



universität
wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„The Observation Bias in Performance Monitoring“

verfasst von / submitted by

Monika Chavushyan, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

Wien, 2024 / Vienna 2024

Studienkennzahl lt. Studienblatt /
degree programme code as it
appears on
the student record sheet:

UA 066 915

Studienrichtung lt. Studienblatt /
degree programme as it appears on
the student record sheet:

Masterstudium Betriebswirtschaft

Betreut von / Supervisor:

Uni.-Prof. Dr. Thomas Pfeiffer

Acknowledgements

Through this acknowledgement I would like to express my gratitude towards Univ.-Prof. Dr. Thomas Pfeiffer as my supervisor and to Markus Eigruber, PhD as my co-supervisor. The completion of my thesis would not have been possible without their guidance and support. I would also want to thank the management of the company from the experiment in my thesis for allowing me to conduct my research there, work with real data and gather valuable insights on my topic. Special thanks to all five participants who agreed to take part in my experiment and lastly, to all my friends and family for supporting me throughout this journey.

Abstract

Performance monitoring, a broadly used tool for the assessment of employees' achievements on individual and organizational level, has proven to be beneficial on multiple occasions. While it is often applied in combination with pay-for-performance models, monitoring performance exhibits a self-contained functionality as well. The observation bias describes the phenomenon that people tend to adapt their behavior when they know that they are being observed. The first part of this thesis guides through existing literature and evidence on the effect of monitoring by itself by introducing the three possible consequences of observation: a positive change, a negative change and no change in behavior. The literature review is followed by a field experiment, aiming to answer the question whether performance monitoring can have an impact on overall productivity and effectiveness. A multi-variable evaluation model has been established and utilized for the hypotheses' verification. The findings of the research are summarized and discussed in the last two chapters "Results" and "Discussion". The summary and the statistical analysis of the results demonstrate a significant change in performance after the implementation of a monitoring system, supporting the hypothesis that employees tend to positively alter their behavior under observation. A variety of adaptations has been proposed in the chapter "Limitations" in order to enable a differentiation between the impact of qualitative and quantitative criteria.

Abstract - Deutsch

Das Monitoring von Mitarbeiterleistung, eine breit eingesetzte Methode zur Bewertung von Performance auf individueller und organisatorischer Ebene, hat sich in mehreren Hinsichten als vorteilhaft erwiesen. Während es oft in Verbindung mit Pay-for-Performance Modellen angewandt wird, weist das Monitoring von Mitarbeiterleistung auch eine eigenständige Funktionalität auf. Der Observation Bias beschreibt das Phänomen, dass Menschen ihr Verhalten unter Beobachtung tendenziell anpassen können. Der erste Teil dieser Arbeit bietet einen Überblick über die vorhandene Literatur, indem die drei möglichen Folgen der Beobachtung vorgestellt werden: eine positive Veränderung, eine negative Veränderung und keine Veränderung im Verhalten. Nach dem Literaturüberblick folgt ein Feldexperiment, das die Frage beantworten soll, ob die Leistungsüberwachung eine Wirkung auf die Gesamtproduktivität und -effektivität haben kann. Ein mehrdimensionales Modell zu Performance Messung wurde erstellt und für die Überprüfung der Hypothesen angewandt. Die Ergebnisse der Forschungsmethode wurden in den letzten beiden Kapiteln „Results“ und „Discussion“ zusammengefasst und erörtert. Die Zusammenfassung und die statistische Analyse der Daten zeigen eine signifikante Veränderung der Leistung nach der Einführung eines Monitoring-Systems. Das unterstützt die Hypothese, dass Mitarbeiter ihre Leistung unter Beobachtung verbessern können. Unterschiedliche Anpassungen wurden im Kapitel „Limitations“ vorgeschlagen, um die Differenzierung zwischen dem Einfluss qualitativer und quantitativer Kriterien zu ermöglichen.

Table of Contents

1. ABSTRACT	3
2. ABSTRACT - DEUTSCH	4
3. TABLE OF CONTENTS	5
4. LIST OF TABLES	6
5. LIST OF FIGURES	7
6. INTRODUCTION	8
7. LITERATURE REVIEW	10
7.1. MONITORING	10
7.2. THE HAWTHORNE EFFECT	15
7.3. BEHAVIORAL CHANGES DUE TO MONITORING	17
7.4. PERFORMANCE MEASUREMENT	22
8. RESEARCH METHOD	31
8.1. EXPERIMENTAL DESIGN	31
8.2. EXPERIMENT EXECUTION	53
8.3. RESULTS	55
9. DISCUSSION	73
10. CONCLUSION	82
11. LITERATURE	83
APPENDIX A – GRADE DEFINITION	87
APPENDIX B – CHECKLIST COLLABORATIVE INTERFACES, PROJECT TAKEOVER AND -MANAGEMENT	88
APPENDIX C – CRITERIA INFORMATION	89
APPENDIX D – DECLARATION OF CONSENT (CEO)	90
APPENDIX E – DECLARATION OF CONSENT (PARTICIPANTS)	91
APPENDIX F – SHAPIRO-WILK TEST OF NORMALITY	93
APPENDIX G – QUANTILE-QUANTILE PLOTS	95

List of Tables

Table 1: Grades per participant for 2022 and 2023	61
Table 2: Statistical data, grades per category for 2022 and 2023	64
Table 3: Inter-category comparison	65
Table 4: Statistical data, grades per participant for 2022 and 2023.....	67
Table 5: Inter-participant comparison (grades).....	68
Table 6: Inter-participant comparison (points)	69
Table 7: Wilcoxon signed-rank significance test.....	72
Table 8: Inter-participant comparison (robustness check)	80

List of Figures

Figure 1: ERP-system Home page	45
Figure 2: Tableau report “Comparison between target and current calculations”	45
Figure 3: Boxplot inter-category comparison.....	66
Figure 4: Boxplot inter-participant comparison	70

Introduction

Employees' engagement is crucial for company's success and is therefore one of the highest organizational priorities. Companies address that topic in several different ways, most common of which are improved efficiency, communication, interaction and feedback. Although there is no perfect formula for encouraging employees' engagement, studies suggest that positive attitudes towards the workplace and active involvement in working activities often transform "the employee potential into employee performance" (Gupta & Sharma, 2016).

Monitoring is an organizational controlling instrument used by companies for supervising employees' behavior. It can be executed by providing regular feedback, setting specific performance goals or through direct supervision by the line manager. However, performance monitoring systems (PMS) are not just used to observe certain actions but also aspire improvement and development. And as personnel performance is directly linked to organization success, we study the effect of performance monitoring on employees' performance in order to provide an alternative perspective for enhancement of organizational success.

Performance monitoring is used to measure the quality of work and its outcome (Wholey & Hatry, 1992). Usually, those outcomes are used to make a comparison between different departments, geographic locations or different points in time. Performance monitoring aims to gain insight on employees' actions and characteristics and helps analyze those over time. There are often certain goals that need to be met and monitoring helps keeping track of them.

When the company's objective is the well-being of employees and a long-term organizational success, managers need to address intrinsic motivation. Intrinsic motivation comes from within and helps employees engage in working processes because they are satisfied with their surrounding conditions. On the other hand, extrinsic motivation is provoked by external factors and focuses on the outcome such as money or power. For managers it might be challenging distinguishing those two types of motivation. Furthermore, when employees receive additional attention, they might adjust their performance just because they are being observed and not due to enhanced motivation.

That fact by itself carries several risks and this needs to be considered carefully. One specific study – the Hawthorne experiment, has demonstrated that people act and perform differently when they know that they are being observed. The original Hawthorne experiment took place between the years 1924 and 1932 at the Hawthorne Works factory.

It aimed to determine whether employees' productivity and motivation change under different illumination conditions. However, that study gathered evidence and reached conclusion on a more complex topic: the importance of social and psychological factors at the workplace and created the term "The Hawthorne effect". The Hawthorne effect describes the fact that people act and perform differently when they know that they are being observed.

The understanding of the observation bias in monitoring is essential for personnel management and controlling. There are many controversial opinions on the topic whether employees should be closely observed or should feel that they are being observed. According to modern studies there are some "healthy borders" of monitoring and control which shouldn't be crossed in order to show trust in the decisions of employees, make them feel more autonomous and avoid counterproductive work behavior (Yost et al., 2019). On the other hand, the Hawthorne effect might be used by itself to improve performance, just by regularly letting employees know that their performance is being measured.

The main focus of this thesis lies on the possible Hawthorne effect that may result from performance monitoring and how that might be used by organizations to enhance motivation. The first question to be answered is:

"Is it possible to improve employees' overall performance by implementing a multi-variable monitoring system compared to a setting without monitoring?".

