CHALMERS
UNIVERSITY OF TECHNOLOGY

An investigation into how error taxonomies can generate actionable data

Master's thesis in Management and Economics of Innovation

Gustav Bergling

Péter Áron Herszényi

DEPARTMENT OF TECHNOLOGY MANAGEMENT AND ECONOMICS
Division of Service Management & Logistics

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2022
www.chalmers.se
Report No. E2022:024

REPORT NO. E2022:024

An investigation into how error taxonomies can generate actionable data

Gustav Bergling
Péter Áron Herszényi

Department of Technology Management and Economics
Division of Service Management & Logistics
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2022

An investigation into how error taxonomies can generate actionable data
Gustav Bergling
Péter Áron Herszényi

Report no. E2022:024
Department of Technology Management and Economics
Chalmers University of Technology
SE-412 96 Göteborg
Sweden
Telephone + 46 (0)31-772 1000
Gothenburg, Sweden 2022

Abstract

Errors are phenomena that individuals and organizations are all familiar with and must accept to be inevitable. In an organizational context, errors are often feared because of the potential negative outcomes they may lead to. Therefore, much thought and effort has been invested by both academics and practitioners into understanding errors and mitigating their negative effects. One way in which organizations do this is to use error taxonomies, that are essentially systems for categorizing errors, in hopes of highlighting error trends such as error recurrence. However, a major concern is how actionable the data that these error taxonomies produce is. This study was done in collaboration with an anonymous case company at their R&D department and this company has been the sole source of empirical data. The purpose of this study is to understand the usefulness of the current error taxonomy, at the case company, in assisting to measure error recurrence patterns. The research design of the study is mixed, using both quantitative and qualitative data collection methods. The results highlight several aspects of the current error taxonomy that make it difficult to make sense of the data that is being generated. Most importantly, the data is not deemed to be actionable, and it does not provide good support for measuring error recurrence. Therefore, this study makes actionable recommendations on how the error taxonomy can be improved, based on a variety of error related literature and the empirical findings from this study. The recommendations are primarily aimed at the case company, but it is deemed that they are relatively generalizable to any organization that is working with an error taxonomy. The case company was very pleased with the results of this master's thesis, and they assured that it will act as a valuable source of input for management in the discussions on how to proceed with their error taxonomy.

.

Keywords: Error, Taxonomy, Category, Categorization, Error Management, Culture, Innovation, Error patterns, Error trends, Actionable research, Recurrence

Acknowledgements

We would like to foremost thank our supervisor and examiner, Hendry Raharjo. His experience and guidance have been invaluable for our master's thesis and has been a blessing during periods of struggle. Further, we would like to thank Anette and the case company for giving us the opportunity for the collaboration. In addition to this, we would like to thank the QI-team for helping us with internal case company related matters and for the auditors taking time from their busy schedule to participate in our interviews. Lastly, a special thanks to our case company mentor, Martin, whose help, and expertise has been radiant, and has made sure we felt welcomed at the company.

Table of contents

1. Introduction

This initial chapter presents the background of this study, outlining the interfacing research fields. Furthermore, the case company that has been at the core of the study will be presented along with the overall purpose statement and accompanying research questions. Finally, remarks will be made on economic factors, societal factors, and delimitations.

1.1 Background

Errors are phenomena that all of us can relate to in one way or another. They occur in the everyday life of people, and they are inevitable. Errors usually have a negative implication and people tend to be afraid to commit them due to potentially undesired consequences. Errors are also relevant in organizational settings where they can range from an employee accidentally spilling their cup of coffee over their keyboard, to an entire nuclear power plant melting down, as a consequence of human error, which was the case in Chernobyl. Considering the potentially severe impact errors might have on individuals and organizations, it is not surprising that scholars have tried to make sense of the concept. Starting at the most fundamental level, scholars try to define what an error actually is. An example shall follow. Do note that the following presented definitions will not be used in the later sections of this study, they simply aim to provide context and give a brief introduction to the research field on errors.

Carroll et al. (2021) defines an error as: *"An error is typically defined as an unintended and potentially avoidable deviation from organizationally specified goals and standards that can yield adverse, positive, or no organizational consequences" (p. 451).*

Anu et al. (2018) provides a basic understanding of errors, although they argue that a more specific explanation is needed, breaking errors down into comprehensible elements. Zapf et al. (1992) explain the process of errors using action-based theory. Firstly, in order for an error to exist, there first has to be an action with a goal. Without the goal, the action would simply be random. Second, an error is therefore the non-fulfillment of the goal previously decided. Lastly, the error cannot be beyond human control, such as losing data due to a natural disaster.

Correspondingly, Frese & Keith (2015) separate errors and risks. Risks are present within the environment whilst errors have the possibility to occur during interaction with the environment. Moreover, the environment can be seen as a factor increasing the likelihood of an error occurring. Frese & Keith (2015) argue that if the risk of not reaching a goal is higher than 50%, it cannot be seen as an error. However, it is possible that the risk calculations are incorrect and that there is indeed an underlying error present. Nevertheless, when these factors are taken into consideration, in order for an error to occur, it has to be an unintentional, non-fulfillment of a goal. Moreover,

violations occur when the process of reaching the goal is intentionally not followed. Hence, a person breaking the rules by choice rather than accident is committing a violation, not an error.

Separating errors, violations, and risks as well as differentiating between human factors and environmental factors leading to the non-fulfillment of a goal, sets a baseline for understanding errors. However, as the previous paragraph explained, an error itself is not the outcome of an action but rather the deviation from the planned action. Thus, the outcome of the error can be negative or positive. Carroll et al. (2021) stresses the importance of understanding that the outcome of an error does not necessarily have to be a failure. For instance, giving a patient the wrong dosage of a drug is an error. However, the outcome may be positive or negative. For example, the outcome may be that the patient suffers even more or that the new dosage leads to the invention of a new drug (Frese & Keith 2015, Carroll et al. 2021). Hence, the outcome of an error is what is most likely experienced when the error occurs. The error itself, however, is simply the deviation from the planned action.

Furthermore, Carroll et al. (2021) distinguish between different levels at which errors can occur. Foremost, they propose three levels: Individual, Team, and Organization and Systems. The Individual level assumes that an increased level of complexity will increase the likelihood of errors when performing tasks. In addition to this, the most common behavior of individuals when errors occur is to stop their current actions and try to figure out a solution to the errors, as errors will most likely result in bad emotions. The Team level highlights how the team-design may impact the possibility of errors. A poorly designed team is likely to negatively impact communication and coordination which increases the probability for errors to occur. Lastly, the organization and system level are about the error culture. The consensus within an organization of how to manage errors will impact the overall ability to learn from errors. Thus, Caroll et al. (2021) emphasize that open communication and acceptance around errors has to be established. In principle, all identified errors should be brought forward in order to be solved and to generate learnings from the solutions. In contrast to this, a culture where errors are blamed on individuals will result in tense emotions, which is counterproductive for the entire organization. To sum up, errors should be regarded on an organizational level and considered to be great learning opportunities.

1.2 Introduction to the case company

In order to better understand how errors impact organizations in practice, a case company has been closely collaborated with during this study. All the empirical data generated in this study derives from the case company. This company wishes to stay anonymous, which is not deemed to limit the study in any way. Do note however, that certain data throughout the report, related to the case company, is displayed in a general fashion. The benefit of this is that the results of the study are expected to be more generalizable. The case company is a large international player that is heavily engaged in both manufacturing and R&D activities. The collaboration is with a subdivision of

R&D and therefore the scope of the study is also limited to only investigate errors that take place within the context of R&D. The case company operates in a regulated industry where both internal and external audits take place to ensure that regulations are properly adhered to. This study is limited to only focus on errors that are detected during internal audits. Please note that the terms “error” and “finding” will be used interchangeably throughout this report. Both terms refer to the detected outcomes of errors, during internal audits, with the difference being that a finding refers to the documented entry in the error database of the case company. The finding provides a description of what has occurred, and various complementary data points related to the error. Furthermore, the finding also shows how the error has been categorized, using the error taxonomy of the case company.

An error taxonomy is essentially a system for categorizing errors. Many different terms are used throughout the research literature that refer to systems of error categorization such as, error taxonomy, error categorization system, error classification system, error classification scheme or error categorization framework, to name a few. Naturally, there are different contextual details or other subtle nuances that are different when these terms are used but they all essentially refer to the same thing, namely systems for categorizing errors. Throughout this study, the term error taxonomy will be used except for when referring to other studies where a different term might be used.

The error taxonomy of the case company is currently categorizing findings into main and subcategories. Some of these categories are rather broad which is a consequence of a recent update that was initiated based on feedback that the categories were too narrow, which made the taxonomy difficult and tedious to use. However, broadening the categories is believed to have brought forth other problems. In particular, as the new error taxonomy has been rolled out and started generating error data, concern has emerged, when observing the data, that the implications of the data are ambiguous.

Furthermore, there are concerns regarding the validity of the data, in particular, it is questioned whether the error taxonomy is used as intended and whether it is capable of producing the desired data. The ultimate goal is for the error taxonomy to produce data that identifies error patterns that can guide managerial interventions. The particular error pattern of interest at the case company is error recurrence. This is problematic to clarify as there is no internal definition of error recurrence. This might seem somewhat strange as the notion of error recurrence has a substantial intuitive component to it. This is not least observable in the research literature where a majority of studies do not attempt to further define the term but instead assume its intuitive meaning, which is that a recurrent error is simply the same error occurring a second time.

When examined further, the notion of error recurrence might not be as obvious as it may seem, primarily because the level of detail to describe an error can vary a lot. For example, if an employee accidentally blocks a fire escape route, and then on a different day, the same happens again, should the error be seen as recurrent if a different employee commits the error the second time? What about if it is a different door being blocked at a different department of the organization or even at a different site in a

different country. The example should illustrate the complexity of the problem and, of course, one can proceed to add even more details.

The case company is currently attempting to define error recurrence from a top-down perspective, meaning that upper management is trying to find a definition that is rigorously supported by academic research. This definition would then be implemented into the error taxonomy in hopes of assisting in generating the desired data about error recurrence patterns.

The case company has stated clearly that they have found limited publications that further dissect the concept of error recurrence in ways that they would have hoped. The closest example of what is being looked for is a proposition made by Cahill (2017), who separates errors as "repeat" or "recurring". These terms are intended to be used specifically in the context of audits. Repeat errors are defined as those that occur due to a flaw in the management system. Specifically, repeat errors are those which have been detected in a previous audit, but action taken to solve the error has not been taken. Therefore, the same error is identified in the next audit. An example would be that a certain workstation is not designed in an ergonomically favorable way. This might be noted in the audit report with clear instructions on how to improve the situation. If no action is taken to improve the situation and the workstation is found exactly as it was during the previous audit, it is an indication that the management system has failed to direct the necessary resources to fix the problem.

Recurring errors, on the other hand, are viewed as isolated occurrences meaning that they cannot be conclusively attributed to a faulty management system. These types of errors are also ones that have been found in previous audits, but the cause cannot conclusively be attributed to the lack of managerial intervention or a weak management system. Instead, these errors are deemed to be natural, or random, in nature and are expected to occur within any system to some no-zero degree, regardless of how good the management system is; an example shall follow. All fire extinguishers need to have a pin that must be removed before they can be used. There is a non-zero statistical probability that the pin falls off, for no obvious reason, and therefore, a small percentage of extinguishers might naturally have their pins missing. This might be noted at an audit and, subsequently, expected that the pins are put back in place. Assume that the management system then manages to coordinate resources accordingly and actually replace the pins. However, at a different audit, the pin on a different extinguisher has fallen off, which gets detected. According to Cahill (2017), this would be a recurring error.

The high context specificity of the proposition made by Cahill (2017) makes it difficult to directly apply in practice. Furthermore, the observation by the case company that similar publications are limited makes it difficult to find rigorous academic support for a top-down definition of error recurrence. During the early stages of this study, it was verified that it is indeed difficult to find such publications and therefore, a slightly different approach had to be taken. Acknowledging the lack of a top-down definition of error recurrence, several imaginable alternative ways of assisting the measuring of

error recurrence were discussed. The limitation was ultimately made that this study shall focus on the error taxonomy and develop insight of how it can be optimally used to inform the measuring of error recurrence. An important aspect of errors that is conducted by the case company but excluded from the scope of this project, is root cause analysis.

1.3 Purpose & research questions

The purpose of this study is to understand the usefulness of the current error taxonomy, at the case company, in assisting to measure error recurrence patterns. The study aims to answer two research questions in order to develop the understanding implied by the purpose statement:

1. What are the benefits and potential problems of the current error taxonomy?
2. How can the current error taxonomy be improved to better fulfill the stated purpose?

The term "usefulness" in the purpose statement, requires some further clarification. As stated, it mainly refers to the ability of the error taxonomy to assist in measuring error recurrence but there is more to it than that. The case company is asking for a project that delivers highly actionable results rather than theoretical implications and suggestions for further research alone. For this reason, a body of scientific literature has been consulted on how to generate research results that are highly actionable for practitioners. This literature is presented in the theoretical framework in chapter 2. Thereafter, the methodology that has been used in this study will be presented in chapter 3. The findings of the study will be outlined in chapter 4, results. Chapter 5 will make connections between chapter 2 and 4. Finally, chapter 6 will conclude and propose a couple of recommendations to the case company.

1.4 Economic and societal considerations

Errors have an impact economically as well as socially. Errors can cause harm to people and equipment and the consequences can be expensive. Furthermore, errors may impact the products that are being produced, which in turn might negatively impact the end-users. For these reasons it is important for organizations to identify errors and understand why they occur so that appropriate interventions can be coordinated. The hope for this study is that it can help in taking a small step in that direction.

1.5 Delimitations

Finally, this chapter will be ended with a note on delimitations. The authors are aware that the challenges faced by organizations are multidimensional. The limitations of scope imposed when conducting research can therefore result in findings that misrepresent or underrepresent certain aspects of reality. Therefore, the authors would like to emphasize that no harmful criticism is being directed towards the case company

or any employee within. It is very possible that certain vital aspects have been missed due to the limitations in scope and in such cases the relevance of the findings must be carefully evaluated. All presented findings are based on a most genuine attempt by the authors to delineate a fair representation of reality as suggested by the empirical findings. In some cases, this may create the perception of certain aspects being black or white when in reality it is much more of a gray zone. In such cases, the authors would like to apologize as there is absolutely no intention to portray the situation at the case company, and the employees which it may concern, in an unjust or disingenuous manner. The authors therefore strongly implore readers to keep this important aspect in mind.

2. Theoretical framework

This chapter presents the theory that is used as a basis for the final recommendations of this study after being compared to the empirical findings. Three sections will be presented where Sections 2.1 and 2.3 cover different aspects of errors in organizations. Section 2.2 is more general in the sense that it presents literature related to the actionability of research results.

2.1 Error taxonomy

An error taxonomy is essentially a system for categorizing errors. A primary concern related to error taxonomies is what the purpose of them is. Several researchers have come up with different propositions. Stanton & Salmon (2009) explain that the purpose of an error taxonomy is foremost to identify and gain a deeper understanding of errors. Furthermore, an error taxonomy should aim to inform strategies on how to manage errors and prevent them from happening in the future. Hooper & O'Hare (2013) explain that the purpose of error classification is to generate and comprehend information related to errors in order to gain a deeper understanding of why errors occur, which can aid in preventing them from happening again. In particular, an error classification system that fulfills its purpose should guide managerial interventions by precisely highlighting in which areas improvements should be implemented (Hooper & O'Hare, 2013). Another purpose statement is provided by Zhang et al. (2004) who argue that the purpose of an error taxonomy is to gain insights about the system within which humans operate. This is done by identifying strengths and weaknesses of the system by identifying where errors are occurring frequently and guiding remedial interventions to those areas. Zhang et al. (2004) argue that this purpose statement is relevant in any context. Walia and Carver (2012) argue that an error taxonomy can help inspectors identify and understand errors and thereby generally increase their knowledge about the system.

Love et al. (2013) stresses the importance of being able to understand recurrent error patterns, termed "error traps" within the system. Error traps are fundamentally errors that keep appearing regardless of who is performing the job. Thus, they are regarded to be system-wide flaws and are expected to keep appearing until the system is upgraded. Therefore, errors cannot fundamentally be solved unless the underlying causes are known. An error taxonomy can aid in finding these patterns and thereby generate valuable information to better understand the flaws of a system. Moreover, Reason and Hobbs (2003) stress the importance for organizations to identify error traps. The way to identify error traps is to look for error patterns as these may be indicative of an error trap being present within the system.

Walia & Carver (2012) also provide several insights related to important factors to consider when designing an error taxonomy. One of these insights is that the perceived ease of use of the taxonomy is critical and therefore Walia & Carver (2012) recommend that the users of the error taxonomy undergo proper training in how to use it. Another

insight is that the abstraction level of the error taxonomy needs to be considered, a too high degree of generalization is not likely going to be effective. Therefore, the different contexts within the organization need to be considered and an appropriate abstraction level should be selected for the error taxonomy (Walia & Carver, 2012).

More factors to consider related to error taxonomies are outlined by Tamuz et al. (2004). Firstly, there has to be agreement between the users of the taxonomy regarding definitions within the taxonomy, otherwise the reliability of the taxonomy might suffer. A potential consequence when consensus is not reached regarding definitions is that certain users under-report error-related events. Furthermore, it is highlighted that misinterpretation of the categories within the taxonomy which in turn influences the generated data can have significant impact on the course of action the organization takes and can therefore carry high risk.

