39.691	1.560	8,209	36.695
5	1.315	6.920	46.611	1.315	6.920	46.611	1.448	7.619	44.314
6	1.276	6.716	53.328	1.276	6.716	53.328	1.390	7.316	51.630
7	1.178	6.199	59.527	1.178	6,199	59.527	1.351	7.108	58.738
8	1.133	5.963	65.49	1.133	5.963	65.490	1.283	6.752	65.490
Extraction method: Principal component analysis									

Table 5 Factor contribution for explaining the observed variance

#	Variables	Factors							
		1	2	3	4	5	6	7	8
1	A transaction is executed with administrator access rights	-0.067	**-0.855**	-0.242	0.089	-0.155	-0.129	-0.078	-0.116
2	The complete process instance is executed by a single user	0.088	0.136	0.118	-0.025	0.008	-0.048	0.050	**0.764**
3	An accounts payable subsidiary ledger account is adjusted using the transaction to record an incoming invoice and then concealed via general ledger entries by the same user	-0.026	**0.463**	-0.106	0.16	-0.151	-0.175	0.005	0.044
4	An incoming invoice is created and payment for it initiated by the same user	-0.020	**0.839**	-0.074	-0.084	0.156	0.128	-0.036	-0.03
5	The customer master data is changed and received cash modified by the same user	0.313	0.04	0.026	-0.012	-0.053	-0.011	**0.678**	-0.038
6	A sales document is entered and released by the same user	0.017	0.049	0.195	**0.799**	0.003	-0.036	-0.035	-0.015
7	A purchase order is maintained and the delivery of services is recorded by the same user	0.007	0.037	-0.009	**0.845**	-0.012	-0.054	-0.005	-0.002
8	An accounts receivable subsidiary ledger account is adjusted using the transaction to record payments and then concealed via general ledger entries by the same user	-0.027	-0.061	-0.117	0.397	0.008	0.083	0.020	-0.005
9	A supplier is created and disbursements directed to the supplier by the same user	0.003	0.144	-0.003	-0.020	**0.815**	0.06	-0.012	-0.027
10	A sales document is created and a billing document generated for it by the same user	-0.013	-0.008	**0.875**	-0.019	-0.011	-0.008	0.103	0.014
11	An unauthorized item is purchased and hidden by recording just partial deliveries	-0.003	-0.162	-0.011	0.073	-0.064	**0.842**	0.011	0.034
12	A supplier is set up and payments initiated to this supplier by the same user	0.003	-0.02	-0.019	0.027	**0.826**	-0.034	0.01	0.033
13	The customer master data is changed and an unauthorized invoice entered by the same user	**0.713**	-0.033	-0.041	0.013	0.09	0.005	0.265	0.069
14	Expenses are settled from an unauthorized order	-0.033	-0.042	-0.056	0.003	-0.002	0.03	-0.031	**0.816**
15	The customer master data is changed and payments posted to that customer by the same user	-0.046	-0.006	0.063	0.005	0.04	-0.001	**0.892**	0.048
16	An accounts receivable subsidiary ledger account is adjusted using payment runs and this is then concealed with general	**0.688**	0.041	0.045	-0.026	-0.057	-0.015	0.01	-0.01

	ledger entries								
17	A sales invoice is changed and payments are changed for it by the same user	**0.876**	-0.015	-0.016	-0.001	0.003	0.006	-0.018	0.012
18	The shipment of goods is concealed by maintaining a fictitious sales document	0.012	0.008	**0.896**	0.014	-0.013	-0.015	-0.016	0.044
19	An item (not complete purchase requisition) is requested and a purchase order is created from that requisition	-0.007	0.31	-0.019	-0.032	0.102	**0.773**	-0.025	-0.057
Extraction method: Principal component analysis									
Rotation method: varimax with Kaiser normalization, the rotation converges after 6 iterations									

Table 6 Extracted factors

Value range	Name of the constellation	Coloring
an individual indicator ≥ 0.75	*'strongly dominates'*	***Black***
one indicator with a medium value $0.4 \leq x < 0.75$	*'weakly dominates'*	***Grey***
multiple indicators ≥ 0.75	*'collaboratively strongly dominate'*	***Striped downwards***
multiple indicators with a medium value $0.4 \leq x < 0.75$	*'collaboratively weakly dominate'*	***Striped upwards***
one or several indicators *'strongly negative'* ≤ -0.3	*'antipode'*	***Checkered***

Table 7 Threshold values for the interpretation of factors and deficiency profiles

Factor	strongly dominates	collaboratively strongly dominate	antipode	weakly dominates	collaboratively weakly dominate
1	#17				#13, #16
2	#4		#1	#3	
3		#10, #18			
4		#6, #7			
5		#9, #12			
6		#11, #19			
7	#15			#5	
8		#2, #14			

Table 8 Identified deficiency profiles