Second, this thesis aims to answer the question:

"Does the Hawthorne effect in employees' monitoring have a bigger impact on qualitative criteria like customer satisfaction and collegiality or on quantitative criteria such as the improvement of contribution margins or the adherence to invoice submission deadlines? ".

Literature review

Monitoring

Monitoring in an organizational setting can be referred to as the process of obtaining information on performance through observation. While there are several types of monitoring that can be chosen by organizations, such as financial or project monitoring, this thesis offers an in-depth analysis of the literature on employees' monitoring. The objective of personnel monitoring is to obtain information on their performance and efficiency and provide foundation for management decisions (Liao & Chun, 2016). Furthermore, the gathered information can be used by the company to allocate resources and establish rewarding systems (Liao & Chun, 2016).

Similar to almost every topic nowadays, we need to consider the fast-evolving technology in the context of monitoring as well. Monitoring is usually performed by a monitoring agent. That agent can be a person or an automatized software. A person has the opportunity to evaluate a situation in a more subjective way, considering human interactions and exceptions, while the computer has the ability to analyze bigger data for shorter time in a more precise way. Both of them have advantages and disadvantages, which are discussed in the next chapters. Several researches have indicated that depending on the company's structure and strategy, a combination of both computer and personal monitoring might deliver optimal results.

Computer Monitoring

Computer monitoring is associated with an electronic surveillance of employees that aims to gather information about important working activities. This data can be utilized to gain knowledge and recognize process inefficiencies. It can also be used for training new employees or even machines. Often it is used to control regulations compliances, improve productivity and maintain security (Yost et al., 2019). However, some employees report to feel "overly controlled" by their organization (Jensen & Raver, 2012). According to a survey from 2012 in Australia, 78% of companies monitored the internet usage of their employees, 88% were blocking certain websites, and 49% were monitoring the email activity of their workers (Holland et al., 2015).

Considering the potential negative impact on performance, one should carefully choose its controlling instruments when it comes to computer monitoring. As a manager, it is important to differentiate between relevant and irrelevant information. When employees believe that the collected data is used to improve employees' and customer satisfaction, they are more

likely to show willingness to cooperate (Wells et al., 2007). On the contrary, when they are convinced that the company wants to control their actions and detect regulations defiance, they expressed reduced job satisfaction (Alge, 2001).

The purpose of this thesis is to find whether performance monitoring might have a positive impact on motivation and on which criteria it manifests itself most prominently. And as employees' dissatisfaction is connected to low motivation, we conclude that computer monitoring should be used reasonably, only for work-related activities, without invading the privacy of the workers.

Performance Feedback

Prue and Fairbank (1981) define performance feedback as information reported to employees regarding the qualities and quantities of their work. In other words, letting people know how their past performance was. Many companies use performance feedback as a controlling instrument, aiming to influence a better quality of work and better outputs. Before looking at the effectiveness of that approach, it is important to know if the performance feedback is used alone or in combination with other controlling instruments. Those complementary instruments can be setting performance goals, introducing incentive plans or behavioral consequences (Balcazar et al., 1985). Studies have shown that giving regular feedback on performance might improve certain organizational areas such as punctuality, excessive time off, customer support, and security (Balcazar et al., 1985).

However, performance feedback can be divided into several categories which determine the effectiveness of the method. It can vary between different sources of the message, contents, recipients, and frequency (Balcazar et al., 1985). The feedback source is the person or the machine that provides the data. Although multiple types of sources have been identified, in this thesis we focus on feedback that comes only from supervisors, customers or software devices. Nevertheless, in the final experiment of this thesis we provide enough instruments for unsolicited self-generated feedback as well. The content of the performance evaluation refers to the comparison of current performance with a previous one and different benchmarking methods (Balcazar et al., 1985). The recipients are categorized in individuals, groups or a combination of both (Balcazar et al., 1985). The frequency of performance feedback might be daily, weekly and monthly.

Performance feedback is often used in combination with other controlling instruments. "Feedback-alone" describes the process where information about work performance is provided by itself (Balcazar et al., 1985). "Feedback and behavioral consequences" refers to organizational systems where employees get regular feedback and might suffer positive

or negative consequences of their behavior (Balcazar et al., 1985). An example for a positive consequence would be an additional time off work and for a negative – reduction of bonuses (Balcazar et al., 1985). The third type of combination is the “Feedback with Goal Setting” (Balcazar et al., 1985). This is where certain performance objectives are being defined additionally to the regular feedback.

The study by Balcazar et al. (1985) demonstrates the effectiveness of those combination sets. They found that the feedback-alone brings only 28% of consistent success in employee motivation. Feedback and behavioral consequences have achieved 52% of behavioral change, similar to feedback in combination with goal setting at 53%.

Performance feedback can be made available through direct reports, peers, supervisors and self (Smither et al., 2005). When all of those sources provide their opinions and ratings simultaneously, it is called multi-scored feedback. It is not only important to know your instruments, but also to be familiar with your target audience. Smither et al. (2005) show that in order to achieve an improvement in performance, one needs to know how to indicate feedback acceptance and rejection by the recipient. For example, it is unlikely to achieve a better quality of work, when the employee who has received the feedback initially reacts in an unfavorable way (Ilgen et al., 1979). In that context, it is essential for leaders to know how to navigate performance feedback in the desired direction and outcome.

Setting Performance Goals

Setting performance goals describes the definition of short- and long-term objectives on a company level as well as individually. That process is often combined with incentive plans in order to achieve extra motivation. However, in this thesis we study the effect of performance goals by itself without introducing any further bonus system. Setting objectives often provides employees with direction by helping them focus their attention and ambition on a specific goal. For instance, a laboratory experiment of 214 subjects has gathered evidence on the positive connection between goal clarity and performance (Anderson & Stritch, 2016). Furthermore, having a well-defined purpose on the workplace enhances motivation and therefore improves performance. Nevertheless, there is evidence that focusing on the output of a task and not on the learning process or on value-creation might undermine intrinsic motivation (Heyman & Dweck, 1992).

Setting performance goals promotes a sense of accountability. When people are feeling responsible for a task, it is more likely to be at the forefront of the progress. A study by Itoh et al. (2008) presents empirical findings on the complementarities among delegated authority, accountability, and monitoring. The data has been sampled from several

Japanese companies. The collected information confirms their hypotheses that monitoring approaches are associated with higher accountability among employees (Itoh et al., 2008). On the other hand, increased accountability is found to be favorable for employees' and company's performance (Itoh et al., 2008).

When defining individual or group objectives, managers stimulate team cooperation. Working together towards a common goal often supports a shared sense of purpose at the workplace. As team cooperation is responsible for employees' satisfaction and retention, one may conclude that setting goals has an important role in the long-term success of an organization.

There are several drawbacks which need to be considered with caution when defining team or individual objectives. The literature suggests that setting unachievable goals might lead to frustration and demotivation (Soman & Cheema, 2004). Furthermore, managers need to make sure that objectives are defined thoughtfully, considering the possibility of "goal displacement". Goal displacement is defined by the process of achieving a certain objective at the cost of other important aspects of the work.

Firstly, we can take into consideration the findings of Dishneau (1992). He studied the sales goals in an auto repair company. The goal was 147 dollars per hour (Dishneau, 1922). In order to reach that amount, employees started to raise prices for the customers and perform unjustified repairs (Dishneau, 1922). This on the other hand, has a high probability to end up with unsatisfied customers. The chief executive officer of the company at the time admitted that "goal setting process for service advisers created an environment where mistakes did occur" (Dishneau, 1922).

Another example for misplaced goals is the energy-trading company Enron. Enron's executives were rewarded based on sales volume, not a profit. By prioritizing the wrong goals, Enron was led directly to a financial devastation (Ordonez et al., 2009). Moreover, when setting challenging goals, employees might try to achieve it at the expense of other parts of the works. In the 1960s, Ford's former CEO announced a challenging goal of manufacturing a new car model that weighs less than 900 kilograms, costs less than 2000 U.S. dollars and will be ready to sell in ten years (Ordonez et al., 2009). One team developed the car under all the conditions mentioned above. However, after a safety check has been done and a design fault has been discovered, that team decides to continue working towards the goal. Their argument was that fixing the false design would cost more than the legal costs associated with 53 deaths and many injuries (Ordonez et al., 2009). That ambitious goal costed the company its reputation, ethical behavior and the ignorance

of safety regulations. The authors of “Goals Gone Wild: The Systematic Side Effects of Overprescribing Goal Setting” argue that the drawbacks of goal setting might outweigh their benefits (Ordóñez et al., 2009).

Setting Deadlines

Setting deadlines is similar to setting goals, but it refers to the timeline in which a task is being executed. It is important for companies for two main reasons. Firstly, managers want to finish projects on time, as on-time execution is one of the three most important part of project management. The second reason is the attention to time, that employees acquire when specific deadlines are set. For instance, studies have shown that employees tend to measure progress by the remaining time (Okhuysen & Waller, 2002). Therefore, deadlines are expected to have the multiple function of measuring performance by the management and as a self-controlling mechanism.