Another factor presented by Tamuz et al. (2004) relates to the distribution of information within the organization. If the error taxonomy data is fragmented by being stored in different databases in different parts of the organization, it can significantly hinder access to information. This can impede the detection of negative trends and subsequently the learning that could have occurred if the information was disseminated as intended. Furthermore, if information is only available within certain subdivisions of the organization, it can lead to learning taking place locally. Such practice risks leading to a silo mentality. Cilliers & Greyvenstein (2012) explain that the concept of a silo mentality within an organization entails employees from different subdivisions working isolated from each other. This creates subcultures within organizations that can sometimes even be hostile against one another but most importantly it can hinder the sharing of information and knowledge between divisions. Therefore, it is important that the storing of data is highly coordinated, and that important information gets disseminated throughout the organization (Tamuz et al., 2004).

Next, an evaluation framework of error taxonomies by Wiegmann & Shappell (2001) will be presented. This framework presents criteria against which an error taxonomy can be evaluated with the premise being that the effectiveness of a taxonomy is higher the more criteria it satisfies. These criteria are reliability, comprehensiveness, diagnosticity, useability and validity.

Reliability refers to the ability of the error taxonomy to produce reliable insights, in other words insights that can be trusted. If this criterion is not met, the value of the taxonomy is highly questionable. An important factor for generating reliable data is that the error taxonomy needs to be closely tailored to the reality in which the users of the taxonomy operate because if users are unable to relate the taxonomy to their context, they are unlikely to report reliable data.

Comprehensiveness refers to the extent to which the error taxonomy captures all the relevant information and context related to an error. However, it should be noted that the level of information that is necessary for a taxonomy to be deemed comprehensive can vary. For some types of errors, it might be less important to document lots of

complementary information, whereas for others, a considerable amount of contextual information might be needed in order to make sense of the recorded data. This can be problematic when designing an error taxonomy, especially if attempting to generalize over a number of different contexts where there might be a high variety of errors that require different levels of contextual information. In particular it is common to see error taxonomies that generate data full of missing fields because the contexts got mismatched and the expected input data was therefore inapplicable. This frustrates the users of the taxonomy, and it creates problems with the readability of the data, making it difficult to identify patterns or trends.

Diagnosticity refers to the extent to which the error taxonomy is able to highlight the root causes of the errors. Furthermore, for a taxonomy to be considered to have high diagnosticity it should pinpoint interrelationships between errors, reveal emerging or shifting trends and guide managerial intervention. A taxonomy with high diagnosticity should replace the need for managers to make decisions based on intuition or experience and instead rely on the data to guide their work of trying to prevent errors from happening.

Usability refers to how difficult it is to use the error taxonomy. This can be a subjective matter and designers of the taxonomy might have different perceptions of what ease of use entails compared to those who are expected to use the taxonomy. In order to design a taxonomy that is usable it is advised that the end-users of the taxonomy should be involved in the process of designing it by providing input on how its use can be simplified. Furthermore, technical, or scientific terms should be replaced by simple and intuitive language. It is argued that if the usability of the error taxonomy is deemed low, operators might find it annoying to use it and thereby get demotivated to use it properly.

Validity is an assessment of how well the error taxonomy is fulfilling its purpose which is strongly dependent on how it fulfills the previous four criteria. Validity is divided into three types, face validity, content validity and construct validity. Face validity refers to whether the error taxonomy is perceived valid by the people that use it. Content validity refers to whether the taxonomy covers a satisfactory scope, in this case whether the categories fully cover the spectrum of potential errors that can occur within an organization. Construct validity refers to the extent to which the error taxonomy explores the root cause of the errors.

2.2. Actionability of management research

The actionability of management research, specifically how to produce knowledge that is useful in practice, is a long-debated topic. Over three decades ago, Shrivastava (1987) defined the actionability of management research as:

"This refers to the extent to which research results are operationalizable through concrete actions or decisions. The more abstract and general research results are the more difficult it is to translate them into actions" (Shrivastava, 1987, p. 79)

More recently, HakemZadeh & Baba (2016) explain the concept of evidence-based management as attempting to narrow the gap between practice and knowledge created through management research. The goal is for researchers to generate knowledge that is deemed useful for managers. However, the gap is difficult to bridge in practice as researchers and managers operate in different realities, with considerably different purposes. Moreover, different logic and language is used by the two different groups, further increasing the discrepancy (HakemZadeh & Baba, 2016).

HakemZadeh & Baba (2016) argue further that management research should be regarded as any other professional research field with the principal objective being to generate critical knowledge that is useful to practitioners. The generated knowledge should be such that it can be used when training professionals as well as in university education, ultimately leading to enhanced quality of managerial practice. The balance between keeping management research highly rigorous and simultaneously providing actionable managerial implications has been a primary concern of literature within the subject for a long time (HakemZadeh & Baba, 2016). This process is not free from difficulties, Pearce & Huang (2012) point out that there is a high degree of disagreement between researchers of what actionable research actually is. Pearce & Huang further highlight the problem that if researchers have such difficulty in reaching common ground, teachers of management courses will face an even greater difficulty as they try to convey the concepts to future leaders.

HakemZadeh & Baba (2016) point out that the choice of research questions can enhance the practicality of the research. The research questions should be closely related to the context of the practitioners, which is ideally done by defining the research questions from within a practical setting, such as from within an organization. Furthermore, to reach the highest degree of relevance between research and practice, data is ideally gathered from within organizations HakemZadeh & Baba (2016).

It is not uncommon that managers feel that management research has limited usefulness and Sartof (2008) states that managers quickly ignore research when the relevance and practicality is questionable or ambiguous. Syed et al. (2009) highlights that manager tend to deem research more useful when the dependent variable is highly relevant to their context and the independent variable is actionable. However, Rousseau (2006) stresses the importance of striking a balance between "Big E" and "Little E" evidence. In this context, Big E evidence refers to actionable knowledge that is strongly supported by research and Little E evidence refers to knowledge that lacks clear support from

research and has instead been generated within an organization. Furthermore, Little E can also refer to knowledge that is based on the judgment and experience of managers.

Attempts have been made by researchers to get a deeper understanding of what is perceived to be actionable research by practitioners. Mohrman et al. (2001) conducted an exploratory investigation involving ten participating companies where six hypotheses related to the usefulness of research were evaluated. Their study showed different levels of empirical support for each of the hypotheses. The insights with the highest empirical support will be summarized next.

The first finding that was strongly supported by evidence relates to the concept of self-design. This concept refers to the cognitive and behavioral processes that members of an organization employ whilst redesigning or reorganizing their own organization. The finding from the study is that the more the results of the research inform self-design, the more useful it is perceived to be by members of the organization. In layman terms, this finding can be interpreted as the research being deemed useful to the extent that its findings impact critical activities within the particular organization of the recipient.

The second finding relates to the concept of "perspective taking" which recognizes that academic researchers and practitioners have clearly distinct "thought-worlds", which refers to the two groups operating under different paradigms that require different premises and thought processes. Perspective taking refers to the extension of the thought-world of one community by understanding and recognizing the thought-world of another community. The study provides evidence that perspective taking positively impacts the perceived usefulness of research.

The third finding relates to the concept of "interpretive forums" which can be seen as an extension of perspective taking. In addition to recognizing and interpreting the thought-world of a different community, Interpretive forums involve joining the thought-worlds of two different communities. In practice, the concept involves members from the two communities, e.g., the academic research community and the practitioner community, coming together to jointly interpret and reflect upon information and thereby create knowledge together. The study provides evidence that establishing interpretive forums positively impacts the perceived usefulness of research to a larger extent that perspective taking does.

Next, a framework will be presented that is proposed by HakemZadeh & Baba (2016) that summarizes 7 criteria for research in order to be considered to be actionable. The following criteria are proposed: operationality, causality, contextuality, comprehensiveness, persuasiveness, comprehensibility, and conceptual clarity. HakemZadeh & Baba (2016) argue that a higher number of these criteria being met by a study result in a higher degree of actionability. These criteria provide a basis for comparing the findings of this master´s thesis with the theoretical framework. The attentive reader will notice that there is varying degree of overlap between certain criteria.

Operationality refers to research that makes pragmatic recommendations. Furthermore, the recommendations have to be readily implementable in order to be considered operational. By readily implementable HakemZadeh & Baba (2016) refer to recommendations whose implementation could technically begin instantly in a cost effective, legal, and reasonable manner.

Causality refers to the ability of decision makers to reliably predict the outcomes of managerial interventions and thereby make more informed decisions that are more likely to result in the desired outcomes. For research to be considered causal in this context, it must present the cause-and-effect relationships between the managerial interventions and their expected consequences. Typological theories are deemed to effectively demonstrate such relationships.

Contextuality acknowledges that the validity of a specific theory depends on contextual factors such as culture, industry standards or economic viability. It is therefore difficult to generalize theories about managerial implications since one theory might be perfectly applicable under a certain set of contextual factors but completely non-viable under another set of factors. Research that is considered to be contextual clearly outlines under which contextual factors its findings or theories are relevant.

Comprehensiveness is not a trivial concept. The reality which managers and their organizations face are oftentimes highly complex and dynamic. In parallel, researchers must make limitations in their study to maintain a balance between academic rigor and the practical viability of conducting the study. This essentially results in some form of simplification of reality having to be made. Therefore, there is an inherent conflict between the reality of researchers and that of managers. Because of the need for researchers to simplify the complexity of reality, HakemZadeh & Baba (2016) claim that no single study within the field of management can fully explain a particular phenomenon that challenges organizations. With this background, in order for research to be considered comprehensive it must delineate the complexity of the studied phenomenon in a sufficiently comprehensible manner and at least provide a direction in terms of how to manage the complexity.

Furthermore, the research must acknowledge its limited ability to solve the complexity, perhaps it can serve as a part of a multifaceted, larger scale solution. Providing a direction to managers may for example entail referring them to complementary research that provides further perspective on the complexity. Thereby the manager can eventually reach a state of knowledge where an informed decision can be made. To conclude, sometimes management research cannot directly yield actionable results due to the complexity of the phenomenon being studied. In such cases, research can still be deemed actionable by clarifying scope and limitations as well as providing a direction for how necessary complementary knowledge can be obtained.

Persuasiveness concerns trusting in the quality and validity of the research. Quality is ensured by peer reviews, replications of the study and getting published in esteemed management journals. Validity is ensured by presenting clear evidence for the recommendations being made.

Comprehensibility refers to the difficulty for decision makers to understand the results of the research. A common mistake is to not sufficiently consider that the recipients of the study, in this case decision makers, might have a lower academic literacy compared to the researcher. Presenting complex subject matter knowledge and using highly technical terms can considerably limit the accessibility for non-academics. This often leads to recipients losing interest and deeming the research to have low actionability. Therefore, comprehensible research delivers high quality and high validity balanced with a language level that is appropriate for the target audience.

Conceptual clarity fundamentally refers to obtaining a holistic understanding of the underlying concepts that inform a particular decision. Oftentimes there are several underlying concepts that are related and research that is able to highlight these connections between relevant concepts comprehensively is deemed to provide conceptual clarity. Understanding the different factors that lead to a decision and the interplay between them put managers in a better position to make decisions in highly uncertain environments.

2.3. Error culture and innovation within organizations

Dimitrova et al. (2014) discuss the concept of Error Management and Error Prevention. These are two separated but still related error handling strategies which are defined as processes of dealing with errors either before or after they have occurred. Figure 2.0 shows the relationship between Error Prevention and Error Management. Error Prevention is primarily focused on the time before the error occurs and Error Management is focused on the time after. Whilst primarily being considered to be strategies, these concepts have an impact on the error-related cultures and the degree of innovativeness of organizations.

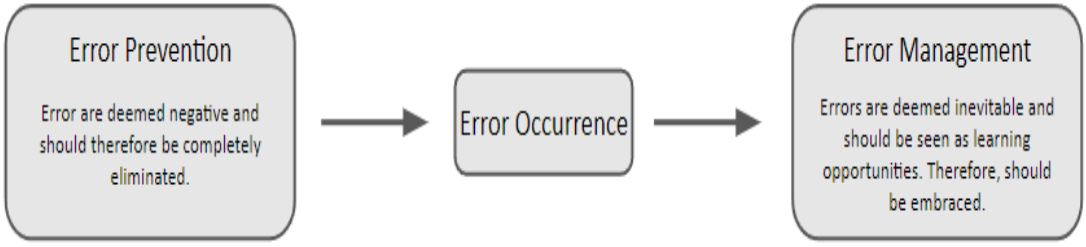

Figure 2.0 - The relationship between Error Prevention and Error Management.

Dimitrova et al. (2014) bring forth that solely focusing on Error Prevention has shown to bring negative results. For instance, it has shown to decrease learning from errors,

people hiding committed errors, poor performance and lower psychological safety. This result does not suggest that Error Prevention is an unnecessary strategy, however. Dimitrova et al. (2014) state that the context of the task or action always has to be considered. In the context of highly complex tasks, a pure Error Prevention strategy would most likely yield the results previously described. However, in a context where the task is rather straightforward and predictable, a pure Error Prevention strategy will likely yield better results than Error Management. Thus, the context always has to be considered.

Dimitrova et al. (2014) bring forth that Error Management is a strategy related to limiting, or completely disconnecting the error from its potential negative effect. As a result, the strategy emphasizes accepting errors as inevitable, and therefore to embrace them. Furthermore, the focus is to increase the positive consequences of errors, such as learning from them and thereby increasing knowledge. Error Management has also shown to reduce bad emotions and positively increase motivation and performance when implemented in complex task environments.

In addition to this, an Error Management strategy has proven to be beneficial in situations where familiar tasks are performed. However, Matthews et al. (2022) emphasize that maintaining a culture of Error Management is easier said than done. Humans are instinctively aligned with error aversion, and therefore Error Management requires people to change their habitual behaviors and mindsets. Dimitrova & van Hooft (2021) state that performance-focused leaders tend to be less inclined to embrace Error Management since they fear that employees would regard them to be incompetent or weak. However, the evidence of the study by Dimitrova & van Hooft (2021) showed the opposite. Employees view leaders that embrace Error Management as highly competent and caring about the needs of the employees. Therefore, an emphasis on Error Management tends to indicate higher job satisfaction, lower employee turnover and increased leader-rated employee performance.

Frese & Keith (2015) argue that the overall most common approach chosen by leaders has been Error Prevention, mainly because errors have shown to lead to some catastrophic consequences. However, they argue that errors can never be reduced to zero. Thus, it is argued that Error Prevention will always be a strategy with limited potential and therefore, Error Management is needed as a complementary strategy. Frese & Keith (2015) draw the conclusion that organizations have to embrace both Error Prevention and Error Management to some degree. However, organizations have to be aware that both strategies differ in the fundamental belief in how errors have to be managed and therefore it can be difficult to harmonize the two. This is especially important to consider because the strategy with which errors are managed will directly influence the organizational outcomes as well as the individual outcomes for the people involved.

Frese and Keith (2015) further explain the goals of Error Management, specifically; Controlling the potential damages errors may cause, Reducing the potential of one error leading to another and to act as a “secondary” Error Prevention.

1. Controlling the potential damage errors may cause: The first goal is related to how an organization should act in the presence of potential errors. In an organization where errors are hidden or not spoken about, the likelihood of errors triggering accidents is higher. Organizations often struggle with quick damage control of errors. For instance, blaming and punishing each other for errors is the prevalent reaction once an error occurs, ultimately hindering free discussions related to errors within the organization. The consequences of not having quick damage control procedures will most likely result in serious long-term problems. Even small errors that are not openly spoken about or reported, may prevent the necessary corrective actions to take place and managers will remain uninformed. These chains of events lead to lost learning opportunities and therefore damage control procedures need to be in place.

2. Reducing the potential of one error leading to another: According to the principles of Error Management, organizations should have a positive approach towards errors in the sense that errors are deemed inevitable, and employees should not be afraid to report it when they occur. If the message can be conveyed through the organization that errors are not something that one should shy away from but rather to regard them to be learning opportunities, employees are more likely to take a task-oriented approach towards errors, striving to fix the problems that the errors cause before they give rise to further errors.

3. Secondary Error Prevention: Learning from errors will directly influence Error Prevention. If the operators or users learn from errors, particularly why the errors occur, they will tend to prevent the error from happening again in the future. Moreover, it may also result in operators or users becoming more aware of errors, identifying areas where errors are more likely to take place and creating strategies to prevent them.

Reason (1997) emphasizes that errors should not be viewed as causes, but rather, as consequences. Therefore, focus should be directed to understanding the deeper fundamental reasons for the errors, regardless of who committed them. However, focus is oftentimes put on the entire organizational culture of errors. One of the most common actions in an organization is explained to be the blame cycle which refers to an individual receiving blame for committing an error. If the same error is committed by the same person, they are issued stronger warnings and sanctions.

However, Reason (1997) stresses the importance of understanding that everyone is capable of committing errors and that doing so does not necessarily indicate that the perpetrator is incapable. The situation could instead indicate that the system within which the person operates is inadequate in some way. In other words, it might not be the person's direct mistake since they were affected by the system, designed by the organization. Focusing too much on investigating who caused the error in order to find a culprit can therefore unnecessarily divert the attention from the detrimental activity of understanding the underlying reason for the error. In conclusion, one of the fundamental cornerstones for Error Management is that anyone can commit an error. Therefore, organizations have to break free from the blame cycle in order to fully

embrace Error Management. As a result, increased organizational learning becomes possible (Reason, 1997).

Wang & Hao (2017) argue that a positive error orientation can directly influence innovation on an organizational and individual level. It is argued that given that a positive attitude towards errors can be established and subsequently a well-functioning Error Management culture, innovativeness can be boosted. This is mainly made possible because an Error Management culture increases the confidence of employees to speak openly about errors and thereby find solutions and root causes to errors. This process ultimately promotes learning and exploration, directly influencing the organizational and the individual's innovative behavior. Wang & Hao (2017) concludes that innovation is indeed related to Error Management. Further support for this statement can be seen, one example being Fischer et al. (2018) who state that an Error Management culture will influence innovation positively. This is supported by their study of 30 companies. Furthermore, Hofmann and Frese (2017) state that embracing Error Management has shown evidence that indicates an increase in the firm's profitability.

It has been established that there is evidence supporting the positive correlation between Error Management and innovation. It might therefore be interesting to consider factors that influence Error Management in an organization. Hofmann and Frese (2017) brings forth three factors that may influence how well an organization may adopt a culture of Error Management. First, if individuals and groups do not believe errors may occur, they might show signs of denial when errors do actually occur. Therefore, groups or individuals have to understand that errors naturally accompany actions and will most likely eventually happen. Second, organizations should embrace open communication about errors. Therefore, once an error is identified the focus should be to spread the knowledge generated by the error, resulting in organizational learning and enhanced knowledge. Third, technical systems have to be designed in a transparent and intuitive manner to help individuals understand it. Thus, it is advantageous to assist the individuals in question in developing solid mental models of such systems.

3. Methodology

This chapter presents the methodology that has been used to gather the empirical data of this study. The chapter begins with a section outlining the chosen research design and the second section explains the practical steps taken to gather the empirical data.

3.1. Research design

Bell et al. (2019) explain that qualitative data allows researchers to gain insights into the personal perspectives of the people being studied. Quantitative data on the other hand is more structured and thereby allows researchers to investigate specific issues more rigorously they might be interested in. The case company has data of both types available, i.e., employees with relevant subject matter knowledge and relevant data within their internal error database. Therefore, it was deemed that in order to get a complete understanding of the issues being investigated, a mix between qualitative and quantitative research was needed. Furthermore, Bell et al. (2019) explains the concept of triangulation, which fundamentally refers to observing and analyzing an issue from multiple angles in order to gain more nuanced insights. In the context of mixing qualitative and quantitative research, Deacon et al. (1998) explain that triangulation can be applied to cross-check the findings and thereby increase the validity of the findings.

Based on the theory above, the research design of this study is mixed, involving both qualitative and quantitative research elements. In particular, the choice of research design is closely based on the "convergent parallel design" proposed by Creswell & Plano Clark (2011) of which an illustration can be seen in Figure 3.0 The convergent parallel design aims to benefit from the positive aspects of both qualitative and quantitative research whilst simultaneously mitigating the negative aspects of the two (Creswell & Plano Clark, 2011).

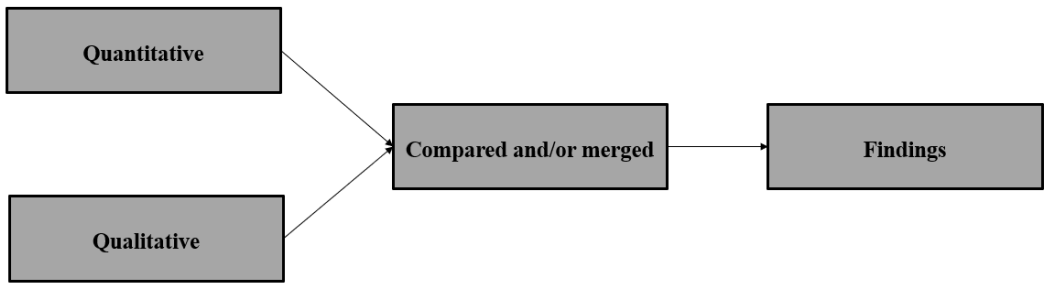

Figure 3.0 - An illustration of the Convergent parallel design by Creswell & Plano Clark (2011).

An illustration of the convergent parallel design applied to this study can be seen in Figure 3.1 The figure shows how the findings, stemming from the data collection, are compared to the theoretical framework in order to develop final managerial implications and recommendations. Similar to how the choice of mixed research design was motivated, the rationale for choosing the convergent parallel design is that it was

deemed most appropriate based on the types of data sources that were available within the case company.

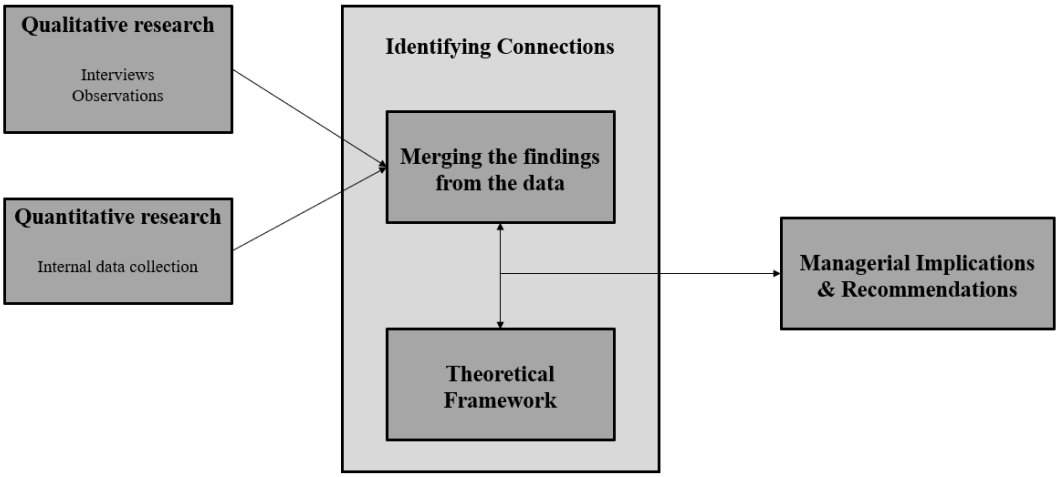

Figure 3.1 - The convergent parallel design in the context of this study.

3.2. Data collection

The data for this study was collected through semi-structured interviews, from the internal error database at the case company as well as observations whilst visiting the site of the case company and interacting with employees. The interview data together with the observations constitute the qualitative data whereas the quantitative data is gathered from the internal error database of the case company. More detailed descriptions of how these three data collection processes were performed will be presented next.

3.2.1. Interviews

Part of the qualitative data of this study was gathered through semi-structured interviews with employees from the case company, more specifically auditors. A sample of 15 auditors were carefully selected with the help of a senior director at the case company. Out of the 15 auditors, 14 accepted to be interviewed and 1 politely declined. Each auditor was asked for consent to record, to which 12 agreed. For recorded interviews, the recordings were used to verify statements that were fading in memory, to extract quotes and to identify emerging themes. For the non-recorded interviews, careful notes were taken by one of the authors.

Regarding sampling, Bell et al. (2019) explain that qualitative research often involves some variation of purposive sampling, meaning that the sample units are selected with the main selection criteria being to answer the research questions. Furthermore, fixed purposive sampling means that the sample is fixed from the outset without adding any further units at a later stage. A fixed purposive sampling approach was taken in this study where the interview guide focused strictly on generating answers to the research

questions. Furthermore, only the initial sample of auditors were interviewed. However, slight adjustments were made to the interview guide as more interviews had been conducted. The reason for this is that certain questions were deemed unclear by the auditors and therefore needed to be slightly rephrased. Furthermore, in some single cases, follow up questions had to be added in order to understand the underlying reason for the initial answer.

The interviews were divided into two parts, one semi-structured part and one part containing a case exercise. Bell et al. (2019) explain that semi-structured interviews involve asking the same overall questions to each interviewee but allowing for some degree of flexibility, in other words, giving the interviewees some leeway in their answers. This way, the interviews can take slightly different directions, but the core questions and structure remain the same. Bell et al. (2019) further argue that semi-structured interviews have the benefit of providing the researcher with valuable additional insight by including the human perspective to a greater degree, vs more structured interviews or surveys. Due to one research question being to investigate how the current error taxonomy works, it was deemed appropriate to let the auditors speak somewhat freely about their experiences, especially considering that they are the most experienced with using the taxonomy. The data from the semi-structured interviews was used to identify any themes regarding the use of the current error taxonomy. The semi-structured interview guide can be found in Appendix A.

The second part of the interviews was a case exercise where auditors were asked to use the current error taxonomy live to categorize three different error findings from the internal error database. When selecting these three findings, a sample of 100 errors were analyzed and narrowed down based on how long the finding description was. Since the auditors were only given roughly 10 minutes per finding, the finding description had to be reasonably brief. Once the findings had been narrowed down, an ex-auditor from within the case company was consulted to give advice on which three findings to pick for the case exercise. There was no formal hypothesis regarding the level of ambiguity of the findings in terms of the category they belonged to in the error taxonomy. However, at least one of the findings was deemed to be clearly ambiguous and it was chosen intentionally to make the exercise more exciting. During the exercise, auditors were asked to think out loud so that their thought process could be observed. Furthermore, auditors were not given any information about the findings beforehand. The data from the case exercise was compiled and used to gauge whether the auditors were using the error taxonomy in a similar manner.

3.2.2. Observations

The rest of the qualitative data of this study was gathered through observations whilst interacting with the case company. The observations mainly derived from formal digital meetings and informal interactions with employees whilst visiting the site of the case company. However, observations were also made whilst analyzing the internal error

data and whilst conducting the interviews. There was no formal plan or theory behind how observations were made other than a general alertness to note down anything that could potentially support the findings in answering the research questions. Observations are limited in the sense that they only originate from the subdivision of R&D with which the collaboration of this study has been.

3.2.3. Internal error-data collection

The quantitative data of this study was gathered from the internal error database at the case company. The goal was to randomly extract 100 error findings from the database to analyze how they have been categorized with the current error taxonomy. The case company has generated a lot of error-data, and therefore, delimitations had to be made. Firstly, only audit findings, in other words, errors that were found during audits were selected because those are deemed to be most critical to the case company. The total number of audit findings in the error database was several thousand, so further limitations had to be made. An associate director from the case company helped to apply relevant filters to appropriately narrow down the data to suit this study. One such limitation was to only look at findings between 2020-2021, the reason for this being that earlier findings were categorized with a different error taxonomy. The population was narrowed down to around 1193 of which a random sample of 100 was selected. The random number function in Excel was used to select the random sample which is a common practice according to Bell et al. (2019).

Bell et al. (2019) also explains that a key concept within quantitative research is generalizability, in other words the question of whether findings that derive from a sample can be extrapolated to also apply to the population from which the sample was taken. Therefore, the randomness of the 100 samples from the error database was double checked by comparing it to the original, filtered population of 1193 entries. The full comparison between the total population and the 100 random samples can be found in Appendix B. The unit of comparison was the percentage of entries in each subcategory. The average absolute difference across all subcategories was calculated between the population and the sample, resulting in 0.89%. This number is deemed to be low enough, meaning that the sample is a fair representation of the population. Therefore, the sample is deemed to be random and the results from the sample should be fairly generalizable. Furthermore, Figure 3.3 shows the frequency of findings within each subcategory among the 100 samples. In addition to this, all categories and subcategories have been renamed to Cx (Category and number) and Sx (Subcategory and number) at the request of the case company.

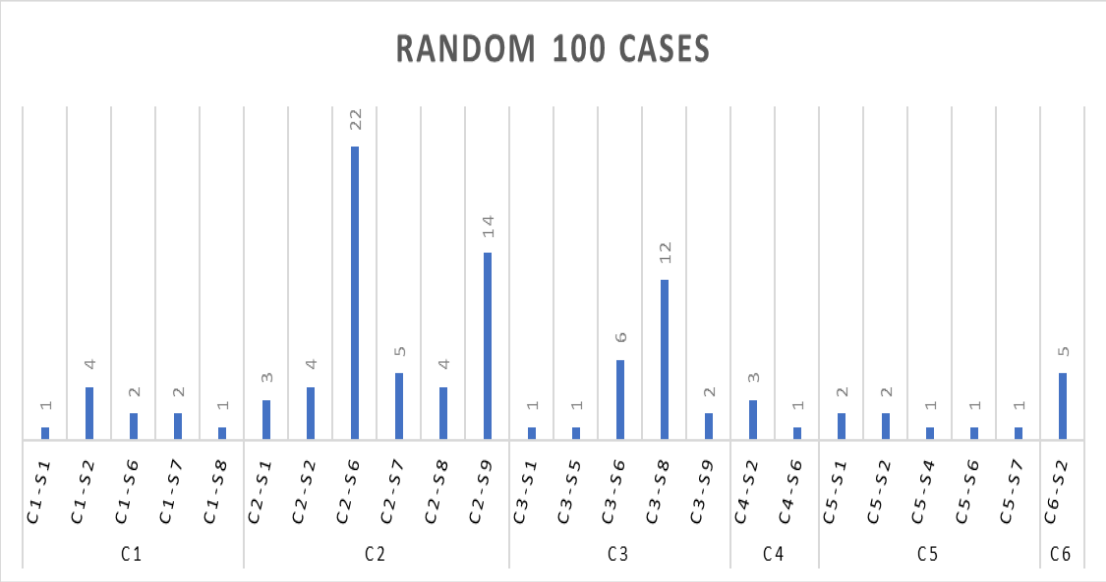

Figure 3.3, the 100 findings in their respective subcategory.

The main purpose of gathering the sample of 100 findings was to measure the number of recurrent findings, also known as recurrent errors, within the sample. In order to be able to do this, a general idea of what is considered to be a recurrent error was needed. The case company does not have a working definition for this as it is currently being investigated internally how it should be defined. Therefore, for the analysis of the 100 cases, the baseline is chosen that a recurrent finding is a finding that occurs at least twice in the sample. The question remains however, of how similar the findings need to be, in order to be considered recurrent. The case company made it clear that it may be difficult to identify recurrent errors solely based on the subcategories. Therefore, the aim was to find certain evaluation criteria when analyzing the findings. More specifically, criteria that could be used to compare findings to get a better approximation of error recurrence. In order to find these criteria, both authors thoroughly analyzed the 100 cases independently and then jointly discussed the potential evaluation criteria that could be used. The interview data was also used as a basis for finding and selecting evaluation criteria. Lastly, the case company was consulted to confirm that the selected evaluation criteria were appropriate.

4. Results

The results of this study are of both qualitative and quantitative nature. This chapter will initially present the qualitative data which derives from interviews and observations. Note that the interviews consisted of two parts, a semi-structured part, and a case exercise. The quantitative data derived from the internal error database of the case company.

4.1. Findings from semi-structured interviews

This section presents the data gathered from part one of the interviews, the semi-structured part. Whilst analyzing the interview notes and recordings, six themes emerged. These themes will be presented next along with quotes that capture the most significant aspects of each theme. As stated in the introductory paragraph of Chapter 4, observations were also made. These will primarily be presented in Section 4.3. However, certain observations were made specifically during the interviews, related to each theme, and therefore it is deemed more convenient to present these in the following sections.

4.1.1. Theme 1: Striking a balance between broad and narrow categories

Interview subjects have different opinions on where the optimal point on the spectrum lies regarding the balance between broad and narrow categories. Some subjects take a clear standpoint in the matter, such as:

> "*The categorization system is clearly too broad, especially certain subcategories, such as* C2-S6. *Many different types of errors get assigned that category and I suspect that it is oftentimes used when someone is uncertain of which subcategory to pick. It would therefore be better if the categorization system was more narrow, with more refined subcategories"*

Other subjects have a neutral standpoint, arguing for advantages and disadvantages with both extremes. As one subject stated:

> *"The problem with too broad categories is obvious, you simply do not know what managerial intervention to apply. However, too narrow categories are not a good option either, we have tried that in the past, with the previous system. There were too many subcategories, and it was difficult and cumbersome to use. The problem is that when you try to analyze the data, you see a forest that consists of tree stumps. In other words, you do not really identify any emerging, actionable trends."*

Observations related to this theme:

The previous system did not seem to be appreciated by any auditor that had used it, they deemed the new one to be better. A reason is that the categorization became even more tedious (several subjects deem it tedious even within the new, narrower system).

Striking a balance is clearly the goal to aim for, but it is difficult to achieve in practice. No subject had a solution in mind other than trial and error.

4.1.2. Theme 2: The role of GxP area in the error taxonomy

GxP area, essentially refers to "Good Practice" within a certain area, which is indicated by the" x". It is basically a system that divides the organization into different areas and outlines how work should be performed within each area. For example, "GMP", might refer to good manufacturing practice, which signifies that the context in question is manufacturing and outlines how the organization expects work to be performed within this context.

The error categorization system is designed to encompass every GxP area. In other words, findings from every context within the firm are categorized with the same system. However, some adaptations have been made. For example, once a certain GxP area is selected in the system, certain subcategories automatically become unavailable for selection, due to being irrelevant. Depending on which GxP area the auditor is specialized within, this can be more or less of an issue. One auditor from PV (Pharmacovigilance) stated:

> *"For me being a PV auditor, the categorization system can be confusing sometimes as I do not think any subcategory is a perfect fit to the context I'm operating in. I think the system is much better suited for GxP-1."*

Another initiative to distinguish between the contexts can be found in the guidance document that outlines how to categorize errors. In this document, examples are given on errors for each subcategory and GxP area. However, several subjects have stated that the examples are lacking. An example follows:

> "*The guidance document is a great tool as it gives examples for most GxP areas, but not all. There are however too few examples and for some subcategories they are missing altogether. Sometimes the subcategory is broad, and several types of errors can fit but there is only one example that does not give me enough information."*

One subject had worked at a different, much smaller, company where they had distinct categorization systems for every GxP area. The subject considered that to be considerably more effective due to the adaptation it made to each context, making it more intuitive to use. However, the subject acknowledged that it is a resource intensive practice to establish such a system, especially at a larger firm.