The literature suggests that as a deadline approaches, people are more likely to become aware of the remaining time and resources. This can help optimize resource allocation and therefore improve overall performance. Waller et al. (2002) promote the idea that with deadlines around the corner, group motivation and pacing effort are being enhanced. They also suggest that groups or individuals with pending deadlines tend to switch the focus of their tasks to the most significant parts and therefore deliver a better performance (Waller et al., 2002).

However, setting deadlines might lead to stress and anxiety associated with the workplace. Very tight deadlines might be often impossible to meet and that can lead to frustration among the personnel. With the growing position of work-life balance in today's society, having to ignore or put aside personal matters might intensify the frustration resulted from work-related deadlines. In fact, a study by Herbert et al. (2014) shows that the top six concerns associated with tight deadlines at work are: “restricting family holidays, time spent on work at home, impact on children, stress at home, impact on family and friends, and impact on partner”.

Moreover, rushing through work tasks might result in critical quality drawbacks. For companies and business areas where certain quality standards are expected, a missed deadline might still be a better outcome than ignoring important legal and safety requirements. In the subsequent experiment, the company of choice is operating in a high-risk sector, that requires all of the safety aspects to be satisfied, therefore setting deadlines is considered just as a complementary controlling instrument.

Direct Supervision

The idea of direct supervision refers to manual monitoring indicated by human involvement. In the means of this work, direct supervision will not be used alone, but in combination with behavior assessment scales. However, there are some benefits and drawbacks of direct supervision that are worth mentioning.

Personal monitoring ensures quality. Human observation can control the aimed quality and offer guidance and feedback through the process. In that way it is also possible to train new personnel and learn from mistakes. Some further benefits of managerial supervision are real-time adaptability and problem-solving(Christensen et al., 2008) of human beings. Unlike a machine, people can quickly adapt to changing conditions, make quick decisions and work in a flexible manner. Considering the expansion of artificial intelligence, it is most likely that the difference between human and computer monitoring in terms of flexibility will be reduced over time.

Like any other controlling instrument, there are some risks associated with direct supervision as well. Human opinions and decisions are very subjective. If employees don't share the same point of view with their supervisors, the feedback of their manager might feel unfair to them. Direct supervision is resource-demanding as well. It requires time, manpower and financial means, which often are better spent elsewhere. Lastly, delegating supervising responsibilities to people can make a company dependable on a few individuals and their know-how.

A quantitative study by Susmadiana et al. (2021) supports the idea with their findings that direct supervision has a positive impact on employees' performance. Nevertheless, this is not the only factor that influences performance. If monitoring is not executed properly, there are many concerns on the part of employees that can be raised. An example for that would be the privacy of people at the workplace, which when violated, might lead to dissatisfaction and poor performance.

The Hawthorne Effect

The Hawthorne effect describes the phenomenon that people tend to adapt their behavior when they know that they are being observed. That change in behavior can be positive or negative and happens regardless of external conditions. For instance, the workers in the Hawthorne factory increased their performance under every condition change they had: increased or decreased illumination (Fernald et al., 2012). There are some researches that

report about an unaltered behavior as well. All of the three possible outcomes are reviewed in the following chapter “Behavioral Changes due to Monitoring”.

Over the past few decades, the Hawthorne effect has been studied and applied in multiple fields. Researchers have used the Hawthorne effect in the context of healthcare interventions such as the use of digital health records, administration protocols and hygiene compliance rules (Chen et al., 2015). Another example for observation and its impact on performance is found in the educational field in the form of non-traditional learning methods such as “Flipped Classrooms” (Christensen et al., 2008). Lastly, the Hawthorne effect has been studied in the context of social psychology to understand how people's behavior is influenced by social groups and situational factors (Asch, 1955).

Jones (1990) reevaluated the original Hawthorne study by using modern statistic methods. He suggested that the outcomes of the original research feature interdependencies between the employees and even their personalities (Jones, 1990). He found that not only the fact whether employees are being observed by their managers has impact on their performance, but also the performance of their co-workers (Jones, 1990). In this thesis, those findings will be used to support the opinion about the importance of peers' feedback and observation. Studies point at both good and bad sides of co-workers' influence on motivation.

According to an article by Smet et al. (2023) there are six types of employees. The first one is “the quitter” and is estimated to build around ten percent of the personnel (Smet et al., 2023). According to the authors, they are not automatically the worst performers but mostly end up leaving the company regardless of the efforts of the management (Smet et al., 2023). Next are “the disruptors”, again at around ten percent. They are the part of the team that is the less committed and often demonstrates its dissatisfaction (Smet et al., 2023). In that way they endanger the performance of the good performers as they might “suck out the motivation out of work and workers around them” (Smet et al., 2023). The next group consists of “the mildly disengaged” and builds up to thirty-two percent of the staff (Smet et al., 2023). They usually deliver just the required work and never more. “The double dippers” makes only five percent of the workforce and consists of workers, who typically have more than one job (Smet et al., 2023). “The reliable and committed” is estimated at thirty-eight percent (Smet et al., 2023). This is where the satisfied and committed employees are executing their daily activities and sometimes bring up their own ideas. The last group is “the thriving stars” and consist of only four percent of the team (Smet et al., 2023). They manage to create value and a safe and productive setting for themselves and for their teams. It is important to understand what contributes to their well-being and try to use it on

the other groups. However, the authors advise that the top-performers also should be protected and taken care of. A successful management should be able to differentiate between the quality of work of top- and low-performers and evaluate it accordingly. If the performance management system (PMS) does not recognize the effort and the value of “the thriving stars”, they might end up feeling unjust and reduce their motivation or even leave the company (Smet et al., 2023).

The Hawthorne effect in monitoring

In this thesis, the focus lays on the observation bias in employees’ monitoring and its impact on performance. As there is no general formula how to distinguish the real change in behavior from the one resulted from the Hawthorne effect, an antonymous term will be used for clarification. In the context of this research, the opposite of the Hawthorne effect is to be understood as intrinsic motivation. Intrinsic motivation describes the drive to improve that comes from within and is connected to growth on the long-run (Heyman & Dweck, 1992).

According to Lied and Kazandjian (1998) there are two ways to recognize the Hawthorne effect resulting from performance measurement systems. One typical feature of motivation enhanced through the Hawthorne effect is its temporary nature (Lied & Kazandjian, 1998). It would normally fade away after a couple of months unlike the genuine, intrinsically enhanced motivation which could last for a longer period of time when being supported by the management. Another approach to differentiate would be to ask the employees for their feedback regarding their change in behavior, whether it was due to the fact that they are receiving more attention or because they believe in the functionality of the new controlling system (Lied & Kazandjian, 1998).

Behavioral Changes due to Monitoring

Monitoring is often used for minimizing costs and optimizing supervision of work. However, it is easier to use when having simple tasks to control. Nowadays, employees get complex work assignments, where controlling performance through incentive plans is more effective than just a simple monitoring (Guthrie, 2000). Nonetheless, monitoring remains an important part of performance measurement models and contributes to change in employees’ quality of work.

Gathani et al. (2015) conducted research on the effect of monitoring on work performance. The authors suggest that monitoring has a positive impact on work ethic and project outcomes by increasing accountability and devotion. They mention three potential effects

that could lead to changes in employees' behavior. The first one is the Hawthorne effect - behavioral changes that occur when individuals are observed during an experiment (McCarney et al., 2007). The second possible reason for change in certain actions is the mere-measurement effect. It is defined by the change in behavior of participants after they have been asked certain questions. For instance, Battaglia et al. (1996) found that mothers who have been asked whether their kids have been vaccinated are more likely to vaccinate them in the next months. The last effect is called "the reminder effect" and refers to the simple act of asking or reminding people about the completion of the task, aiming to raise awareness about its significance (Gathani et al., 2015). The authors have not been able to differ between the three channels. However, this thesis focuses on the Hawthorne effect as a consequence of monitoring and will therefore use the findings of their study to support hypotheses around that specific topic.

According to Holderness et al. (2022), employees that expect to be monitored are more likely to perform better. Another study from 1995 emphasizes the importance of multi-score feedback for the expected quality of work (London & Smither, 1995). They propose that feedback improves self-awareness. On the other hand, it is expected that self-awareness stimulates change in behavior and performance improvement (London & Smither, 1995).

In 2005 Smither et al. continued their studies in the field. They found that not only it is important to provide regular feedback in order to clarify goals and enhance self-awareness, but also it depends on the person who receives the feedback. According to their studies, it is most likely to achieve an improvement in performance when the feedback indicates that an improvement is needed (Smither et al., 2005). Another condition for successful feedback would be the perception of the employee, how they react to feedback and if they believe that change is possible (Smither et al., 2005).

This is where the Hawthorne effect meets monitoring – as a possible result of observation. In other words, monitoring is the action or the observation and the Hawthorne effect is the consequence of the action or in that case – the change of behavior. One needs to keep in mind that monitoring doesn't implicitly lead to a Hawthorne effect. A change in performance due to monitoring in pay-for-performance might and often does happen also due to enhanced motivation and because people know that part of their future income is depending on the quality of their work. In this paper we provide insight on the topic whether the aspired enhancement in motivation and performance happen because of the possibility to earn more money or if it is possible to happen just because of simple observation.

The knowledge of the fact that one's performance is being regularly measured might result in three possible outcomes: a positive change in behavior, a negative change in behavior and no change in behavior.

A positive change in behavior

There are multiple benefits from monitoring – from organizational to employee advantages. For example, customer support companies or call centers often record their customer calls. On the one hand, those recordings can be used for training new employees by showing them real scenarios. On the other hand, companies benefit from recorded calls by reviewing those and learning by possible mistakes. This can help organizations achieve a higher customer satisfaction (Bylinski, 1991).

Another organizational advantage from monitoring would be the cost savings. A Californian company managed to drop their costs by 15% by using a call accounting software for just a month. They tracked all the incoming and outgoing calls and therefore knew where there was a room for optimization and cost cutting regarding communication costs (Tucker, 1992).

Furthermore, Aiello and Kolb (1995) found out through an empirical investigation that many top performers started to complete their tasks faster. Findings by Griffith (1993) also confirm that companies have achieved some crucial improvements in the task completion time by implementing different monitoring systems.

Proponents of monitoring suggest that also employees could benefit from a well-implemented control system. According to Grant and Higgins (1988), providing employees' performance data might help them recognize their current position and productivity and therefore enable them to improve it. Monitoring also improves consistent work and clarity in terms of productivity (Grant et al., 1988). In the case of software controlling systems, Grant and Higgins (1988) found that computer monitoring is seen as fairer towards the workers compared to the subjective opinion of supervisors. The argument of accuracy of electronic monitoring has been confirmed a couple of years later in an experiment by (Fenner et al., 1993).

Another benefit of monitoring is the optimized allocation of resources by the management. By constantly reviewing employees' performance, leaders gain more insight on ongoing projects and have the chance to work on productivity and efficiency improvement (Ukko et al., 2007). In that way, a company gains a financial advantage through a more efficient resource allocation.

Moreover, the study by Ukko et al. (2007) shows that performance monitoring might improve the communication and teamwork between the management and the employees. Although reviewing employees' performance might be often misinterpreted as an excessive form of control, Ukko et al. (2007) found that regular conversation between the management and the employees has contributed to the specification of more exact performance goals. This results in a clarification of organizational strategy and incorporation of employees' input in it.

A negative change in behavior

Monitoring has many forms. In the last few centuries organizations have used the technological evolution to expand their controlling instruments. Telephone call observations, screen sharing networks, recordings of display turn on and offs are just a few examples of the power that employers have gained in the last years (Fickel, 1991). This kind of power is referred as a panoptic power (Mulgan, 1991). Panoptic power implies to a concept of observation in prisons, where there is a huge tower in the center. The guards of the prisons can see every cell and every inmate from the tower without being seen.

Approaches like this can be an unfavorable decision for a company, especially when the targeted outcome is more autonomous and performance-oriented way of completing tasks. Therefore, it is essential to understand which type of monitoring is suitable for a company and what objective it has. Two different types of monitoring are electronic monitoring and observational or interactional monitoring (Liao & Chun, 2016). The biggest difference between both of them is the lack of personal interaction in the first one. Depending on the aspired accuracy and outcome of monitoring, a mix of an electronic and interactional approach is often a suitable solution for companies. For instance, just a plain number of clicks per day or sales volume wouldn't be enough to make a conclusion on the motivation and performance of an employee without talking to them and creating a personal impression. The other way around, without the real numbers, a manager couldn't possibly distinguish the quality of work of two middle-performer. The middle performer are the people in a company who don't stand out with the best performance and tasks accomplishments but also don't belong to the lower bound.

The next question to be answered by an organization is what to monitor. A company needs to know whether they want to collect data about the whole activity of worker or just the performance-relevant activities. Narrowing down the spectrum of observation, together with further conditions, which will be discussed later in this thesis, may help the observer

(employer) gain some positive improvements in performance. However, when monitoring work-irrelevant activities and using the data to provide an invasive feedback, organization might experience disadvantages from excessive monitoring (Alder & Tompkins, 1997).

Unsuitable performance monitoring might result in unwanted consequences such as poor quality of work or disadvantaged teamwork. In their study from 1988, Grant et al. compared the performance of two working groups: one of which was monitored and the other one wasn't. They found out that the monitored group was aiming the quantity of work before the quality. Monitoring also led to a poor cooperation between coworkers and therefore to a deteriorated teamwork. The results pointing at bad quality of work due to monitoring have been confirmed by later studies. For example, directory assistance operators from several telecommunication organizations have reported that they feel incapable of providing a sufficient quality of work due to monitoring standards (Aiello, 1993). Another study provides evidence on that, by showing that 65% of the participant cannot achieve the quality they want to as monitoring is forcing them by administrating pressure of time (Ross, 1992).

Besides the disadvantages on performance, studies have shown that employees' privacy may often be violated by monitoring and stress sensations – provoked. For instance, a study from 1993 by Aiello shows that observing nonwork activities such as bathroom visits of employees might make them feel compromised. Those negative feelings towards the workplace are often associated with dissatisfaction and anxiety and lead to lower employee retention and poor performance. Therefore, it is important to consider not only the direct impact on performance but also the collateral damage that might follow from excessive monitoring.

Furthermore, the stress factor at the workplace shouldn't be underrated. If the objective of a company is to enhance motivation and achieve an improvement in their employees' performance, it should avoid conceptualizing their workplace as a "electronic sweatshop". The term was used in 1986 by Nussbaum and duRivage and in 1988 by Garson. It refers to an organization which monitors the activities of their workers on excessive scale without actually needing the whole information gathered. Moreover, this type of companies often uses the monitoring data to provide mostly critics and less improvement suggestions.

No change in behavior

The third possible outcome would be the lack of change in the observed behavior. That means that the employees who have been observed didn't adapt their behavior in any way. In the years 2009 and 2010 Fernald et al. (2012) conducted a clinical experiment in two groups – "observation" and "comparison". The participants who were receiving additional

attention and feedback were considered as observed. The group where nothing has changed was considered as a comparison. The results of the study show no changes in the methods of work in the “observed” group. Although the Hawthorne effect in employees’ monitoring is a topic of a more economic interest, one needs to be aware of each and every evidence the literature suggests. Therefore, considering the previous chapters, we can now proceed to our first three hypotheses.

Null Hypothesis (H0): Multi-variable monitoring has no impact on employees’ overall performance.

Alternative Hypothesis 1 (H1): Multi-variable monitoring has a positive impact on employees’ overall performance.

Alternative Hypothesis 2 (H2): Multi-variable monitoring has a negative impact on employees’ overall performance.

Performance measurement

In the next chapters we aim to answer the question “Does the Hawthorne effect in employees’ monitoring have a bigger impact on qualitative criteria like customer satisfaction and collegiality or on quantitative criteria such as the improvement of contribution margins or the adherence to invoice submission deadlines? “.

As mentioned, the choice of performance measurement systems (PMS) is crucial for employees’ monitoring to be successful. PMS can be based on quantitative as well as qualitative criteria. While quantitative measures refer more to the efficiency of a process, qualitative criteria describe the effectiveness of a task. According to Chase and Aquilano (1992), effectiveness means to do the right things and efficiency – to do things the right way. The effectiveness describes the achievement of intended results and a certain quality of work. On the contrary, the efficiency is identified as an optimal ratio of output to input (Meredith, 1992). Therefore, qualitative performance criteria are most commonly used for long-term strategic decisions and quantitative – for short-term (Zeydan & Çolpan, 2009).

There are several variations of performance monitoring. Usually, PMS should compare the work of different departments, analyze current performance versus previous one, or calculate the percentage of target achievement (Bartik, 1995). Furthermore, PMS might also be comparing the work of different geographic regions or population groups (Hatry et al., 1990). Another type of PMS works with benchmarks and statistical analysis and

compares groups, working under similar conditions that may have impact on the results (Wholey & Hatry, 1992).

To ensure that all the three pillars of project management are satisfied, it is often advisable to combine both qualitative and quantitative measures. “The Triple Constraints” of project management include quality, time, and cost. Balancing and optimizing those three factors is essential for a project to be defined as successful. In the following chapters those three fundamentals are discussed in the context of employees monitoring.