Observations related to this theme: Not having GxP specific taxonomies bothered certain auditors more than others. Specifically, within certain GxP areas, the generality of the error taxonomy is less of an issue. However, in other GxP areas the relevance was limited which clearly bothered those auditors.

4.1.3. Theme 3: Differences between auditors

Various differences between auditors that created variance in the results can be identified. There are a variety of tangible differences, such as the GxP area the auditor is specialized in, how much experience the auditor has, whether the auditor has experience with other categorization systems, if the auditor has worked at a different company, to name a few. These are examples of concrete, measurable differences between the auditors that impacted their answer to certain questions. However, the findings also indicate that there are more subtle, intangible differences between auditors such as their overall mindset related to their work. Most notably, differences in mindset showed differences in the answer to the question of how they think the categorization system could be improved. Certain subjects were less enthusiastic about this question, such as:

> *"I think the system is good and I don´t know how it could be improved."*

Whereas others were passionate about constantly improving and really played attention to the flaws of the categorization system, for example:

> *"The way we categorize errors right now does simply not result in actionable data. For starters, we need to make sure that every auditor interprets and reports a finding in a unified fashion, I am unfortunately not sure if this is currently the case. "*

Observations related to this theme: It is believed that certain auditors might have felt concerned to express their true, or complete, opinions on the subjects. These auditors gave notably shorter answers to each question and refrained from making any comments of anything being negative. They expressed that everything was working fine, and that they were happy with their jobs. These subjects also showed little interest in why proper categorization needs to be made or what the data will be used for. They just wanted to do their job, nothing less nothing more. The perhaps most problematic observation is that there is a high degree of variance in how auditors write their reports. Some follow certain guiding documents, but others do not. This makes it difficult to compare findings between different auditors. Furthermore, certain auditors expressed that they felt annoyed when encountering a finding that did not follow their own structure of reporting.

4.1.4. Theme 4: Uncertainty regarding the purpose of the error taxonomy

The purpose of the error categorization was often mentioned to be ambiguous. As mentioned in the previous theme, certain subjects were not very concerned with any matters that their job did not encompass. However, some subjects were notably bothered by the purpose being unclear. One subject stated:

> "*Management wants us to do this categorization but I must admit that I personally think it is the least important part of my job. In my opinion it is much more important to get the classification as well as the risk assessment right. Sometimes it feels like the categorization is just a tedious step that must be done but I am uncertain of the value it adds. This unfortunately results in not so much focus being devoted to make sure the choice of category is perfect. I believe that if I had a better idea of what the purpose was, I would be more motivated to put more time and effort into categorizing."*

Another subject had personal ideas of what the purpose should be but strongly believed that the categorization did not fulfill that purpose, they stated the following:

> *"The goal should be to exercise data driven decision making and that is why correct categorization is important. However, I do not believe we are there yet as I question what the data could actually be used for. We sort of just do our job without really being informed about what or how the data we generate is really being used. I believe that management needs to do a better job in defining and communicating the purpose of the error categorization order to make sure that actionable data is being collected."*

Observations related to this theme: Some auditors are clearly bothered by not having a clear purpose with the error taxonomy. Others acknowledge the ambiguity in the purpose but do not seem to care or think too much about it. They just want to do their job, as was also mentioned in the observations of the previous theme. It should also be mentioned that almost all auditors believe that the error taxonomy is easy to use in practice. The purpose might be unclear and there are other issues as well but the practical use itself was not mentioned to be an issue.

4.1.5. Theme 5: Lack of training in how to use the error taxonomy

Auditors were being asked what training they have received related to the error taxonomy and the answer was usually "none". They explained that they learned it by using it in practice and by consulting other auditors that have longer experience in using the system. Some subjects have shown concern about this as they believe it results in the system being used differently by different auditors, ultimately leading to a lack of unity. Simultaneously, as stated in the previous section, auditors believe that the error taxonomy is easy to use, which is contradictory to this finding. This issue is primarily apparent when analyzing audit finding reports. It is clear that there is variance in how

different auditors write the reports. Some auditors were outright frustrated about how some reports were written:

> "*I sometimes get frustrated when I see how differently auditors write their reports. There is a document that outlines how reports should be written but I can clearly tell that some do not follow the framework, which results in their report looking messy. It would be better if everyone followed the same operating procedures".*

The example shows how some subjects closely follow a document whereas others do not. A similar example is the guidance document for error categorization which was mentioned in Section 4.1.2. This document gives examples of errors that would fall under each subcategory and GxP area. However, the results show that half of the subjects that were interviewed did not know that this document exists. Perhaps even more concerningly, one subject was using an outdated categorization system. The subjects that are aware of the existence of the document use it extensively. As one subject quoted:

> *"The guidance document is an excellent tool and I use it all the time. It is not perfect but my job would be considerably more difficult if it did not exist."*

A subject that did not know it existed but was shown the document during the interview, stated:

> *"This sure looks like a very valuable resource and I had no idea it existed. I am perplexed that I have never even been informed about its existence. Please send it over to me"*

Observations related to his theme: Training material related to the error taxonomy clearly exists, but no auditor recalled being formally informed about the material or being encouraged to look them up. Those who knew it existed had either searched for it themselves or heard about it from a colleague. In other words, no coordinated effort was made to distribute the training material and make sure auditors had read the content.

4.1.6. Theme 6: Error recurrence identified during audits

The final theme emerged when subjects were being asked about their thoughts on error recurrence. Every subject stated that recurrence is identified based on memory of previous audit findings or by consulting with colleagues. No system-wide metric or other type of indicator exists currently to notify the auditor that a finding is recurrent. One subject stated:

> "*We simply know, based on our recollection of previous audits if a certain finding has been seen before. Of course we also talk to each other and read through the files from previous audits before revisiting a site. It is an integral*

part of our job to identify when a finding from a previous audit has not been fixed because then we need to raise the classification of the finding to major. This needs to be done regardless of how minor the original finding was, which might seem a little counterintuitive."

Observations: It should be noted that the part of the interview guide that investigated error recurrence was adjusted to be more elaborate, in preparation for the last interview. Therefore, limited data was collected prior. However, the last subject provided an important insight, showing how it is documented manually when a finding is recurrent. Auditors write a remark in the finding description that the same finding was identified in a previous audit. This way of indicating a recurrent finding does not seem to be a properly articulated standard and it is easy to miss when reading the finding description, especially considering that auditors have different styles of writing. It would perhaps be better if there was an option in the system to select that the finding is recurrent.

4.2. Findings from interview cases

The second part of the interviews consisted of a case-exercise where subjects were asked to categorize 3 findings from the error database live. These three findings were chosen with the assistance of an ex-auditor at the firm. The subjects had approximately 10 minutes to categorize each finding into the subcategory(s) they believed to be the best fit. The aim of the case-exercise was to capture the subjects' thought processes of how they were using the error taxonomy in practice. Note that certain subjects selected a number of categories that could potentially be a fit, without ultimately deciding which of these was the most appropriate.

Table 4.0 displays each category and subcategory that were mentioned for each case. The numbers in the brackets indicate how many auditors mentioned that particular subcategory. Note that the numbers in brackets for each case do not add up to 14 because each auditor mentioning a subcategory counts as one, but certain auditors mentioned more than one subcategory as a potential fit. Furthermore, a subcategory was not counted if a subject mentioned it but later changed their mind. However, a subcategory was counted if the subject mentioned the category but was uncertain. Finally, the subcategory marked in green is the "correct" category that has been logged into the error database.

Case 1	*Case 2*	*Case 3*
• C3-S6 (9) • C3-S4 (7) • C2-S6 (4) • C2-S8 (1) • C1-S5 (1)	• C3-S1 (13) • C2-S1 (2) • C2-S6 (2) • C3-S6 (1) • C2-S7 (1) • C5-S6 (1) • C5-S1 (1)	• C3-S9 (14) • C1-S5 (2) • C2-S6 (2) • C3-S1 (1) • C3-S8 (1) • C5-S5 (1) • C2-S7 (1) • C1-S7 (1) • C3-S6 (1)

Table 4.0, The selected subcategories during the case-exercise.

4.2.1. Case 1 summary

This case essentially involves a standard operating procedure (SOP) not being followed precisely. Specifically, the documentation of a non-conformance activity did not happen according to the SOP. The "correct" subcategory for case 1 is C2-S6. The notion of "correct" henceforth refers to the subcategory that is assigned to the finding in the error database. This does not necessarily indicate that the assigned "correct" subcategory within the database is universally true as the auditor who selected the subcategory could have made a mistake. The GxP area of this finding is GxP-5.

Case 1 proved to be the most ambiguous of the three as indicated by only 4 out of 14 subjects selecting the "correct" category. Furthermore, most subjects were advocating for either C3-S6 or C3-S4 rather than C2-S6. This somewhat depends on how the auditor approaches the finding, most auditors automatically select S6- or S4 as soon as

the finding involves not following an SOP correctly. Those that mentioned C2-S6 are not entirely sure but state that the category is quite broad and therefore the finding can likely fit there. Finally, some subjects strongly disagree with the correct category, stating that C2-S6 is completely irrelevant in the GxP-5 context.

4.2.2. Case 2 summary

This case essentially involves a quality agreement, between the firm and a partnering firm, not being updated and executed. The correct subcategory for case 2 is C3-S1. The GxP area of this finding is “GxP-1”.

Case 2 seems to be considerably less ambiguous as 13 out of 14 subjects selected it with a high degree of confidence. It can also be seen that other categories are only mentioned once or twice and this is a result of the subjects being actively asked if some other category could be a fit as well. The argument as to why C3-S1 is the correct category is usually that the finding directly involves an agreement. It should be noted that several subjects mentioned that the report of this finding is quite messy, although this did not affect their choice of subcategory.

4.2.3. Case 3 summary

This case essentially involves guidelines, for handling sensitive data, being too generic and unclear. The correct subcategory for case 3 is C3-S9 under the Organization category (There is a subcategory with the same name under the C4 category). The GxP area of this finding is “GxP-1”.

Case 3 is also deemed unambiguous as every single subject selected the correct subcategory. This was most commonly motivated by no other subcategory being relevant at all, making C3-S9 an easy pick. The auditors mentioning the other potential subcategories were highly uncertain. However, some did mention that no category was a good fit for this finding and argued that there should be a category for “Data Management”.

4.3. Findings from observations

The rest of the qualitative data was gathered through observations. Some of these have already been presented, specifically those that were gathered during the semi-structured interviews. This section will present the remaining observations that were made throughout this study. As mentioned in Section 3.2.2., in the methodology chapter, no formal plan was used when making observations. The observations are simply based on any form of interaction with the case company where relevant information to the study was encountered such as during informal conversations with employees, or team meetings. Furthermore, informal meetings were held early in the project in order to get

to know each team member of the QI department. The most significant observations have been grouped into subsections and will be presented next.

4.3.1. Observation 1 - Minor findings

This observation derives from a discussion that took place during a virtual meeting as well as inspection of the internal error database. One important task of an auditor is to assign a classification to a finding which essentially denotes the severity of the error that has been committed. When looking at the internal error database, the vast majority of findings are classified as minor, which is the lowest degree of severity. Major and critical are the higher degree labels and any finding that has these labels undergo certain procedures to address the underlying problem. These said procedures consist of a root cause analysis, corrective actions and a plan for preventive actions to stop the same error from happening again in the future. Minor findings on the other hand, are left alone, with no further procedures taking place.

Concern has been raised internally that much potential knowledge is lost by not investigating the minor findings and that there might be serious consequences if they are allowed to brew up into more serious errors. However, there is no simple solution to this problem as the remedial procedures are very resource intensive and even the limitation of just investigating major and critical findings further is deemed to be too much by certain employees of the case company. The reason this observation is significant for this study lies in the process of upgrading the classification of a minor finding to a higher level of severity. The criteria for the upgrade process are that if a finding that was classified as minor appears again in a subsequent audit, indicating that it has not been fixed, it has to be upgraded to major or critical. This is similar to what Cahill (2017) calls a repeat finding, referring to a finding found at a subsequent audit that has not been fixed. Such findings are essentially deemed to be recurrent findings by the auditors, which leads into the next observation.

4.3.2. Observation 2 - The difficulty of defining error recurrence

This observation derives primarily from discussions with the senior director from the case company who also supervised this project. It was explained that error recurrence is very difficult to define, foremost because it is not arbitrary which evaluation criteria to apply when assessing whether two findings are the same. For example, would forgetting to sign some type of document be considered a recurrent error even if it is committed by different people, perhaps in different countries? The answer is not straight-forward, and it is a difficult balancing act because including too many details might not be viable whereas too few details might not capture important aspects of the error.

Defining recurrence is only a first step however with the overall goal being to be able to measure error recurrence formally and rigorously in order to identify trends and

direct managerial interventions. In order to reach that stage, the data that is derived from the error taxonomy must be actionable which directly depends on how recurrence is ultimately defined. Furthermore, in order for a definition of error recurrence to be chosen and used in practice, the next level of the managerial hierarchy must approve the definition. The error taxonomy is therefore designed in a top-down fashion and there are strict criteria for suggestions to be rigorously supported, for example by academic research.

4.3.3. Observation 3 - Discrepancy between how management and auditors view error recurrence

This observation derives from comparing observation 2 with theme 6 that emerged from the semi-structured interviews, presented in Section 4.1.6. Observation 2 outlined how the design of the error taxonomy is a top-down process with upper levels of management being closely involved. The definition of error recurrence is one example and there has to be strict theoretical support for management to pass a definition. However, a finding from the interviews is that the users of the system, such as the auditors, are not involved in the process of designing the error taxonomy. Theme 6 in section 4.1.6 showed that auditors have an intuitive understanding of what error recurrence is. Auditors simply know based on experience, memory, or consultation with colleagues when a finding is recurrent, without having to think too much about some detailed definition.

Further, it is even reported sometimes when a finding is recurrent, but this happens informally by simply adding a note in the finding description. The important aspect of this observation is that there is a high degree of discrepancy between how management views error recurrence and how the users of the system, in this case auditors, view error recurrence. Not involving the users of the system in designing the system is a likely explanation for this discrepancy. The experiences of the auditors might indicate that management is somewhat overthinking the definition of recurrence and substantial theorizing around the concept, without taking input from the users might disconnect theory and practice even further. This is a potential danger for managers to be aware of as it resembles the notion of organizational silos as explained by Cilliers & Greyvenstein (2012).

4.3.4. Observation 4 - Concern that non-auditors might struggle to use the error taxonomy correctly

This observation derives from the case exercise of the interviews as well as discussions with the supervisor. The case exercise, in particular case 1, showed difficulty in using the error taxonomy considering that the opinions of auditors varied considerably. Only 3 cases were evaluated but it is reasonable to believe that there are more cases that would prove difficult. As a reminder, this project only examined audit findings, but

there is a much larger pool of error findings that are not reported by auditors, but rather, by personnel from all parts of the organization. Concern has been raised that if the auditors, who are supposed to be the most prevalent in using the error taxonomy, have difficulty using it, non-auditors are likely to have an even more challenging time.

4.3.5. Observation 5 - A note on blame culture

This observation derives from discussions with the project supervisor, various other employees, and visits to the site of the case company. According to the project supervisor the case company is doing an overall good job in avoiding the emergence of a blame culture. This is partly because the company is striving to be a great place to work, and employees are encouraged to speak their mind and create a positive atmosphere. Visits to the site of the case company confirm that the atmosphere is indeed very good, and the employees are easy going, friendly and highly respectful towards one another. Furthermore, during interactions with employees it is often stated that they love working at the case company because the culture is so decent.

However, as stated by the supervision of the project, fear of blame can never be fully eliminated, especially in the context of errors. It is natural to suspect that some percentage of underreporting occurs due to the fear of blame. Furthermore, it has been brought up by employees that they sometimes find reporting errors tedious and simply ignore doing it. However, during the semi-structured interviews, it became clear that some auditors are annoyed that certain auditors write their audit finding reports poorly. There clearly is variation in how auditors write their reports as mentioned in Section 4.1.3. and this might raise concern because a lack of unity in reporting can lead to problems when compiling and analyzing the data.

4.4. Findings from the internal error data

The 100 cases were initially read through thoroughly in order to get a sense of the number of recurrent findings at face value. This is because reading the finding descriptions gives an intuitive sense of the degree of similarity between findings. The most significant result is that, at face value, no two findings are exactly identical within the sample. An overview of the 100 cases can be seen in Appendix C.

However, many findings are in fact similar, to varying degree. Therefore, the question becomes whether error recurrence can be better approximated by categorizing findings based on the degree of similarity between them. One way to do this is by selecting evaluation criteria that must match between findings in order for them to be regarded as recurrent. There are many different potential evaluation criteria that could be selected. Based on relevance, data availability, interview findings, and consultation with the case company, three evaluation criteria were selected. These criteria are the Subcategory, the GxP area and the SOP, of which the former two have already been introduced.

The SOP, "Standard Operating Procedure", refers to the specific part of an SOP that has been violated by the discovered error. This is not a formal variable that is reported for each finding whilst using the error taxonomy. Instead, it is expected to include it in the finding description. However, sometimes it is not applicable, forgotten, or simply omitted for other reasons, as it is not a strict requirement to report it.