Qualitative measures

The importance of qualitative performance measures in terms of organizational and employees' performance has been rising in the last few decades. It provides more detailed picture of a certain performance profile and allows to take into account many important areas which cannot be quantified easily (Zeydan & Çolpan, 2009). Examples for qualitative performance measures are customer satisfaction, team cooperation, quality of data, and documentation. By promoting those areas of work, a company makes sure that the aspired success remains for a longer period of time and ensures a long-term growth. Some researchers even suggest that qualitative criteria are the only way to indicate performance (Hershey & Blanchard, 1969). However, considering only the effectiveness of work without the aspect of efficiency might not deliver the desired results (Sahay, 2005).

Quantitative measures

Quantitative criteria are essential for both manufacturing and service companies. The output of quantitative performance measurements is often indicated by numbers. They might be absolute numbers, percentages and ratios. An example for an absolute number would be the sales volume. A frequently used performance measurement in percent is a contribution margin. A popular ratio used to measure performance would be the service rate or service level. Some researchers suggest that only quantitative measurements deliver reliable results regarding organizational performance (Berman, 1998). However, more recent studies on multi-criteria decision-making argue that the combination of both qualitative and quantitative outputs might be most successful.

Combination of quantitative and qualitative measures

(Zeydan & Çolpan, 2009) propose a multidimensional approach for measuring performance in a manufacturing company based on both qualitative and quantitative criteria. They combine two methods: the fuzzy TOPSIS (technique for order performance (preference) by similarity to ideal solution) method and DEA (data envelopment analysis).

The primary idea of the TOPSIS method is to find a solution in a certain set of points, where the ideal outcome is identified by “the shortest path” (Parkan & Wu, 1999). The fuzzy logic is used to evaluate multidimensional, close to human decision-making systems. Therefore, the fuzzy TOPSIS analysis is used to collect and transform qualitative data into quantitative measures (Zeydan & Çolpan, 2009). These quantitative measures are the being evaluated in the DEA.

Another research by Bessane et al. (2018) supports the conviction that the combination of both approaches might be the most suitable solution. They found that these two methods are complementary to each other, as the one reinforces the weaknesses of the other and vice versa (Bessane et al., 2018). According to their study, qualitative criteria often lack precision. They are more suitable for complex and dynamic environments. Qualitative methods don’t require big sample sizes and can still achieve impactful observations (Bessane et al., 2018). On the other hand, one should expect more subjective opinions rather than exact (Bessane et al., 2018).

This is where quantitative criteria can contribute with their preciseness. They can find exact links and dependencies between variables, analyze big data samples, and offer more universal findings (Bessane et al., 2018). However, complex problems often need further interpretation and remain inflexible if considered only mathematically.

In this thesis, we combine both qualitative and quantitative performance measures. On the qualitative side stands project management, customer satisfaction, team cooperation, and data quality. The change in contribution margins and the adherence to invoice submission deadlines have been chosen for quantitative measures. This multidimensional model aims to secure a short-term, efficiency-focused success by motivating employees to achieve better monetary results in the form of higher contribution ratios. The effectiveness and the long-term success are supported by the enhanced customer satisfaction and employee retention, which result from improving the qualitative criteria.

The effect of monitoring on qualitative criteria

Team performance

Porter et al. (2010) study the impact of performance monitoring on teamwork and team performance. They found that observation on the workplace helps employees prevent mistakes and wrong decisions. In that way, collegiality and cooperation are being promoted. They, on the other hand, are found to be beneficial for overall motivation and performance (Marks & Panzer, 2004). Furthermore, by monitoring the work of a team, one

learns the strengths and weaknesses of the team members and therefore, knows how to optimally allocate tasks and resources (Porter et al., 2010). Considering the allocation theory, it is essential for teams to work together and support each other, as they always have finite resources available. Those resources might be time, money, competencies, experiences, and more. Porter et al. (2010) also suggest that the combination of performance monitoring and backing up behavior might deliver even better results. In that context, backing up behavior is defined by the act of helping colleagues, when they need assistance.

Martin et al. (2018) studied the relationship between shared leadership and team performance. Shared leadership describes the structured delegation of responsibilities among team members, which is the opposite of the traditional hierarchical leader. Martin et al. (2018) have found that an improved team performance is associated with the collaboration of multiple members and with the work of a single leader. They refer to several important characteristics when considering team performance. While the leadership typology, the complexity of the task and personality traits of the team members were positively associated with team performance, no evidence for a relationship between the team type and the methodology has been found (Martin et al., 2018).

Customer satisfaction

According to Cronin and Taylor (1992), “service quality is an antecedent of consumer satisfaction”. This means that companies that pursue an enhanced customer satisfaction need to focus on their service quality. Studies suggest that performance monitoring often has a positive impact on service quality. Therefore, addressing service quality by monitoring employees should lead to an increase in customer satisfaction.

Performance monitoring based on qualitative measures gains importance also in the public sector. More monitoring systems are being used every day for financial planning, resource allocation, enhancing employee motivation, improving service quality, and communication between the government and the citizens (Wholey & Hatry, 1992). According to Wholey and Hatry (1992) even the government tends to offer an unsatisfying service quality if not being monitored. By measuring the service quality and the outcomes accomplished at least annually or quarterly, one makes sure that employees are staying on track with their goals.

In 1992 The National Academy of Public Administration has adopted the following resolution. They made a strong recommendation for all public organizations to adopt performance monitoring on service quality and its outcomes (Wholey & Hatry, 1992). By the time, PMS should already be seen as an essential part of administration and financial

planning. Further declarations have been made regarding the implementation of PMS. These include consultation and involving of all interested parts when designing the system and setting goals, definition of KPIs of the system itself, and monitoring those (Wholey & Hatry, 1992). Moreover, they suggest that the establishment of realistic measurements supports the success of PMS and that it is important to use the collected data to actually improve performance and not only to collect it (Wholey & Hatry, 1992). Lastly, a reporting system should be integrated as well and should be informing about changes in performance (Wholey & Hatry, 1992). The further recommendations of the National Academy of Public Administration are not being included in this thesis as they are not considered to be relevant for the topic.

Although the last two paragraphs refer to the public sector, those findings can be used for a deeper understanding of performance monitoring in private entities as well. However, sometimes organizations don't manage to reach the aspired results and end up with unexpected negative consequences of monitoring. For instance, an excessive focus on current financial success might result in a decreased customer satisfaction. Frank & Enkawa (2009) found that customer satisfaction is positively related to economic growth but negatively related to short-term monetary expectations. In other words, if managers focus too much on gaining profit and cutting costs, there is a high probability that this will have a negative impact on customer satisfaction. Another example for the negative effect of monitoring on customer satisfaction would be the stress on the workplace caused by the pressure to meet performance goals. Aiello (1993) conducted research and showed that stress and burnout among employees due to monitoring is a possible outcome if monitoring systems are not implemented properly. That stress might affect their service quality in a negative way and subsequently the customer satisfaction. Lastly, a study by Coelho et al. (2021) identifies controlling mechanisms in a company as partly responsible for employee's creativity as well. If monitoring systems are not designed accordingly, this might result in a loss of creativity. Creativity of work is required for frontline employees in order for them to be able to respond to every unexpected customer concern. If they are feeling under pressure due to the current monitoring system, their creativity might suffer and result in poor service quality and customer satisfaction.

Quality of data and project management

The quality of documentation is crucial in various industries such as healthcare, legal, technical, and customer service. It is important for the compliance with legal regulations, knowledge sharing, decision-making, and therefore, the quality assurance. Managers can

help increase the quality of documentation, the structure of work, and time management by including a suitable monitoring system.

A study on visual search accuracy aims to find out whether the quality of employees' work might be influenced by the fact that they are being watched (Miyazaki, 2013). Three decision-making experiments have been conducted. The first one includes "watched" and "unwatched" conditions with twelve participants in total (Miyazaki, 2013). The participants had to perform a certain task once being observed and once – not. The second experiment was similar but instead of "watched", the participants have been videotaped (Miyazaki, 2013). The participants were aware that the video tapes might be watched in the moment or in the future. The third experiment included again the "watched" and "unwatched" conditions, but expanded to a predefined range of success in the decision-making process (Miyazaki, 2013).

The results of all three experiments show that participants tend to work slowly but more accurately when being observed, compared to when not being observed. In the context of this thesis, an accurate and more precise operation is defined as a prerequisite for quality assurance. Hence, observation might contribute to a better, more accurate documentation, and motivate employees to avoid errors and omissions that could reflect negatively on their performance metrics and their overall performance. While Miyazaki (2013) does not explicitly reference the term 'Hawthorne effect,' his findings are being utilized in this thesis to support research questions pertaining to that specific observational bias.

As consistency in documentation is crucial for clarity and continuity, for organizations it is important to encourage that kind of behavior. PMS might be the right answer to that, considering the fact that studies have found that monitoring consistency behaviors might have many advantages for the company. For instance, Bermbach and Tai (2014) support the opinion that evidence found by consistency monitoring might be used as a foundation for future decision-making. Taking responsibility for the accuracy and completeness of documentation might improve overall quality of work as well. Studies have found that those particular qualities can be addressed by performance monitoring. For example, Itoh et al. (2008) show that perceived accountability rises with monitoring intensity.

In this thesis, project management, customer satisfaction, team cooperation, and quality of data are utilized as a controlling instrument for quality assurance. They are supporting the quality aspect of the "Triple Constraints" of project management in the main experiment of this work.

The effect of monitoring on quantitative criteria

Adherence to invoice submission deadlines

The adherence to invoice submission deadlines refers to one of the three pillars of project management: time. By managing the chronological plan of a project, an organization avoids paying penalty fees, reputation damage, and assures a better and more efficient resource allocation. Furthermore, providing the final billing of a project is essential for the liquidity and solvency of the company. Nonetheless, a lot of managers have difficulties finishing their projects on time. Eik-Andresen et al. (2016) studied the tendencies and the reoccurring patterns regarding significant delays in the separate milestones of project management. Besides their main findings on portfolio management, their research on over 2000 project milestones also shows that managers, working on multiple projects at the same time show the tendency for deadline delays. In order to prevent liquidity risks, companies try to address those problems in different manners. One possible way to keep the time aspect under control is performance monitoring and considering deadlines as a performance metric.

On the contrary, Waller et al. (2002) assume that the simple existence of deadlines motivates employees to keep track of time and allocate resources wisely. They made an experiment on that how people perceive time in the different phases of a task completion and found that when deadlines are being set, people tend to become more aware of time and to pay more attention to it nearly before the deadline ends (Waller et al., 2002)

One alternative approach is the imposition of hard deadlines (Zhang, 2016). Hard deadlines are defined by a rather early than late goal of time completion. Zhang (2016) argues that setting earlier deadlines might help motivate employees more than paying bonuses after the end of a project, as the rewarding of success after the whole work has been done undermines ongoing motivation. In other words, setting timeline goals might be used as a monitoring instrument by itself, which encourages a better quality of work in addition. Saez-Marti and Sjögren (2007) also support the theory that strict deadlines might encourage a better quality of work, higher profits, and a better task completion.

However, rushing a project closure carries a likelihood of adverse consequences as well. While manufacturing entities might need to focus on perfect time and cost efficiency, many service companies might not want to risk important customer relationships or strategic decision. Those might often take longer than expected and be considered as inefficient, but they entail a long-term success and a company value. Therefore, when designing a PMS, it is important to consider the nature of work, the company culture, size and industry.

In this thesis, we focus on the end billing process of a projects as a deadline due to two main reasons. The initial perspective considers the financial aspect of the problem, as billing for products and services is crucial for ensuring the company's financial liquidity and its ability to endure. Secondly, the core activities of the company, where the experiment takes place, are associated with long-term customer relations. Moreover, many security requirements need to be met, which obscure the possibility to rush a project completion. Hence, the time aspect in that thesis is considered with a moderate importance.

Improvement of contribution margins

As profitability is one of the objectives with the highest priority that companies have, we want to consider that as well in the context of performance monitoring. Profitability is not only crucial for financial stability but also for building reliability and trust in the firm's name. Ling and Peh (2005) conducted research on fifty-five key performance indicators. They found that profitability is among the fourteen most important metrics for measuring company's performance. Contribution margins are a performance indicator, calculated by subtracting the variable costs from the revenue. It is a KPI (key performance indicator) preferred by many organizations as a measure for efficiency and effectiveness. Both company profitability and project profitability are essential. However, as employees should be measured only based on parameters which they have influence on, this thesis focuses on project profitability by studying the effect of monitoring and the possible Hawthorne effect on contribution margins.

The most common way to improve the financial state of a business by monitoring is cost control and optimization. It can help allocate resources efficiently and identify where improvement is needed. However, we want to study how monitoring impacts the employees' attitudes and whether it might bring improvement in contribution margins just by the fact that they know that they are being observed. The aim is to answer this question without the interference of further controlling instruments like incentive plans.

As mentioned in the previous chapters, the Hawthorne effect might contribute to increased productivity. The very first experiment that took place in the Hawthorne works of the Western Electric Company in 1924 showed that the productivity of the workers improved as the lighting in the working rooms has been increased (Gale, 2004). Nevertheless, the productivity of the control group and of the group where the lighting was reduced also improved (Gale, 2004). Even when the experimenters haven't had anything changed, the workers still continued to increase their motivation just due to the fact that they received additional attention. Therefore, in the context of quantitative performance measurements,

contribution margins might rise with enhanced productivity without implementing further controlling instruments.

Another reason for suggesting that monitoring might help improve contribution margins is the enhanced quality control. Miyazaki (2013) showed in his experiment, that people who are being watched or know their work is being recorded in some way, tend to pay more attention to details and therefore deliver a better quality of work. By ensuring a more controlled and concentrated working process, it is expected to reduce error rate in daily tasks. For the organization as a whole, this means that fewer resources will be needed for reworking or correcting mistakes, which on the other hand leads to cost reduction.

However, there is a drawback of quality assurance concerning the improvement of contribution margins. While working on quality management might ensure a better team cooperation and workplace atmosphere, it often doesn't show any positive effect on financial efficiency. In fact, Fisher (1990) has found that quality management has only moderate effect on profitability. He compared further approaches for productivity improvement such as tighter financial control and product rationalization and showed that they were more successful in enhancing profitability in the short-term (Fisher, 1990).

Thus, considering the findings on qualitative and quantitative performance measurement, the following three hypotheses have been formulated.

Null Hypothesis (H0): The Hawthorne effect in employees' monitoring has no impact on qualitative and quantitative performance measures.

Alternative Hypothesis 1 (H1): The Hawthorne effect has a bigger impact on qualitative criteria than it has on quantitative criteria.

Alternative Hypothesis 2 (H2): The Hawthorne effect has a bigger impact on quantitative criteria than it has on qualitative criteria.

Research method

A field experiment has been used as a research method to answer the two research questions and to broaden the findings from the literature review. The experiment took place in a middle-sized Austrian company from the energy and telecommunication sector. The company works mostly with public contractors who have very strict tender terms and qualifications that need to be met. Competing for the lowest price against multiple other suppliers, while maintaining a certain quality that meets the requirements of a public contractor, were two important characteristics of the setting in which the experiment took place. In addition, some of the projects take several years to complete. Keeping in mind the duration, the price competition, and the quality requirements, a company working under such complex conditions needs to keep track on internal agencies. The management of the company has decided to increase the attention on employees' performance by introducing a new performance monitoring system.

Experimental design

In the beginning of 2023, a management decision has been met to invest more resources in the monitoring of employees. The aim is to achieve an increase in financial success by ensuring an efficient workload and high quality of work. The controlling department designed a multi-variable evaluation model for monitoring of project leaders' performance. The objective of the performance measuring system is to provide insight on quality of work, enhance motivation and ensure a better allocation of resources. The evaluation is based on nine criteria, seven of which measured qualitatively and the rest – in quantities. There is a different controlling instrument for each category. The frequency of the evaluation varies among the criteria as well. A responsible party for the design and evaluation has been defined for each category. According to the model and the requirements of the management, not every measure has the same importance. Therefore, the weighting of the individual categories' ranges between 0.5 and 1.

The responsible parties have to assign grades for each category and each participant. No categories should be left blank. The grading is done according to the Austrian grading system, with one being the best and five being the worst. After all categories have been completed, one final score is being calculated based on the individual grades and their weightings. The definition of the individual criteria and the ratings follows in the next chapter.

In order to investigate whether a Hawthorne effect can be found in an organizational setting like the one in the experiment, the exact same categories have been rated for 2022. After

that, the research questions regarding the presence of the Hawthorne effect and its influence on qualitative and quantitative criteria have been answered based on the comparison between the results from 2022 and 2023. The year 2022 represents a setting without monitoring and the year 2023 – a setting with monitoring. The objective of this thesis is to demonstrate that monitoring performance, without additional motivators like monetary incentive plans, can increase performance and motivation. This assumption implies that an observation bias exists and it can contribute to the improvement of overall organizational success.

Several forms of monitoring are being combined. The first one is computer monitoring. This is where most of the data is gathered. The most important information sources for the model are: the ERP (enterprise resource planning) system of the company (FileMaker), the BI software Tableau, the personnel database in Oracle, and the company server. No monitoring on real-time activities has been performed, as privacy and independence are important for the company culture.

Furthermore, regular feedback on performance is provided by the controlling department. Individual meetings with the project managers regarding single projects have been set up. In addition, emails with updates and reminders have been sent to the project leaders frequently. The controlling department provided all the necessary working materials for self-evaluation in the meantime. The most common medias used were Excel sheets, checklists, Tableau reports, and custom-made functions in the ERP system.