Next, the selected evaluation criteria were applied in layers, ontop of eachother, forming three layers. For each layer, a better approximation of error recurrence can be measured because each evaluation criterion requires a higher degree of similarity between findings. Continuous feedback from several stakeholders of this study has indicated that the purpose of this layering process is ambiguous and difficult to comprehend. To clarify, the purpose is simply to illustrate a methodology that can be used to find a better approximation of error recurrence. This is done by applying several criteria that have to match between findings and recategorizing them based on these criteria. This is done to gradually increase the similarity between findings and therefore also gradually improve the approximation. As opposed to the current way of working, which is to attempt to precisely define error recurrence from a top-down perspective, the presented methodology takes a bottom-up approach. The idea is to circumvent the obstacles that the top-down approach presented yet still gain a good enough approximation of error recurrence patterns by approaching it with the proposed bottom-up methodology.

Detailed explanations of each layer will be presented next.

Layer 1 - Subcategory

The evaluation criterion of the first layer is chosen to be the subcategories of the current error taxonomy. The current error taxonomy has two levels of abstraction, namely main categories, and subcategories, which need to be assigned to each finding. Therefore, the subcategory is the most fine-grained level of the current error taxonomy, and findings within the same subcategory should naturally share some degree of resemblance. It is therefore deemed that the subcategory is a relevant evaluation criterion.

For layer 1, all findings within a certain subcategory are considered to be sufficiently similar and hence an approximation of recurrence. Within the 100 samples, this involves excluding any finding that is the sole finding within a subcategory. For example, if there is only one finding in the C3-S1 subcategory, it is deemed unique, rather than recurrent, and therefore excluded.

Layer 2 - Subcategory + GxP area

The second layer adds the GxP area as an evaluation criterion in addition to the criterion of layer 1.

If a finding is reported to be from a certain GxP area, it means that the error was discovered within that context of the organization and is related to some activity or procedure that takes place within that context. Two different GxP areas, or contexts, might be very different and therefore, comparing findings within different GxP areas can sometimes be problematic. An informal, yet effective analogy would be to attempt to compare apples and oranges. For this reason, the GxP area is deemed to be a rather obvious choice of evaluation criteria.

For layer 2, all findings within the same GxP area as well as the same subcategory are considered sufficiently similar and hence an approximation of recurrence. Within the 100 samples, this involves excluding any finding that is the sole finding within the same GxP area and subcategory. For example, if there is only one finding from GMP in the C3-S1, it is deemed unique, rather than recurrent, and therefore excluded.

Layer 3 - Subcategories + GxP area + SOP

The third layer adds the SOP as an evaluation criterion in addition to the evaluation criteria of layer 2. If two findings violate the same SOP, they must have a quite high degree of similarity by nature, as the SOPs are quite specific. For this reason, it is not likely to find two findings violating the same SOP to be substantially different. Therefore, the SOP is deemed to be a highly relevant evaluation criterion.

For layer 3, all findings that violate the same SOP, are within the same GxP area as well as in the same subcategory, are considered sufficiently similar and hence an approximation of recurrence. Within the 100 samples, this involves excluding any

leftover finding from layer 2 that is the sole violator of a certain SOP. For example, if there is only one leftover finding from layer 2 violating SOP section 4.1.5, it is deemed unique, rather than recurrent, and therefore excluded.

It should be noted that many other potential evaluation criteria could have been chosen instead, to do a similar analysis. Other than what has already been specified for each criterion, an important reason for the choices is that these were data points that were readily available. Furthermore, the case company was consulted, and all parties agreed that these three criteria were relevant considering the scope of the project. With that established, it is believed that, if necessary, an even higher degree of similarity can be found between findings by applying even more criteria.

The findings from the layering process will be presented next, in particular how many recurrent findings there are among the 100 samples when categorized according to the aforementioned three layers.

Layer 1 - Subcategory

The approximation of error recurrence in the sample is **92 out of 100** when applying this evaluation criterion.

This means that every finding within a certain subcategory is deemed to be the same and therefore as soon as there are 2 or more findings within a category, all are considered recurrent. Findings that are the sole finding within a subcategory are discarded. The number of findings fulfilling the criteria can be seen in Figure 4.1. which only displays categories containing at least two findings.

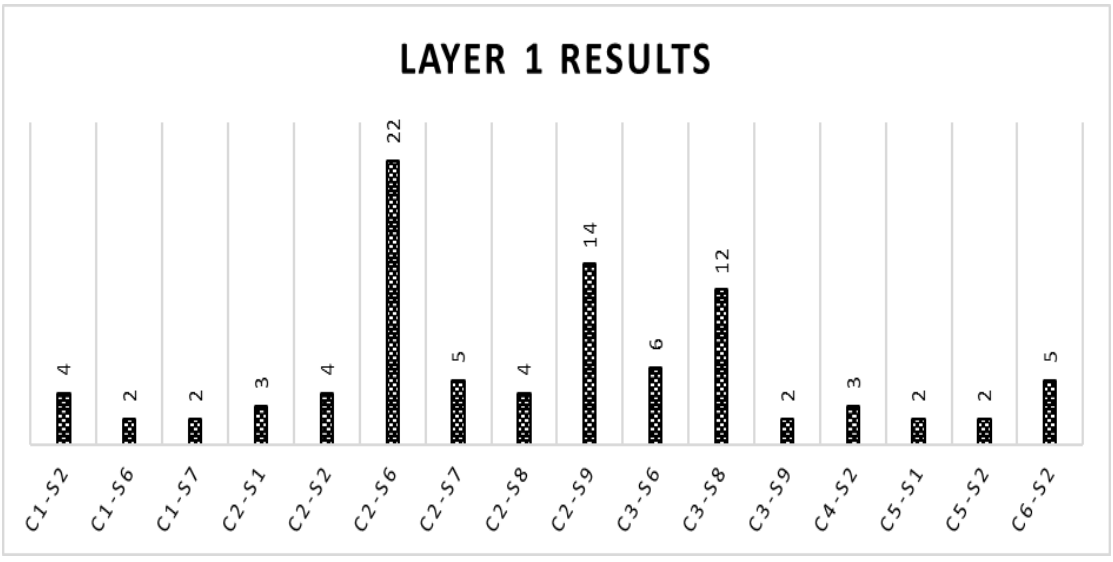

Figure 4.1, Layer 1 results.

Layer 2 - Subcategories + GxP area

The approximation of error recurrence in the sample is **80 out of 100** when applying these evaluation criteria.

Not surprisingly, adding a second criterion results in a lower approximation. In order for a finding to pass the layer two evaluation criteria and be included in the count, there has to be at least one more finding within the same subcategory and the same GxP area. Findings that do not fulfill the criteria are discarded and not included in Figure 4.2.

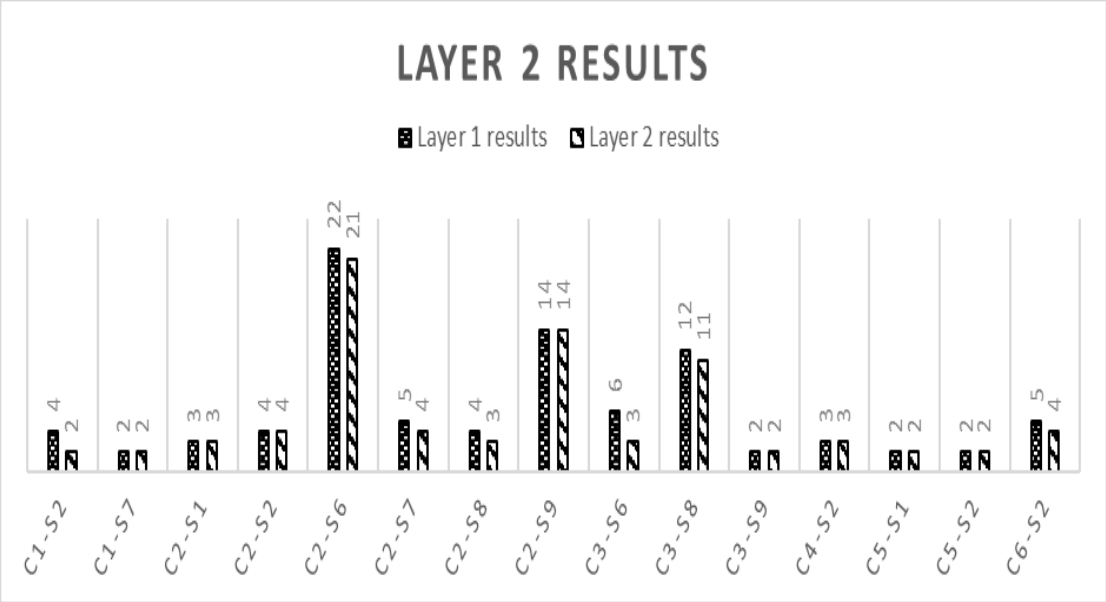

Figure 4.2, Layer 2 results.

Layer 3 - Subcategories + GxP area + SOP guideline

The approximation of error recurrence in the sample is **36 out of 100** when applying these evaluation criteria.

This difference, compared to the recurrence measurement of 80, from layer 2, is significant. This is mainly because certain subcategories are broad, even when looking at findings within the same GxP area. The SOPs on the other hand, track back to specific violations of an established SOP. Therefore, findings that violate the same SOP are considerably more similar. Figure 4.3 summarizes the results from all 3 layers but only shows subcategories that contain findings passing the layer 3 criteria. In order for a finding to pass the layer 3 evaluation criteria, it must pass the criteria of layer 2 and at least one more finding must violate the same SOP. Note that certain findings violate several SOP. For the evaluation criteria to be fulfilled however, it is considered enough if at least one SOP is the same between two findings. Figure 4.4 shows the breakdown of the layer 3 results into the specific SOPs that were violated and their frequency.

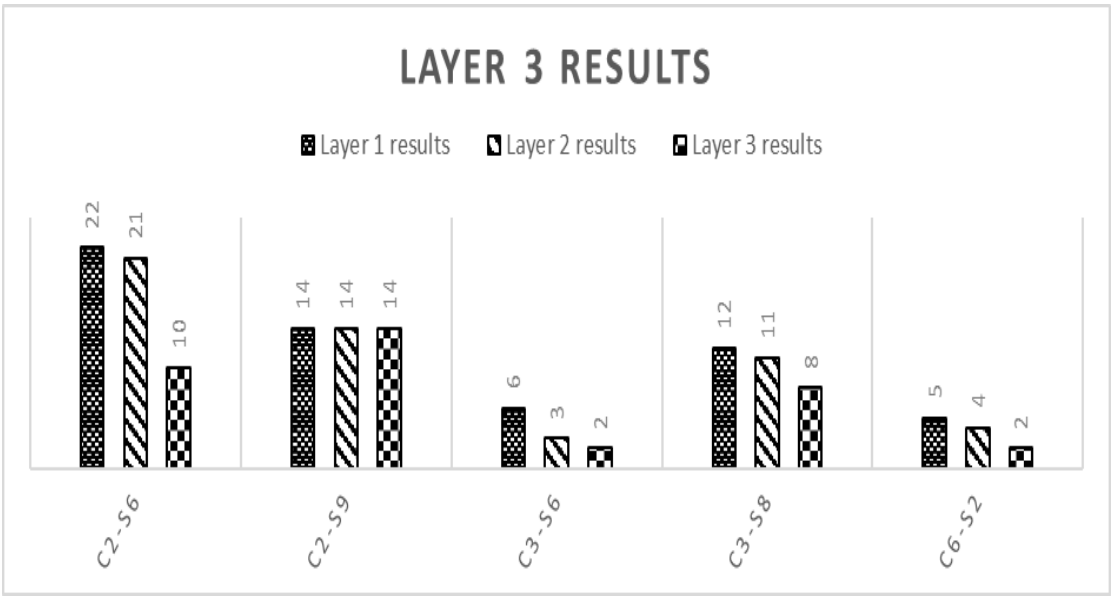

Figure 4.3, Layer 1-3 results.

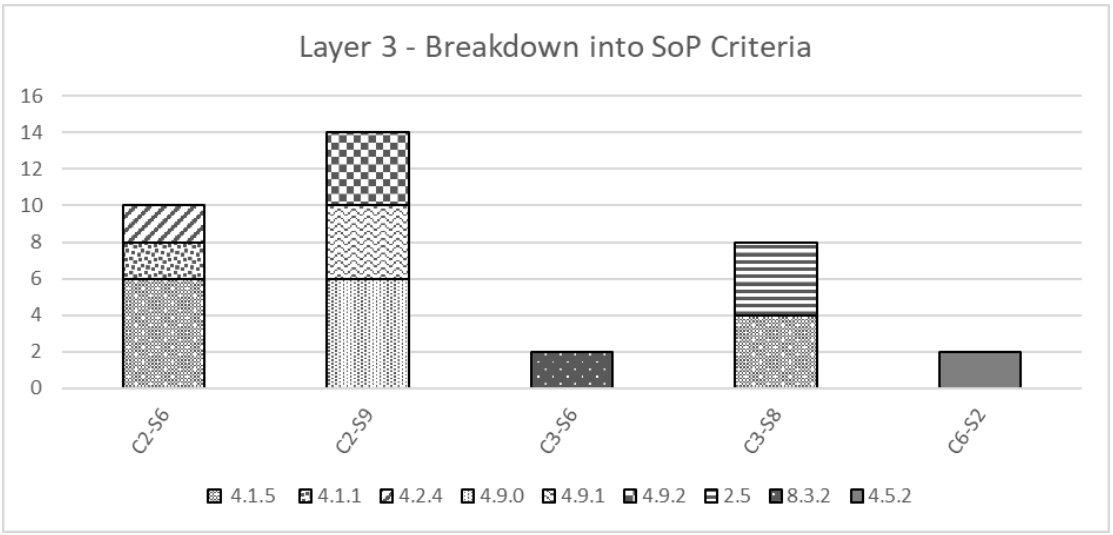

Figure 4.4, Specific SOP violations of layer 3 findings.

5. Analysis and discussion

This chapter is divided into three sections. The first section presents the key connections that can be made between the literature presented in the theoretical framework and the empirical finding. The second section evaluates the current error taxonomy of the case company. The third section proposes a modification to the current error taxonomy and provides evidence that supports the proposition. These three sections lay the foundation for the recommendations that will be presented in the next chapter.

5.1. Connections between empirical data and literature

This section presents the key connections that can be made between the empirical findings of this study and the literature presented in the theoretical framework. The connections are related to three identified contexts, “error patterns and trends”, “error culture”, and “purpose of error categorization”. In order to avoid a high degree of repetitiveness, these connections will be presented in Table 5.0 rather than in a continuous text format. Table 5.0. highlights the relevant context, the parts of the literature and empirical data between which connections can be made, and a description of each connection. All connections relate to the error taxonomy in some way and aim to assist in answering the research questions of this study.

Context	Literature	Empirical data	Connection
Error patterns and trends	**Section 2.1.** - Love et al. (2013): The concept of error traps. **Section 2.1.** - Reason & Hobbs (2003): Looking for patterns to identify error traps.	**Section 4.1.1.** - Theme 1: Striking a balance between broad and narrow categories. **Section 4.4.** - Findings from the internal error data: Insights from the layering process.	The discussion in Section 4.1.1. acknowledges the difficulty in balancing broad vs narrow categories. If categories are narrow, it risks being perceived as difficult and tedious to use. If categories are too broad, variance might be too high between findings in certain subcategories. In both cases, identifying patterns and thereby error traps might become more difficult. The layering process data suggests that comparing SOPs is an effective way to pinpoint errors that are similar and should therefore be an effective measure to detect error patterns and thereby error taps.
Error Culture	**Section 2.3.** - Wang & Hao (2017): Positive correlation between error culture and innovativeness. In other words, using errors as learning opportunities boosts innovation potential.	**Section 4.3.1.** - Observation 1 - Minor findings.	The literature explains how learning from errors boosts the innovativeness of the organization and that not doing so might result in lost opportunities. This might be the case to some extent because minor findings are not followed up. However, it should be underlined that this issue is complex to solve due to resource constraints. It is believed that improvements in the error taxonomy might help in generating more actionable data, which can be used to efficiently gain insights from patterns within the minor findings as well.

	Section 2.3. - Fischer et al. (2018): States the same as Wang & Hao (2017), used to further support their standpoint.		
Error Culture	**Section 2.3.** - Reason (1997): Describes the negative consequences blame cycles can have on organizations. In particular, it leads to learning opportunities being lost and bad emotions among employees.	**Section 4.3.5.** - Observation 5 - A note on blame culture.	It should be stressed that the case company is not deemed to have a blame culture. On the contrary, they are taking several measures to be considered a great place to work and encourage openness and collaboration. However, the fear of blame can never be fully eliminated, especially when it comes to the subject of errors. It is therefore likely that a percentage of errors are not reported.
Error Culture	**Section 2.1** - Walia & Carver (2012):	**Section 4.2.1.** - Case 1 summary.	There are several indicators in the empirical evidence suggesting that too little training is undertaken by auditors in how to use the error taxonomy. They learn it by simply using it and get advice along the way from more experienced

	Stressing the importance of providing proper training to those that are expected to use the error taxonomy.	**Section 4.1.5. -** Theme 5: Lack of training in how to use the error taxonomy. **Section 4.1.3**.- Theme 3: Differences between auditors.	auditors. One the one hand this might suggest that there is no need for more sophisticated training. On the other hand, the empirical data of this study might suggest otherwise. Auditors use the error taxonomy differently which also creates variance in the data being produced. This in turn is believed to make the data less standardized and therefore more difficult to analyze.
Error Culture	**Section 2.1. -** Tamuz et al. (2004): The importance of sharing information between departments and having a unified view of important concepts. **Section 2.1.** - Cilliers & Greyvenstein (2012):	**Section 4.1.5. -** Theme 5: Lack of training in how to use the error taxonomy. **Section 4.1.6. -** Theme 6: Error recurrence identified during audits. **Section 4.1.4. -** Theme 4: uncertainty regarding the	Several pieces of empirical evidence suggests that there is no unified consensus in how to use the error taxonomy between auditors. For example, this is an obvious consequence of inconsistencies related to the purpose of the error taxonomy and no formal training. This is further indicated by the variance in the use of the error taxonomy and the occurrence of non-standard practices such as how recurrence is currently reported. All taken into consideration two fundamental weak points are implied. The first relates to communication between management and auditors and also between auditors. The second issue derives from the first and suggests that a consensus has not been established. This is very similar to the organizational phenomenon known as organizational silos.

	The concept of organizational silos.	purpose of the error taxonomy. **Section 4.1.3**.- Theme 3: Differences between auditors.	
Purpose of the Error Taxonomy	**Section 2.1:** Walia & Carver (2012): An error taxonomy can help inspectors identify and understand errors and thereby generally increase their knowledge about the system. **Section 2.1:** Stanton & Salmon (2009): To understand and gain a deeper understanding of errors.	**Section 4.1.4. -** Theme 4: uncertainty regarding the purpose of the error taxonomy. **Section 4.1.4. -** Observations related to this theme. **Section 4.1.3**.- Theme 3: Differences between auditors. **Section 4.2.1. -** Case 1 summary.	There is a strong consensus in the literature that it is important to have a clear purpose with the error taxonomy. Purpose suggestions by different authors are compiled to the left. The purpose of the error taxonomy at the case company, however, is highly ambiguous at least as perceived by auditors. The empirical data suggest that management has perhaps not been able to effectively communicate this purpose to auditors which is seen to be a problem as it hurts the motivation of certain auditors, making them reluctant to put a lot of time and energy into using the error taxonomy. This is also likely part of the explanation to the variance between how auditors write their reports. It is clearly observable that this sense of lacking purpose is confusing and affects the error data that is being generated as seen in the case exercise. This is a legitimate reason to question the legitimacy of the error data.

	Section 2.1: Zhang et al. (2004): To understand how humans interact with the system within which they operate and identify strengths and weaknesses. **Section 2.1:** Hooper & O´Hare (2013): To understand why errors occur and prevent them from happening again.		

Table 5.0 - Key connections between the literature and the empirical findings of this study.

5.2. Evaluating the current error taxonomy

In this section, the current error taxonomy of the case company will be evaluated. The evaluation is based on the empirical findings and the error taxonomy evaluation framework by Wiegmann & Shappell (2001). This framework uses five criteria to evaluate an error taxonomy, refer back to the end of Section 2.3 for details.

5.2.1. Reliability

Reliability fundamentally concerns being able to trust the data that is being generated by the error taxonomy. From the outset this notion is problematic as one main reason that this study is being conducted is that the case company does not fully trust the data, which has clearly been articulated by several representatives from the company. Furthermore, evidence from the interviews suggest that there are problems with reliability such as the results from Case 1, see Table 4.0. The results from Sections 4.1.5. and 4.1.3. also raise concerns regarding reliability since it is indicated that no coordinated training is provided in how to use the error taxonomy and that auditors use it differently.

A second important aspect to ensure validity is for the error taxonomy to be clearly context specific. This is arguably not the case at the case company, which is supported by Section 4.1.2. where it is established that the error taxonomy is general for all GxP areas. Certain auditors clearly raised concern that this creates problems in certain contexts in terms of the error taxonomy not being properly applicable.

In conclusion, there is clear evidence suggesting that there are several reasons to be concerned regarding the reliability of the current error taxonomy.

5.2.2. Comprehensiveness

Comprehensiveness fundamentally concerns the ability of the error taxonomy to provide all the necessary complementary data related to an error.

The main problem is that a measurement of error recurrence cannot be deduced from the current data and hence it can be argued that the most important complementary data is not displayed. Section 4.3.2 discusses that there is no definition of error recurrence which obviously complicates the issue. However, as illustrated in the layering process of Section 4.4, a relatively high degree of similarity, and hence recurrence, between errors can be seen when comparing SOP violations. However, this data-point is not formally stored in the error taxonomy but is instead sometimes noted in the finding description. Lack of formalization in how to report this important metric arguably hurts

the comprehensiveness of the error taxonomy. Further support for this derives from Sections 4.1.6 and 4.3.3 where it is seen that the understanding of error recurrence is highly informal and can vary between auditors and managers. Specifically, the notion of recurrent findings is primarily memory-based and only certain auditors report recurrence in the finding descriptions informally.

Another problem that arises is that different contexts, in this case different GxP areas, might call for different contextual information to be displayed. As can be seen in Section 4.1.2., having a general error taxonomy for all GxP areas creates problems in terms of relevance within certain contexts. This has implications on the data that the auditor is forced to report as they need to force-fit their findings into the more general context of the error taxonomy.

Furthermore, the strong uncertainty regarding the purpose of using the taxonomy, as displayed in Section 4.1.4., particularly that certain auditors feel unmotivated to use the taxonomy, should raise concern regarding the data that is being reported. For this reason, auditors find it tedious to report all the expected complementary data, which should raise concern about the comprehensiveness of the data.

In conclusion, there is clear evidence suggesting that there are several reasons to be concerned regarding the comprehensiveness of the current error taxonomy.

5.2.3. Diagnosticity

Diagnosticity imposes two main criteria onto the error taxonomy. Firstly, it should generate data that shows the root cause of the error. Secondly, the data should be of high enough quality to autonomously guide decision making, without the need to rely on the intuition and experience of senior managers.

The first criteria concern root cause which is outside of the scope of this study. The case company does apply root cause analysis to certain findings, but this practice has not been investigated and therefore it is impossible for this study to comment on how well it fulfills its purpose.

However, evidence has been found that contradicts the second criteria, which can be seen in Section 4.1.3. Certain auditors clearly state that they believe that there is a long way to go to reach such a level of data driven decision making. This point is further reinforced by the discussion in Section 5.1.1., which shows concerns regarding the reliability of the error taxonomy data.

In conclusion, there is clear evidence that shows concern regarding the second criteria of diagnosticity. This is deemed to be enough reason to question the diagnosticity of the current error taxonomy.

5.2.4. Useability

Useability requires the error taxonomy to be easy to use. As mentioned in the observations of Section 4.1.4, the practical use of the taxonomy is not deemed to be difficult but that does not mean it is deemed to be problem-free.

It is difficult to disprove the useability of the error taxonomy, but evidence suggests that it might not be used correctly. The most illustrative example is the result from Case 1 in Section 4.2.1., which arguably shows a rather chaotic spread between how the finding is interpreted. This indicates that there is reason to believe that there is lacking consensus in how to use the error taxonomy.

Further concerns are indicated in Sections 4.1.3, 4.1.4. and 4.1.5. Auditors come from different professional backgrounds and contexts and do not receive formal training on how to use the system which might explain the variance in how the taxonomy was used in Case 1. This might be further explained by the sense of lacking purpose. Furthermore, as stated in 4.1.5. not all auditors follow, or even know about, the guidelines for writing finding-reports. Some do not know about the existence of the error taxonomy guide.

Whilst the error taxonomy is deemed easy to use by auditors, there is clear evidence suggesting that they use it differently. The findings provide several plausible explanations for this variation. The existence of guidance documents implies that there is a correct, or at least recommended, way to use the taxonomy. However, since several auditors do not even know about the existence of these documents, it should be questioned whether a proper standard of how to use the error taxonomy is established and communicated properly.

In conclusion, due to the high variation in its use, the useability of the error taxonomy can arguably be questioned. However, the evidence suggests that the underlying reason is not necessarily because it is difficult to use in practice but instead due to ambiguities related to how it should be used.

5.2.5. Validity

Validity fundamentally refers to how well the error taxonomy fulfills its purpose. Weighing together all previous criteria, is argued by Wiegmann and Shappell (2001) to give a good indication of validity. For each previous four criteria, the evidence suggests considerable weaknesses and therefore the validity of the current error taxonomy should be considered low based on the framework by Wiegmann and Shappell (2001).

Furthermore, it would be highly questionable to claim that the error taxonomy fulfills its purpose when there is strong evidence suggesting that the purpose is unclear by several auditors as outlined in Section 4.1.4.

Validity is divided into three parts, face, content, and construct validity. Construct validity is related to root cause and is therefore outside of the scope of this study.

Face validity refers to whether the error taxonomy seems to fulfill its purpose on face value. As stated previously, this is problematic to claim based on the discussion in 4.1.4. Another piece of evidence against face validity would be the observation of the high degree of skepticism towards the error taxonomy expressed by certain auditors as well as non-auditors from within the case company.

Content validity means that an error taxonomy covers the full scope of possible errors that can occur. It can be argued that this criterion is utopian from the outset. Section 4.1.1. provides insight by highlighting the dilemma of balancing broad vs narrow categories. Only two abstraction levels are used within the current error taxonomy and certain subcategories are rather broad. This means that it is difficult to comment on content validity because on the one hand, the broad categories might cover the scope of all possible errors but on the other hand they may compromise the level of granularity that is needed to generate sufficient details about the findings. Nevertheless, this particular dilemma is difficult to balance and should be treated with an appropriate amount of respect.

In conclusion, the validity of the current error taxonomy is low which naturally leads to the question of how it could be improved.

5.3. Proposed improvements of the current error taxonomy

When considering the empirical evidence of this study and applying the error taxonomy evaluation framework by Wiegmann & Shappell (2001), many shortcomings of the current error taxonomy of the case company can be identified. However, the empirical data also informs potential solutions, primarily based on the insight gained from the layering process of Section 4.4 and the themes that emerged from interviews. Applying the 3 layers of evaluation criteria in Section 4.4., resulted in a much higher perceived actionability of the data, which was confirmed by a senior director at the case company. This increase in perceived actionability mainly took place when applying SOPs as an evaluation criterion to determine whether findings were sufficiently similar to be a good approximation of error recurrence. It can also be seen clearly that layer 1 and layer 2 both result in significantly higher approximations vs layer 3 and those approximations are clearly overestimations of error recurrency, for all effective purposes.

Furthermore, Theme 4.1.1. indicates that the dilemma of balancing broad vs narrow categories is a daunting challenge. Theme 4.1.2. indicates that the categories are too general. Theme 4.1.3. indicates that auditors categorize differently due to several factors. Theme 4.1.4. indicates that auditors perceive the categorization process to be an obstacle that adds limited value and that the purpose is unclear. Theme 4.1.5. raises concerns about training, ultimately causing even more variance in how auditors

categorize. This variation is seen in practice from the results of Case 1 in Section 4.2.1. Theme 4.1.6 indicates that there is no formal reporting of error recurrence even though auditors seem to know intuitively when a finding is recurrent. In fact, auditors must know if a finding is recurrent because it is used as a basis for raising the risk-classification of the finding.

All themes indicate that the categories only introduce problems without adding much value. Therefore, the following recommendations are made concerning the error taxonomy.

1. Reevaluate the purpose of having categories and subcategories.
2. Make the GxP area and SOP the cornerstones of the error taxonomy.
3. Add a checkbox for recurrency which is based on auditors' intuition and experience.

The SOP should not only be a sporadic comment in the finding description but a proper metric in the error database. Other than that, it is recommended to add a checkbox for recurrence that auditors can use if they recognize that they have encountered the same finding before, based on their memory and intuition.

To clarify, it is proposed to categorize findings into columns based on GxP area and SOP. Please note that this does not result in a graph such as the one seen in figure 4.3 which shows the final result of the layering process. The layering process simply illustrates how the selected evaluation criteria were applied to find a higher degree of similarity between findings. Instead, figure 5.1 is a visualization of what the data would look like using the proposed error taxonomy. As a result, all findings within a certain column will be highly similar and should therefore serve as a better approximation for error recurrence vs the current error taxonomy, which categorizes findings into subcategories. To illustrate, Figure 5.0 shows the 100 findings categorized into columns based on GxP area and SOP, using the proposed error taxonomy. In addition to this, Figure 5.1 only shows SOPs that were violated by at least 2 findings and hence, the frequency of findings within each column are approximate measures of error recurrence for every particular SOP. A further breakdown of the 100 sample findings using GxP-area and SOP can be found in Appendix D. Note that in figure 5.0 and 5.1, the frequency on the y-axis does not show the number of findings but rather the number of times the particular SOP is violated. This number can be different from the number of findings because certain findings violated several SOPs, in which cases it can be argued that the finding should have been separated into two different findings. Nevertheless, measuring and illustrating the number of SOP violations, circumvents this dilemma and highlights the most vital aspect of the data.
.

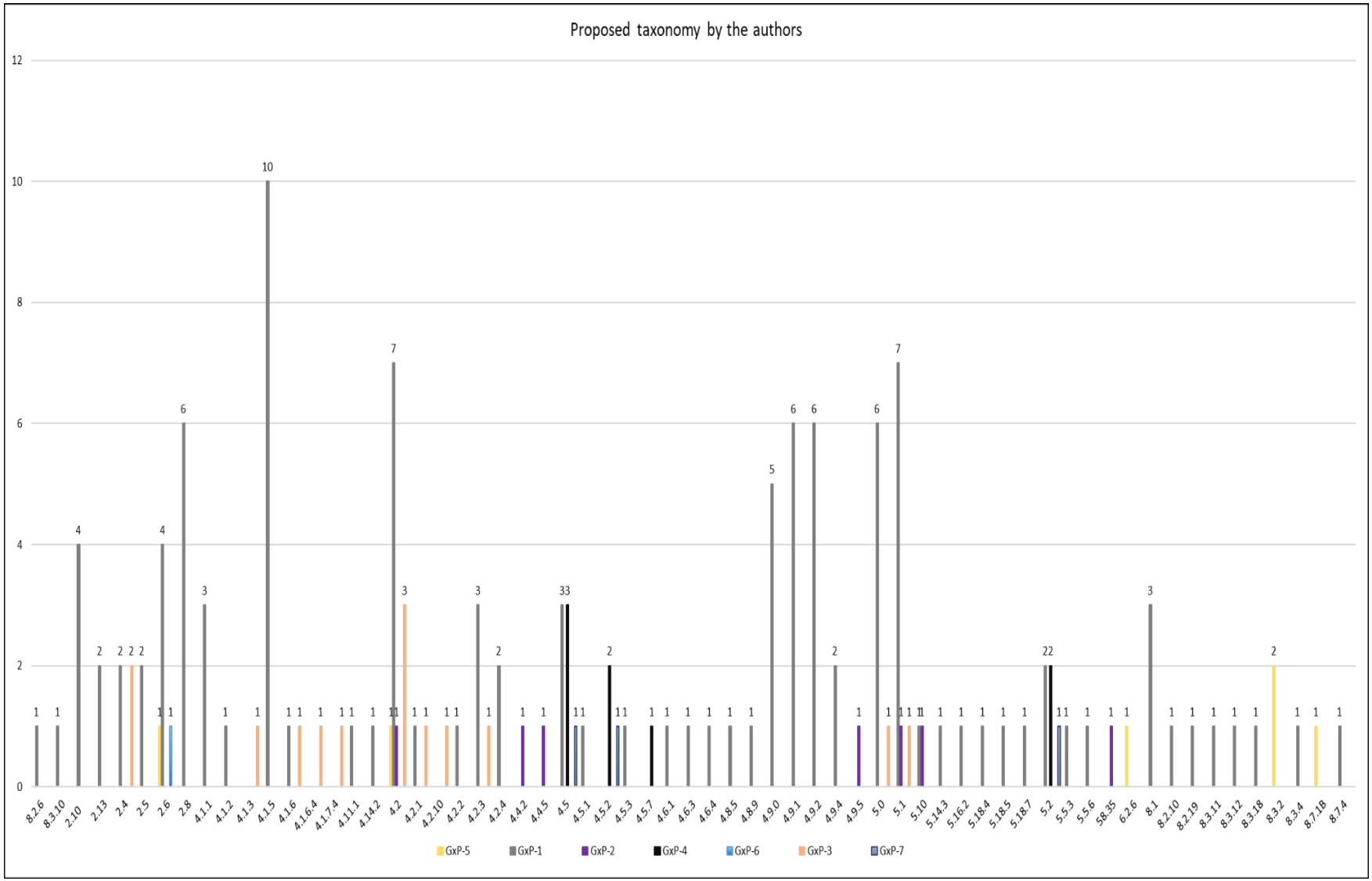

Figure 5.0, The proposed error taxonomy - Based on GxP area and SOP. Unique violations included.