Another monitoring approach used during the experiment is goal setting. All of the goals are defined in Appendix A and will be discussed in a further chapter. Every grade in the definition of the performance measurement plays the role of an objective. For instance, if a project leader manages to earn a contribution margin by ten percentage points higher than the target calculation, they receive a one as a grade on the budget category. Setting deadlines is also a common approach in the experiment. Similar to goal setting, setting deadlines is integrated in the grading system. This means that normally project team members would get a better grade the sooner they manage to finish the billing process. However, at this criterion there is a room for interpretation as the project staff is always being asked why the delay happened. If there is a plausible reason for the late billing process, the management shows willingness to adjust the payment schedule and the evaluation in the interest of the employee.

Although optimizing processes isn't explicit defined as part of monitoring activities, the CEO of the company and the controlling department believe that in order to optimize the output

(performance), one needs to optimize the process as well. Therefore, different functions in the ERP system have been improved by the cooperation of the IT company, which provides the infrastructure of the ERP system, and the controlling department. The desired functionalities of the program have been defined by the management and controlling and the implementation part was conducted by the software engineers. Examples for the additional functions are the new tools for interim assessment of project development such as a table for a comparison of offered and actual quantities at a certain point of time and a table for invoice submission deadlines. Another improvement in the software was the possibility to import offer items instead of entering them manually. All of those new functions should make the working process easier and, as a result, enhance employee engagement.

Performance measurement criteria

The performance measurement criteria were chosen in alignment with the company's culture and strategy. The management wants to ensure a fair and effective model which considers every aspect of work. Furthermore, there should be no measurement of criteria that cannot be influenced by the employees themselves. A mixture of qualitative and quantitative categories has been introduced which are discussed in the following chapters. While the qualitative categories include more of the subjective opinions of the responsible parties, quantitative criteria are based mainly on numbers. However, numbers often need interpretation, therefore regular meetings have been set up by the controlling department to align the position of the employee with the analysis of the numerical data. On the other hand, qualitative data is often being quantified to achieve a better resource allocation such as time invested in the evaluation.

Quality of data

The quality of data is a qualitative measurement and is one of the ways to evaluate the organizational skills, structure- and transparency of work of the project managers. They are required to enter essential project information into the ERP system. This information includes working time records, target calculations, current calculations, and payment plans.

Time tracking

The time tracking of projects in terms of working hours of the project leaders plays an important role in many personnel and financial processes and is incorporated in the category of data quality. Project leaders are required to enter their working time on the exact projects they have worked, with the exact duration, as the time they needed to

complete the project is then transformed into an expenditure. This step is crucial for the further comparison between target and actual project execution which is the second to last criteria in this multi-variable model. It is also important for additional processes such as forecasting which are not included in the performance monitoring but still essential for the organizational success and development. The history of the company has shown that employees tend to neglect that part of their job, as they rarely see a benefit of time tracking. The management and controlling aim to prove that time records offer transparency and by being done regularly, it reduces the stress and the frustration of doing it with delay.

The controlling instrument for that subcategory is the BI (business intelligence) software Tableau and is operated solely by the controlling department. Once a day, the whole data of the ERP software is being exported in CSV format and is available to update all the ready-to-use reports in the BI software. Therefore, the responsibility of measuring and assessing the quality of time tracking belongs entirely to the controlling division. The quality of time records is measured monthly. As data quality is essential for monitoring of subsequent aspects of work and provides data for further analysis, that category has a weighting of 0.8 out of 1. This is considered to be moderate to high importance (see Appendix C).

The building, where the company is located, is equipped with a time clock. The time clock allows the organization and the personnel to keep track of their working time by swiping cards when entering and leaving. Nevertheless, those times represent just the basic working hours without project assignments. The duration of the recorded time is used as a comparison for the completeness of time tracking and is evaluated as follows. By entering 70-100% of the time clock entries in the following month an employee earns one point as a grade, which is the highest rating, as the aim is to have as lower score as possible. If they have recorded 70-100% of the time clock entries in the first three following months, the employees get two points. For 70-100% of the time clock entries in the first six following months, three points are being assigned. The grade four is given when 50-70% of the time clock entries are recorded with a delay of up to six months. Regardless when the time entry has been entered, if there are less than 50%, the employee is graded with five. Furthermore, for delays longer than six months, a five is assigned as well (see Appendix A).

Target calculations

Another important aspect of data quality is the entry of expected expenses and turnover. The generic term used for that data is “target calculation”. It introduces the basics for the

comparison between target and actual project execution. The target calculation needs to be adjusted by the project manager every time they believe something fundamental has changed. A reason for adaptation would be in that case an extra order by the client or higher costs due to errors. It is important to mention that those changes should be beyond the control of the accountable project leader, otherwise it shouldn't be allowed to adjust the metrics. For example, if the client decides to purchase an additional material for the project and its delivery and assembly contributes to a project delay, an adjustment of the target calculation needs to be made. Another example would be an error caused by a faulty product of a vendor and needs to be repaired. That extra cost shouldn't downgrade the results of the project manager as they have no influence over the events. Therefore, the target calculation needs to be updated as well.

The controlling tool utilized for the evaluation of target calculation is again the ERP system of the company in combination with the business intelligence software Tableau. The responsible party is again the controlling department with a weighting of 0.8 out of 1. This parameter is measured monthly (see Appendix C).

Entering the target calculation in the beginning of the project and updating it, if necessary, brings one point. If the evaluated employee enters the target calculation in the beginning of a project but does not adjust it if later needed, they get a grade of two. By entering the data at a later point than the beginning of the project, a grade three is assigned. Four points are given, when the target calculation is in the systems, but is not entirely accurate. For a missing entry of the target calculation the project leader gets the grade five. This means that the sales department and the project management didn't provide any information at all, and the controlling department needs to take care of that alone (see Appendix A).

Current calculations

To compare the planned numbers with the current ones, the company needs to ensure a high accuracy of the current calculation. In the following experiment, the current calculation includes the purchases of materials and external services, labor and salary costs and the sales of materials, external services and personnel. As the company is reselling high-voltage cables, where the prices change every day according to special indices from the metal industry, there cannot be any fix prices in the system, as they change every day. Therefore, the cable prices should be calculated directly from the purchase invoices. The rest of the materials are being calculated from the internal shipping documents. The ERP software provides a solution that helps employees to compare the purchase invoices and the internal shipping documents. Although the system manages to mark around 80% of

the double entries, that solution needs monitoring as well. Therefore, in order to support the quality management of the company and further monitoring techniques, it is required for project leaders to control their expenditures on a monthly basis. This step aims to promote financial awareness as well.

Another aspect that needs the control of the project leader is the working time of the mechanics at the building lots. As there are always multiple projects going on simultaneously, all of them in different parts of Austria, the only possible way to control the working time of their workers is in a digital form. The mechanics write down the duration of their work and eventually assign it to a project. The project manager must control whether the provided data is correct and if it is assigned to the right project. Only then the work order is considered as approved. This step is essential for further processes such as the payout of overtime, additional legal compensations and resource allocation analysis. It is again analyzed entirely by the controlling department with the help of the ERP system and Tableau BI software. The weight of that category is set at 0.8 out of 1 (see Appendix C).

When no duplicate material entries exist and work orders are released on schedule, the project manager gets graded with one point. For few duplicate material entries and punctual release of work orders, grade of two is being assigned. If there are a few duplicate material entries and a delayed release of time tracking (up to 1 month), the employee gets the grade three. The grade four is assigned if various duplicate material entries and delayed release of work orders exist (up to 1 month). For a delay longer than a month, the project leader gets graded with a five (see Appendix A).

Payment plans

The last subcategory of data quality aims to improve time planning and evaluates the quality of the entered payment plans. Payment plans help the management and controlling to build internal and external reports and forecasts and are essential for growth and success. Moreover, they are the base for further parts of the monitoring system introduced in the experiment. The project leader must enter the month of the planned payment and the expected volume of sales. Similar to target calculations, payment plans can and must be adapted due to several reasons. The adjustment is incorporated in the evaluation of the category. An example for such adjustment would be a shift in the time plan due to a reason that cannot be influenced by the project manager such as a certain delay on the customer side or severe health issues. The evaluation is again conducted with the ERP system and Tableau BI software in monthly intervals and has a weighting parameter of 0.8 out of 1 (see Appendix C).

Entering payment plans on time and updating them, if necessary, brings one point. If the project leader enters the data but does not update it, if necessary, they get a grade of two. By entering the payment plan later than the beginning of the project, a grade three is assigned. Grade four is given, when the billing plan is in the system, but is not entirely accurate. For a missing entry of the payment plan the project leader gets the grade five. If the data is missing it does not mean that the further evaluation of the adherence to invoice submission deadlines will be rearranged. It means that the controlling department needs to find that information and invest more time resources into that process than the project manager would have needed. Hence, not providing the needed information does not exclude employees from the rating, but it makes it hard for them to get a good grade (see Appendix A).