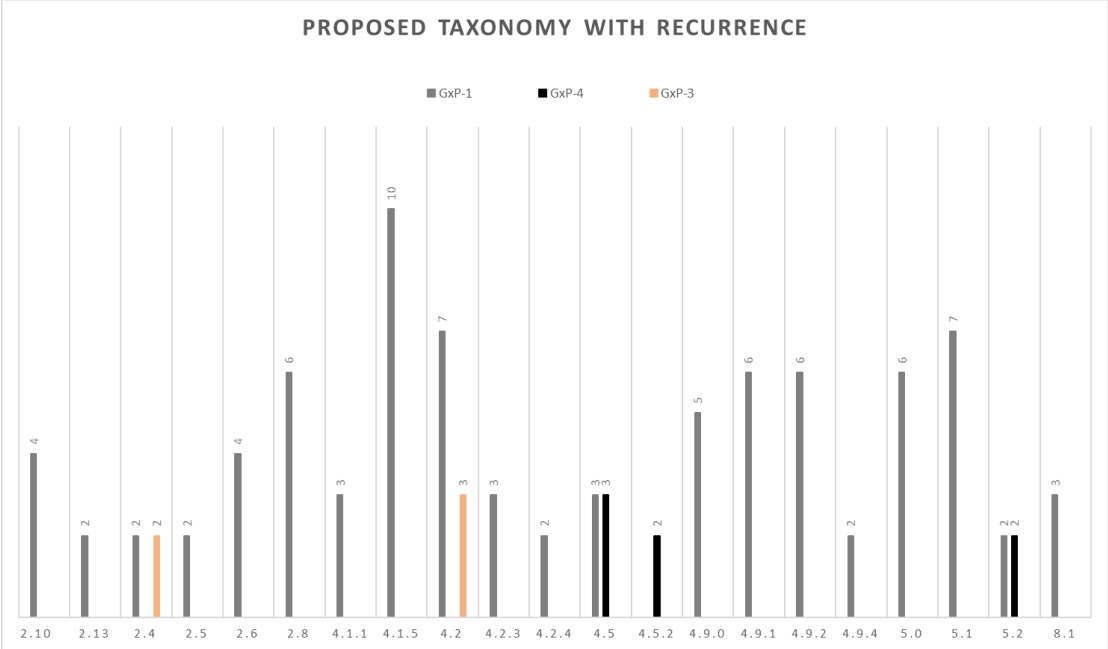

Figure 5.1, The proposed error taxonomy and recurrence approximation- Based on GxP area and SOP. Unique violations excluded

Next, further arguments will be presented as to why the proposed error taxonomy is deemed superior by once again using the error taxonomy evaluation framework by Wiegmann & Shappell (2001). Furthermore, in Section 5.3.6, the framework by HakemZadeh & Baba (2016) will be applied to justify the actionability of the proposition.

5.3.1. Reliability

When assigning the SOP that has been violated, to a finding, a deep understanding of what has occurred is needed because the SOPs are very specific and broken down for each GxP area. This increases the likelihood of the data being correct because it leaves less room for interpretation compared to assigning a subcategory.

Furthermore, the issue of auditors potentially categorizing differently is eliminated and it is believed to be considerably more unlikely for auditors to assign different SOP violations due to their high specificity.

It is therefore believed that better reliability could be achieved with the proposed error taxonomy.

5.3.2. Comprehensiveness

The proposed error taxonomy does not necessarily provide more complementary data vs the current error taxonomy, but GxP area and SOPs are deemed to generate the most significant data. The most important addition with the proposed error taxonomy is to make the SOP a standardized metric compared to only being mentioned in the finding

description as free text. The checkbox for error recurrence is also believed to add value to the data that is being generated. To clarify, the notion of error recurrence in this context refers to the cases where the auditors remember that it is a recurrent finding. If, in the future, an organization-wide definition is established, that should naturally serve as the evaluation criteria for selecting the checkbox. Due to these additions, the proposed taxonomy is deemed superior in terms of comprehensiveness.

5.3.3. Diagnosticity

The relevant aspect of diagnosticity is the degree to which the error taxonomy allows autonomous data driven decision making. Whilst that might be a high ambition and the proposed error taxonomy might not fully reach that stage, it is deemed to be a step in the right direction. Not trusting the reliability of the data, which is the case with the current error taxonomy, is the main limiting factor for diagnosticity. It is believed that the proposed error taxonomy will enhance reliability as discussed in Section 5.3.1. and thereby also increase diagnosticity. The SOPs should give good indications of where interventions need to be made.

5.3.4. Useability

As mentioned in Section 5.2.4., the practical useability of the current error taxonomy is not necessarily deemed difficult. However, the useability of the proposed error taxonomy is still argued to be superior because it removes the burden of categorizing into subcategories which can sometimes be deemed tedious and ambiguous by auditors. The useability is especially enhanced for those auditors that work within a GxP area that the subcategories are less relevant for. The proposed error taxonomy works equally well within any GxP area.

It is therefore believed that superior useability could be achieved with the proposed error taxonomy.

5.3.5. Validity

Since validity is regarded to be an accumulated measure of how well the other four criteria are met, and the proposed error taxonomy is believed to be superior to the current error taxonomy across all criteria, it follows logically that the proposed error taxonomy has higher validity.

Validity was furthermore divided into 3 parts:

Face validity - It is believed that categorizing based on SOP violation will increase the face validity of the error taxonomy instantly and significantly. If findings are categorized based on SOP, an observer will directly notice that findings within the same SOP category are highly similar, patterns will be much easier to identify, and it will be

significantly more clear what intervention to apply to prevent the error from happening in the future. Combined, these factors are highly likely to give a sense of face validity.

Content validity - Due to the fact that SOPs are highly specific and practically every action that takes place within the organization has an associated SOP, the proposed error taxonomy arguably does a good job in covering every possible scenario where errors might occur.

Construct validity - Outside the scope of this project.

In summary, the proposed error taxonomy arguably has higher validity than the current error taxonomy.

5.3.6. Actionability

The 8-criteria framework by HakemZadeh & Baba (2016) essentially evaluates whether research findings are actionable to practitioners. See detailed descriptions of each criterion at the end of Section 2.2. The framework is used to argue for the actionability of the proposed error taxonomy.

The criteria of **operationality** are fulfilled as the proposed error taxonomy could be readily implemented. This is because SOPs are mostly already being reported and the same is true for the GxP area. All that has to be done is to add a formal metric where the auditor can select an SOP criterion as well as a checkbox if the auditor intuitively believes the finding is recurrent.

The criterion of **causality** is fulfilled because the SOPs would give those who analyze the data a very clear understanding of what the error is about. Conversely, it will give a much better indication of what must be done to solve emerging error patterns.

The criteria of **contextuality** are enhanced as a consequence of eliminating the subcategories that did not consider the context. The subcategories forced the auditors to categorize findings regardless of the GxP area, even if the relevance of the taxonomy for that particular GxP area was limited. This practice would risk thwarting the data unnecessarily. The proposed framework is fully applicable in any GxP context and does not risk negatively influencing the data by introducing ambiguity.

Comprehensiveness is fundamentally lost when research simplifies reality too much. The proposed error taxonomy does the exact opposite in the sense that the SOP guideline imposes a strict requirement to portray reality exactly as it was.

Persuasiveness concerns trust in the research findings. It is believed that strong empirical evidence has been presented arguing for the proposed error taxonomy being superior to the current error taxonomy in several ways. Furthermore, the layering process in Section 4.4., illustrates how further evaluation criteria can be tested to potentially find an even better way to find recurrent errors.

The **Comprehensibility** criteria is fulfilled automatically because the recipients of this study are highly familiar with the current error taxonomy as well as what the proposed error taxonomy would entail. It is argued that the comprehensibility of the proposed error taxonomy is higher than that of the current error taxonomy from the perspective of the auditors. This is basically because it simplifies their work process by eliminating a step that they deem tedious and unnecessary.

Conceptual clarity is arguably the only criteria that the proposed error taxonomy does not necessarily fulfill in a superior manner vs the current error taxonomy. No further conceptual insights are gained by using the proposed error taxonomy. However, it can be questioned if more conceptual clarity is needed in the case of the error taxonomy. The proposed taxonomy is highly practical in nature, and it can be argued that a good understanding of the GxP area is sufficient for its effective implementation and use.

To conclude, it is believed that 7/8 criteria of what HakemZadeh & Baba (2016) argue to be criteria for actionable research, are better fulfilled by the proposed error taxonomy. The conceptual clarity criteria are not necessarily fulfilled better but not necessarily worse either. More importantly, the significance of the conceptual clarity criteria is questioned in this context.

A potential weakness with the proposed error taxonomy is that it is yet to be tested in a practical setting. SOPs might not always be easy to identify mainly because findings can violate multiple SOPs which raises questions in how exactly to approximate recurrence. Furthermore, categories are not solely negative, they might serve a purpose such as providing a higher-level overview of where errors occur. Such an illustration might be more favorable when presenting to higher level managers within the firm even though the categories themselves do not aid in approximating recurrence. It is therefore important to consider who the recipient and the pros and cons must be weighed against each other when it comes to having categories. Due to time constraints, this study did not conduct a more thorough investigation into why the categories were designed, it is possible that more insights could have been gathered by doing so to get a more righteous understanding. Lastly, as stated before, it is possible that even better evaluation criteria exist, and experimenting is needed in order to find those.

6. Conclusion and recommendations

The purpose of this study was to understand the usefulness of the current error taxonomy, at the case company, in assisting to measure error recurrence patterns and the empirical findings have illuminated several concerns. Furthermore, the current error taxonomy does not hold up well against frameworks from the selected literature. The first research question was:

What are the benefits and potential problems of the current error taxonomy?

Regarding benefits, it has been found that the error taxonomy is deemed easy to use in practice. This is a direct consequence of updating the error taxonomy from having narrow categories to more broad categories. However, the empirical findings of this study overwhelmingly highlight problems, rather than benefits with the current error taxonomy. A fundamental pain point for the case company is that they are unable to come up with a satisfactory definition for error recurrence which, in turn, means that error recurrence cannot be measured. This study has ultimately attempted to shed light upon the concept of error recurrence by consulting the error taxonomy. It is however unfortunately concluded that the current error taxonomy is not very useful when it comes to assisting in measuring error recurrence. This is primarily because the data generated by the current error taxonomy is not considered to be actionable, meaning that appropriate interventions cannot be identified based on the data. The second research question was:

How can the current error taxonomy be improved to better fulfill the stated purpose?

The study has provided many insights into the plethora of problems with the error taxonomy but also into how to solve the problems. The answer to the second research question will be presented next in the form of recommendations. The first recommendation is directly aimed at the case company. Subsequent recommendations are primarily aimed at the case company but can also be useful in a more general context to inform the design of actionable error taxonomies.

The first recommendation concerns the recommended improvements to the error taxonomy that were presented in section 5.3. Firstly, to reevaluate whether categories and subcategories are truly needed. Second, to base the error taxonomy on GxP area and SOP. Finally, to add a checkbox for error recurrence that auditors can use when they believe a finding is recurrent based on their intuition and experience. It is firmly believed that the proposed error taxonomy improvements would generate considerably more actionable data vs the current error taxonomy and it would also assist in measuring error recurrence. Categories and subcategories have not shown to provide much support in finding error recurrence patterns but instead seems to slightly complicate the job of auditors. With that being said, it does not mean that the categories are completely

useless. They fit nicely into PowerPoint presentations to give a neat overview of the high-level distribution of errors. To certain audiences, this type of data is more appropriate to display. It is certainly possible to keep the categorization system as a part of the error taxonomy if this is deemed to be an important factor, but it is firmly advised to weigh the pros and cons against each other.

The second recommendation is to include auditors when discussing changes to the error taxonomy because they have first-hand experience in using the system. This can be as simple as asking for their opinions about how they are experiencing the error taxonomy, much like this study has done through the interviews. Furthermore, the opportunity should be taken to define and communicate the purpose of the error taxonomy to auditors in order to help them appreciate the bigger picture of their work and strengthen the communication between managers and auditors.

To reduce variation in the data that is being produced, it is recommended to firmly establish standards in how to structure reports and what data points to include, rather than simply having guidance documents and relying on sporadic comments in the finding descriptions. These standards need to be communicated effectively, and it is highly recommended that they are accompanied by a basic level of training. An important benefit of these practices is that they reduce the risk of organizational silos forming between managers and auditors as well as among auditors.

The third recommendation is to refrain from trying to define error recurrence in a highly detailed manner from a top-down perspective. Instead, it should be approached from a bottom-up perspective, with the ultimate goal being to build an error taxonomy that generates actionable data and assists in measuring error recurrence. Taking a bottom-up approach starts with asking the question “what is considered to be actionable data?”, and then, evaluation criteria can be selected to categorize errors in a way that generates the desired data and assists in measuring error recurrence. It is very possible that better evaluation criteria can be found compared to the ones proposed in this study. This is especially true when considering the organization from a more holistic perspective compared to the rather limited scope of this study.

By definition, recurrent errors must be highly similar. It is believed that a sufficient level of similarity between findings can provide a good indication of error recurrence. An acceptable level of similarity can be found by experimenting with evaluation criteria and analyzing the output data until it seems reasonable at face value. The layering process has illustrated how this can be done and the process can easily be replicated and continued to find even better evaluation criteria.

Furthermore, the concept of error recurrence has a significant intuitive component which should be embraced and taken advantage of, as suggested by the first recommendation. For example, the memory and intuition of auditors, indicated by clicking a checkbox, can be used as an additional evaluation criterion. The goal to aim

for should be to have categories within which findings are sufficiently similar to inform appropriate interventions. If that stage can be achieved, a top-down definition of error recurrence will not be necessary since the overarching goal will be achieved and it could therefore be said that the categories of the error taxonomy give a sufficient, or good enough, measure of error recurrence.

To conclude, error taxonomies can be useful given that they are designed in a way that generates actionable data. In order to be able to design such an error taxonomy, organizations must have a good understanding of what they mean by actionable data. They can then build the taxonomy with an experimentative bottom-up approach until the taxonomy generates the desired data. It is highly important that all relevant parties are involved in the design of the error taxonomy in order to establish a strong common ground in how to use it.

6.1 Further research - Academic perspective

Further research should foremost focus on how to develop a taxonomy taking into consideration the Generic error modeling system (Reason, 1990; Reason, 1997) and/or Skill, Rule, Knowledge-based behavior theory (Rasmussen, 1986). Foremost, effort should be put on how to apply these two theories in the setting of R&D and how it could benefit the understanding of error recurrence. The initial idea of this study was to apply Rasmussen's or Reasons theories as final layers. However, the authors deemed it difficult to apply these concepts due to the study having a too wide approach, focusing on all process areas.

Thus, research centered around Reason or/and Rasmussen's theories should foremost be limited to one process and one operation to be able to capture the real cognitive process. Moreover, Rosch et al. (1976) work in categorization could be beneficial for the case company if they deemed it necessary to develop a new taxonomy. Due to the fact that Rosch et al. (1976) theory would be interesting to apply in the context of R&D and highly skilled personnel.

6.2 Further research - Organizational perspective

The study was limited to only investigate the error taxonomy when applied to audit findings. These findings only constitute a small subset of the total findings within the error database of the organization. Further studies could be made to investigate how the taxonomy is used in different parts of the case company's organization. Requiring personnel without auditing experience to categorize errors based on SOPs would most likely be completely non-viable because this practice requires subject matter knowledge and experience. Therefore, it is highly likely that different parts of the organization would require adaptations to the error taxonomy. Even though the

limitation of scope in this study creates many question marks and challenges related to the generalizability of the findings, the hope is that the study has at least provided a direction as to where to start the improvement work and how it can be done.

Bibliography

Anu, V., Hu, W., Carver, J. C., Walia, G. S., & Bradshaw, G. (2018). Development of a human error taxonomy for software requirements: A systematic literature review. *Information and Software Technology*, *103*, 112–124. https://doi.org/10.1016/j.infsof.2018.06.011

Bell, E., Bryman, A., & Harley, B. (2019). *BUSINESS RESEARCH METHODS 5E* (5th ed.). Oxford University Press.

Cahill, L. B. (2017). *Environmental Health and Safety Audits: A Compendium of Thoughts and Trends* (Second ed.). Bernan Press.

Carroll, J. S., Christianson, M. K., Frese, M., Lei, Z., Naveh, E., & Vogus, T. J. (2021). FROM THE EDITORS—Errors in Organizations: New Frontiers in Research and Practice. *Academy of Management Discoveries*, *7*(4), 451–460. https://doi.org/10.5465/amd.2021.0217

Cilliers, F., & Greyvenstein, H. (2012). The impact of silo mentality on team identity: An organizational case study. *SA Journal of Industrial Psychology*, *38*(2). https://doi.org/10.4102/sajip.v38i2.993

Creswell, J. W., & Clark, V. P. L. (2011). *Designing and Conducting Mixed Methods Research* (Second ed.). SAGE Publications, Inc.

Deacon, D., Bryman, A., & Fenton, N. (1998). Collision or collusion? A discussion and case study of the unplanned triangulation of quantitative and qualitative research methods. *International Journal of Social Research Methodology*, *1*(1), 47–63. https://doi.org/10.1080/13645579.1998.10846862

Dimitrova, N. G., van Dyck, C., van Hooft, E. A., & Groenewegen, P. (2014). "Error Prevention, Error Management, or Both?" *Academy of Management Proceedings*, *2014*(1), 15702. https://doi.org/10.5465/ambpp.2014.15702abstract

Dimitrova, N. G., & van Hooft, E. A. J. (2021). In the Eye of the Beholder: Leader Error Orientation, Employee Perception of Leader, and Employee Work-Related Outcomes. Academy of Management Discoveries, 7(4), 530–553. https://doi.org/10.5465/amd.2019.0184

Fischer, S., Frese, M., Mertins, J. C., & Hardt-Gawron, J. V. (2018). The Role of Error Management Culture for Firm and Individual Innovativeness. *Applied Psychology*, *67*(3), 428–453. https://doi.org/10.1111/apps.12129

Frese, M., & Keith, N. (2015). Action Errors, Error Management, and Learning in Organizations. *Annual Review of Psychology*, *66*(1), 661–687. https://doi.org/10.1146/annurev-psych-010814-015205

HakemZadeh, F., & Baba, V. V. (2016). Measuring the actionability of evidence for evidence-based management. *Management Decision*, *54*(5), 1183–1204. https://doi.org/10.1108/md-01-2015-0001

Hofmann, D. A., & Frese, M. (2017). *Errors in Organizations (SIOP Organizational Frontiers Series)* (1st ed.). Routledge.

Hooper, B. J., & O'Hare, D. P. A. (2013). Exploring Human Error in Military Aviation Flight Safety Events Using Post-Incident Classification Systems. *Aviation, Space, and Environmental Medicine*, *84*(8), 803–813. https://doi.org/10.3357/asem.3176.2013

Love, P. E., Lopez, R., & Kim, J. T. (2013). Design error management: interaction of people, organisation and the project environment in construction. *Structure and Infrastructure Engineering*, *10*(6), 811–820. https://doi.org/10.1080/15732479.2013.767843

Matthews, J., Love, P. E., Ika, L. A., & Fang, W. (2022). Error aversion or management? Exploring the impact of culture at the sharp-end of production in a mega-project. *Developments in the Built Environment*, *10*, 100074. https://doi.org/10.1016/j.dibe.2022.100074

Mohrman, S. A., Gibson, C. B., & Mohrman, A. M. (2001). Doing Research that is Useful to Practice a Model and Empirical Exploration. *Academy of Management Journal*, *44*(2), 357–375. https://doi.org/10.5465/3069461

Pearce, J. L., & Huang, L. (2012). Toward an Understanding of What Actionable Research Is. *Academy of Management Learning & Education*, *11*(2), 300–301. https://doi.org/10.5465/amle.2012.0130

Rasmussen, J. (1986). *Information Processing and Human-Machine Interaction: An Approach to Cognitive Engineering (North-Holland Series in System Science and Engineering, 12)* (1st Printing ed.). Elsevier Science Ltd.

Reason, J. (1990). *Human Error* (1st ed.). Cambridge University Press.

Reason, J. (1997). *Managing the Risks of Organizational Accidents*. Ashgate.

Reason, J., & Hobbs, A. (2003). *Managing Maintenance Error: A Practical Guide*. CRC Press.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439. https://doi.org/10.1016/0010-0285(76)90013-x

Rousseau, D. M. (2006). Is there Such a thing as "Evidence-Based Management"? *Academy of Management Review*, *31*(2), 256–269. https://doi.org/10.5465/amr.2006.20208679

Sarof, D. (2008). *Does business ignore the business schools?* Http://Davidsartof.Blogspot.Com/. Retrieved May 8, 2022, from http://davidsartof.blogspot.com/2008/01/does-business-ignore-business-schools.html

Stanton, N. A., & Salmon, P. M. (2009). Human error taxonomies applied to driving: A generic driver error taxonomy and its implications for intelligent transport systems. *Safety Science*, *47*(2), 227–237. https://doi.org/10.1016/j.ssci.2008.03.006

Syed, J., Mingers, J., & Murray, P. A. (2009). Beyond rigour and relevance: A critical realist approach to business education. *Management Learning*, *41*(1), 71–85. https://doi.org/10.1177/1350507609350839

Tamuz, M., Thomas, E. J., & Franchois, K. E. (2004). Defining and classifying medical error: lessons for patient safety reporting systems. *Quality and Safety in Health Care*, *13*(1), 13–20. https://doi.org/10.1136/qshc.2002.003376

Walia, G. S., & Carver, J. C. (2012). Using error abstraction and classification to improve requirement quality: conclusions from a family of four empirical studies. *Empirical Software Engineering*, *18*(4), 625–658. https://doi.org/10.1007/s10664-012-9202-3

Wang, P., & Hao, Y. L. (2017). What Role Does Error Orientation Play in the Relationship Between Error Management Climate and Innovative Behavior? *DEStech Transactions on Social Science, Education and Human Science*, *emse*. https://doi.org/10.12783/dtssehs/emse2017/12780

Wiegmann, D. A., & Shappell, S. A. (2001). Human Error Perspectives in Aviation. *The International Journal of Aviation Psychology*, *11*(4), 341–357. https://doi.org/10.1207/s15327108ijap1104_2

Zhang, J., Patel, V. L., Johnson, T. R., & Shortliffe, E. H. (2004). A cognitive taxonomy of medical errors. *Journal of Biomedical Informatics*, *37*(3), 193–204. https://doi.org/10.1016/j.jbi.2004.0

Appendix A - Semi-structured interview guide

Appendix A consists of the developed semi-structured interview guide that was used during all interviews. In addition to this, the mail that was sent out before the interviews.

Interview guide - Semi structured - Audits

Structure:

The interviews are divided into three different themes and a case exercise at the end. The timeframe for the interviews is assumed to be 60 minutes.

Theme 1 and 2 are more informal and aim to "warm up" the interview subject as well as getting an understanding of their work. These parts are intended to be less structured and kept brief.

Theme 3 and the case exercise are more formal and constitute the most critical parts of the interviews. These parts are intended to be more structured. 3-4 key questions should be identified here.

This interview guide will be refined as the interviews progress to account for any learning that occurs on how the following interviews could be optimized. The focus on theme 2 is expected to be reduced because the information provided there is likely highly uniform, i.e. doesn't vary much between the subjects.

Facesheet:

General info about the person we interview (can perhaps be done beforehand to save time). Name, age, position, number of years employed, numbers of years involved in the auditing process.
However, each interview will be **anonymous**, this face sheet is intended for internal use only to help manage and sort data during the analysis phase.

Theme 1 - Warm up & getting to know the subject

- Introduction, job description, current work, previous work.
- Making sure they know they are anonymous.
- Keep it brief and effective to save time fo theme 3 and case

Theme 2 - Getting to know the audits process in general, the work flow and structure

- What is the process like once a finding is identified? How long does the process take and how often does it occur?
- Can likely be kept very brief for subsequent interviews to save more time for later parts.

Theme 3 - Direct questions about the categorization system

Due to time constraints, 3-4 questions should be specified for this part. The following points are some thoughts that are relevant for the project/interview, 3-4 questions can be constructed based on these.

- What is the thought process when categorizing a finding?
    - Potential follow up: Ask about the training/education they've received around the categorization system.
    - Ask about the role of gut feeling in the process
- Do you believe that auditors share a common view of how to categorize findings?
    - Do you work alone or with another auditor(s)?
    - How often do you agree/disagree with other auditors?
- How would you define error recurrence
    - Do you think the error categorization system is a viable basis for the measuring of error recurrence? why/why not.
- What is your general opinion of the categorization system based on your experience?
    - Is it easy to use?
    - Are there ambiguities between certain categories?
    - How is ambiguity handled? Does “category-favoritism” occur?
    - Would you say that the current categorization fulfills its purpose adequately?

- Follow up (sort of): Do you feel that the categorization system needs to be improved? Why/why not?

    - Why is it important to categorize correctly?
    - Do you have any ideas on how it could be improved?

Cases - Let the auditors divide findings into different categories.

- Pick 3 cases from our sample and observe the process of how the auditors use the categorization system.

End: Ask if there is something they want to add or have any questions.

Interview mail template

Hi NAME!

My name is Peter, and I am writing my Master's thesis together with Gustav Bergling at—-. As I am sure — informed you, part of our study is to look at the error categorization framework in practice and therefore we are conducting interviews with auditors. We have scheduled an interview with you as you can see, hopefully you are interested in participating and that this time works for you. The interview is expected to take 60 minutes but we've scheduled 75 to leave room for potential discussions that arise around our questions. The content in brief is that we start with a short introduction followed by a couple of semi structured questions. Then we will look through a few cases (audit findings) and discuss how the categorization framework is used when categorizing the findings. Hopefully this works for you, please get back to us if you would prefer a different time.

Kind Regards

Peter & Gustav

Appendix B - Comparison of total case company findings versus 100 random findings

Appendix B consists of a comparison between all the findings given to the authors by the case company and comparing them to the random 100 samples, calculating the differences for each subcategory.

C1	1000 Case provided by case company	100 Random cases	% in Sample	% in total population	Difference	ABS Diff
C1-S1	12	1	1,01%	1,00%	0,01%	0,01%
C1-S2	31	4	2,60%	4,00%	-1,40%	1,40%
C1-S3	7	0	0,59%	0,00%	0,59%	0,59%
C1-S4	5	0	0,42%	0,00%	0,42%	0,42%
C1-S5	11	0	0,92%	0,00%	0,92%	0,92%
C1-S6	33	2	2,77%	2,00%	0,77%	0,77%
C1-S7	71	2	5,95%	2,00%	3,95%	3,95%
C1-S8	13	1	1,09%	1,00%	0,09%	0,09%
C2						
C2-S1	18	3	1,51%	3,00%	-1,49%	1,49%
C2-S2	46	4	3,86%	4,00%	-0,14%	0,14%
C2-S3	10	0	0,84%	0,00%	0,84%	0,84%
C2-S4	12	0	1,01%	0,00%	1,01%	1,01%
C2-S5	1	0	0,08%	0,00%	0,08%	0,08%
C2-S6	194	22	16,26%	22,00%	-5,74%	5,74%
C2-S7	56	5	4,69%	5,00%	-0,31%	0,31%
C2-S8	33	4	2,77%	4,00%	-1,23%	1,23%
C2-S9	205	14	17,18%	14,00%	3,18%	3,18%
C3						
C3-S1	14	1	1,17%	1,00%	0,17%	0,17%
C3-S2	8	0	0,67%	0,00%	0,67%	0,67%
C3-S3	2	0	0,17%	0,00%	0,17%	0,17%
C3-S4	10	0	0,84%	0,00%	0,84%	0,84%
C3-S5	8	1	0,67%	1,00%	-0,33%	0,33%
C3-S6	67	6	5,62%	6,00%	-0,38%	0,38%
C3-S7	4	0	0,34%	0,00%	0,34%	0,34%
C3-S8	102	12	8,55%	12,00%	-3,45%	3,45%
C3-S9	18	2	1,51%	2,00%	-0,49%	0,49%
C4						
C4-S1	4	0	0,34%	0,00%	0,34%	0,34%

C4-S2	*19*	*3*	*1,59%*	*3,00%*	*-1,41%*	*1,41%*
C4-S3	*1*	*0*	*0,08%*	*0,00%*	*0,08%*	*0,08%*
C4-S4	*5*	*0*	*0,42%*	*0,00%*	*0,42%*	*0,42%*
C4-S5	*1*	*0*	*0,08%*	*0,00%*	*0,08%*	*0,08%*
C4-S6	*3*	*1*	*0,25%*	*1,00%*	*-0,75%*	*0,75%*
C4-S7	*2*	*0*	*0,17%*	*0,00%*	*0,17%*	*0,17%*
C5						
C5-S1	*25*	*2*	*2,10%*	*2,00%*	*0,10%*	*0,10%*
C5-S2	*9*	*2*	*0,75%*	*2,00%*	*-1,25%*	*1,25%*
C5-S3	*2*	*0*	*0,17%*	*0,00%*	*0,17%*	*0,17%*
C5-S4	*2*	*1*	*0,17%*	*1,00%*	*-0,83%*	*0,83%*
C5-S5	*3*	*0*	*0,25%*	*0,00%*	*0,25%*	*0,25%*
C5-S6	*4*	*1*	*0,34%*	*1,00%*	*-0,66%*	*0,66%*
C5-S7	*4*	*1*	*0,34%*	*1,00%*	*-0,66%*	*0,66%*
C6						
C6-S1	*2*	*0*	*0,17%*	*0,00%*	*0,17%*	*0,17%*
C6-S2	*33*	*5*	*2,77%*	*5,00%*	*-2,23%*	*2,23%*
C6-S3	*1*	*0*	*0,08%*	*0,00%*	*0,08%*	*0,08%*
C6-S4	*4*	*0*	*0,34%*	*0,00%*	*0,34%*	*0,34%*
	1193	*100*		***AVG Difference***	*-0,15%*	***0,89%***

Table B.1: Comparison of total case company findings versus 100 random findings.

Appendix C - Table of evaluation for case company system

Appendix C consists of the table that was used to evaluate the case companies' current categorization system with layering.

	Total Findings	***GxP-1***	***GxP-2***	***GxP-3***	***GxP-4***	***GxP-5***	***GxP-6***	***GxP-7***
C1								
C1-S1	*1*							
8.7.1B						*1*		
C1-S2	*4*							
4.2.3		*1*						
No Criteria		*1*	*1*					*1*
C1-S6	*2*							
2.13		*1*						
5.1				*1*				
C1-S7	*2*							
5.5.6		*1*						
2.10		*1*						
C1-S8	*1*							
4.14.2		*1*						

5.5.3		*1*						
C2								
C2-S1	*3*							
2.10		*1*						
2.5		*1*						
5.10		*1*						
4.5.1		*1*						
2.13		*1*						
C2-S2	*4*							
4.8.9		*1*						
4.8.5		*1*						
8.3.12		*1*						
No Criteria		*1*						
C2-S6	*22*							
4.1.5		*6*						
4.1.1		*2*						
4.2.4		*2*						
8.3.4		*1*						
4.1.2		*1*						

4.2.3		*1*						
8.3.10		*1*						
8.3.11		*1*						
8.2.19		*1*						
8.2.6		*1*						
2.8		*1*						
4.9.4		*1*						
8.1		*1*						
5.18.7		*1*						
5.18.5		*1*						
5.18.4		*1*						
8.2.10		*1*						
5.2		*1*						
No Criteria		*1*				*1*		*1*
C2-S7	*5*							
5.10			*1*					
2.6		*1*						
4.6.1		*1*						
4.6.3		*1*						

4.5.3		*1*						
5.0		*1*						
No Criteria		*1*						
C2-S8	*4*							
4.5.7					*1*			
No Criteria					*2*		*1*	
C2-S9	*14*							
2.10		*1*						
4.9.5			*1*					
58.35			*1*					
4.9.0		*6*						
4.9.1		*4*						
4.9.2		*4*						
4.9.4		*1*						
C3								
C3-S1	*1*							
No Criteria		*1*						
C3-S5	*1*							
No Criteria								*1*

C3-S6	*6*							
4.1.3				*1*				
4.4.2			*1*					
8.3.2						*2*		
4.4.5			*1*					
8.7.1b						*1*		
No Criteria								*1*
C3-S8	*12*							
4.2.10				*1*				
4.1.5		*4*						
2.8		*4*						
2.4				*1*				
4.1.6		*1*						
4.2						*1*		
6.2.6						*1*		
C3-S9	*2*							
4.2.1		*1*						
4.2.2		*1*						
4.2.3		*1*						

4.5		*1*						
C4								
C4-S2	*3*							
8.7.4		*1*						
5.14.3		*1*						
4.6.4		*1*						
No Criteria		*1*						
C4-S6	*1*							
No Criteria				*1*				
C5								
C5-S1	*2*							
4.9.1		*1*						
4.9.2		*2*						
C5-S2	*2*							
No Criteria				*2*				
C5-S4	*1*							
8.3.18	*1*							
5.16.2	*1*							
C5-S6	*1*							

4.11.1		*1*						
C5-S7	*1*							
4.1.6.4				*1*				
4.1.7.4				*1*				
C6								
C6-S2	*5*							
4.5.2					*2*			*1*
No Criteria		*1*						*1*

Table D.1: Table of evaluation for case company system

Appendix D - Table of evaluation of the authors developed system

Appendix D consists of the table that was used in evaluation of the developed system by the authors using layering.

SOP CRITERIA	***GxP-5***	***GxP-1***	***GxP-2***	***GxP-4***	***GxP-6***	***GxP-3***	***GxP-7***	***Total***
	7	*65*	*6*	*5*	*1*	*10*	*6*	*100*
8.2.6		*1*						
8.3.10		*1*						
2.10		*4*						
2.13		*2*						
2.4		*2*				*2*		
2.5		*2*						
2.6	*1*	*4*			*1*			
2.8		*6*						
4.1.1		*3*						
4.1.2		*1*						
4.1.3						*1*		
4.1.5		*10*						
4.1.6		*1*				*1*		
4.1.6.4						*1*		

4.1.7.4						*1*		
4.11.1		*1*						
4.14.2		*1*						
4.2	*1*	*7*	*1*			*3*		
4.2.1		*1*				*1*		
4.2.10						*1*		
4.2.2		*1*						
4.2.3		*3*				*1*		
4.2.4		*2*						
4.4.2			*1*					
4.4.5			*1*					
4.5		*3*		*3*			*1*	
4.5.1		*1*						
4.5.2				*2*			*1*	
4.5.3		*1*						
4.5.7				*1*				
4.6.1		*1*						
4.6.3		*1*						
4.6.4		*1*						

4.8.5		*1*						
4.8.9		*1*						
4.9.0		*5*						
4.9.1		*6*						
4.9.2		*6*						
4.9.4		*2*						
4.9.5			*1*					
5.0		*6*				*1*		
5.1		*7*	*1*			*1*		
5.10		*1*	*1*					
5.14.3		*1*						
5.16.2		*1*						
5.18.4		*1*						
5.18.5		*1*						
5.18.7		*1*						
5.2		*2*		*2*			*1*	
5.5.3		*1*						
5.5.6		*1*						
58.35			*1*					

6.2.6	*1*							
8.1		*3*						
8.2.10		*1*						
8.2.19		*1*						
8.3.11		*1*						
8.3.12		*1*						
8.3.18		*1*						
8.3.2	*2*							
8.3.4		*1*						
8.7.1B	*1*							
8.7.4		*1*						
No criteria	*1*	*7*	*1*		*1*	*3*	*5*	

Table E.1: Table of evaluation of the authors developed system