Collaborative interfaces, project takeover and -management

Collaborative interfaces are essential for an efficient workload and a healthy working environment. This variable of the model considers the technical aspect of collaboration. In other words, collaboration is divided into two parts: technical and personal collaboration. The technical part refers to the way in which team members communicate with each other in a digital form. There are certain points that need to be considered when evaluating that category. They are introduced with an assessment sheet that covers all of the essential requirements for collaborative interfaces between the sales team and the project management team and internal regulations for the project management itself. As the project leaders are the ones being evaluated, the focus lies on the internal interfaces within the project management team. Appendix B shows the requirements used for the measurement.

The assessment sheet is divided in nine subcategories which combine the most important aspects of the project management function. The first group, consisting of three questions, refers to the collaborative interfaces between the sales department and the project management. It is essential for an efficient workload, as well as for internal and external communication. If the project leader takes over the project properly in the beginning, it is most likely to reduce uncertainty and unnecessary communication with the clients or with further members of the team during project execution. It is often required to show initiative when acquiring information about the project, therefore it is considered as an obligation that falls within the purview of a project leader's responsibility.

Another important aspect of the project management function is the digital document storage. All of the project team members are required to save relevant documents following

a predefined folder structure. This is crucial for the “digital communication” between the team members, as it reduces searching time and increases independence of work. That group contains of only one question (see Appendix B). Scheduling of personnel resources and their coordination is a specific requirement for the company of the experiment, as it is providing services which need to be planned in advance. One part of this planning includes the working personnel on the construction site. Another important aspect is the procurement of materials and external services. Both, personnel and outsourced goods and services, represent separate categories with two questions each.

In order to execute a project successfully, one needs to prepare before the project begins. This includes paperwork, operating instructions, documentations and organization of necessary work equipment – all of which have been incorporated in three separate questions in subcategory five. The execution itself is covered in subcategory six.

After completion of the project, the team leaders are sometimes required to create documentation for different purposes. It might be required for employees’ training or by the customers. In both cases it is essential to consider that a responsibility of the project leader and is therefore included in the assessment sheet of project management under point seven. Another important interface is between the project managers and the order processing employees and is therefore evaluated in point eight of the form.

The last subcategory, consisting of three questions, refers to the billing process. While in other categories of this monitoring system, the quality of data or the adherence to invoice deadlines are being measured, the focus here lies on “acting in the best interest of the company”. This includes, among other things, a good communication with the clients aiming at customer retention and higher margins.

The quality of project handover is measured and assessed by the head of department once a year. A qualitative analysis has been conducted on each project manager und then used for grade assignment. Although this is considered to be a qualitative criterion, there was still some data that has been quantified for the sake of simplification. All of the single points in checklist were considered with the same priority. The controlling department evaluated the answers of the assessors by assigning one point to the answers with “insufficient”, two points to “average” and three points for “excellent”. The higher the score – the better was the final grading for the category. This category was weighted with the highest priority of 1 (see Appendix C).

For 46 or more accomplished points of the assessment form, the employee gets one point. That means that they managed the technical collaboration perfectly and contributed to a

positive work atmosphere. For less than 46 but more than 36 requirements fulfilled, the employee gets a grading of two points. From 26 to 35 points of the checklist a grade of three is being assigned. Between 21 and 25 points of the sheet one gets graded with a four. Everything below 20 points means a grade of five (see Appendix A).

Collegiality and team cooperation

The personal collaboration incorporated in the model was evaluated using a behavior assessment scale. This type of scales is often used in organization to assess employees' performance, skills or specific behaviors. In this experiment, the goal is to obtain a simple evaluation of a complex topic such as collegiality and team cooperation. Five different levels of collegiality have been defined, starting at one as the best and five – as the worst. The assessment is considered to be entirely built on the subjective opinions of the management of the company and the head of department and represents a qualitative criterion, measured once a year. The controlling department supported the definition of the levels one to five in an individual meeting with the head of department. They first outlined the profile of a person who shows the “perfect” collegiality and team cooperation and then scaled it down to the last level gradually.

Under the best circumstances, collegiality and team cooperation were defined as follows. The person is always reachable on the phone, both during and if needed after working hours. They respond quickly to emails and are available in person at appropriate intervals. They are helpful and actively support solving internal challenges and engage in the achievement of departmental and company goals. If a person meets all or most of the above, they are assigned with a one.

The level two reads as follows. The project manager is generally reachable on the phone and reacts to their emails promptly, their face-to-face availability is sufficient. Some willingness to support internal problems is demonstrated and working towards strategic goals is present, but not always proactive.

The level three reads as follows. The employee is mostly reachable on the phone and responds to emails in an acceptable time frame. They are sometimes available in person as well, although not in optimal intervals. The willingness for problem solving and goal achievement is demonstrated occasionally, but not proactive.

The level four reads as follows. The person is only occasionally reachable on the phone and reacts to emails with delays. Face-to-face availability is not regular and their

willingness to contribute to problem solving is infrequent. The involvement in goal achievement is limited.

The level five reads as follows. The project leader is very rarely reachable on the phone and tends to neglect emails. They are almost never meeting their colleagues in person. An apparent willingness to support problem solving is missing and they contribute to departmental or organizational goals just minimally (see Appendix A).

This description is based on the experience of the CEO and the head of department with unsatisfactory behavior of certain employees. The company has a history of non-compliance with the organizational rules and wants to improve in that area. Therefore, this category has a weighting 0.8 out of 1 (see Appendix C).

Customer satisfaction

The category of customer satisfaction has two sources for evaluation. On the one hand, several customer satisfaction feedback forms have been collected on the website of the company. At the end of every year customers are being asked to fill out a form, stating how satisfied they are with the service of the person in charge for their projects. In that way, the controlling department was able to collect specific data on customer satisfaction for the project leaders separately. However, not many clients have returned that feedback form and it was necessary to implement a second way to measure satisfaction. A behavior assessment scale was defined by the CEO, the head of department and the controlling department. The scale is built up the same way as the scale for collegiality and team cooperation. Five different levels of satisfaction have been described, starting at one as the best and five – as the worst. The assessment is considered to be influenced by the subjective opinions of the head of department and the CEO together with the verbal feedback they have received from their clients throughout the year. Therefore, in this experiment customer satisfaction is categorized as a qualitative criterion and is measured once a year.

The best scenario regarding customer satisfaction and service quality reads as follows. The customers are extremely satisfied with the support of the project manager. They receive quick and efficient solutions to their concerns. The work performed is considered excellent.

The grade two is being assigned if the customers are very satisfied with their support and most of their inquiries are processed quickly and to their satisfaction. The work is recognized as high-quality and reliable.

The middle of the scale describes the simple satisfaction with the service. The concerns of the clients are generally handled appropriately, although there may be occasional delays. The service performed is considered acceptable.

The level four reads as follows. The customers are not completely satisfied. The handling of customer requests is irregular, and customer support is not always consistent. There is room for improvement.

Level five is characterized by an extreme customer dissatisfaction with the received support. Requests are often neglected, and the work is reported to be unreliable and unsatisfactory. Urgent revision is required (see Appendix A).

This definition is based on the strategic positioning of the company in the market. With rising competition and prices in the last few years, customer satisfaction has become a crucial factor for retaining customers. There are various approaches used for maintaining customer relations in the company. However, they will not be closely discussed in this thesis, as the focus lies on the performance of project managers and their potential contribution to customer retention. Despite its big role in the accomplishment of departmental and company goals, that criterion has a weighting of just 0.5 and is considered a low priority (see Appendix C). The reason for that is the low participation rate of the customers in the feedback surveys and the highly subjective nature of the behavior assessment scale.

Improvement of contribution margins

The next criterion covers the financial aspect of the model. It refers to the crucial economic prosperity of single projects, the department and the whole firm. The company considers the improvement of contribution margins to be an accurate measurement of financial success and extraordinary performance. The contribution margins are key financial metrics that represent the difference between total sales revenues and variable costs. The improvement of the contribution margin is calculated by comparing target calculations with current calculations at the end of each project.

In this category, the company refrains from setting specific targets for the contribution margins. One of the reasons is because the project managers often must work with predefined and non-negotiable selling or buying prices, which reduces their influence on the size of the contribution margin. Another reason for the abstention of target margins is the fact that the company is engaged in a high volatility market, where prices are often

influenced daily by external factors such as economic crises, inflation, fluctuating commodity prices.

Instead of entirely neglecting the financial aspect of day-to-day business activities, the CEO defined the percent-wise improvement of contribution margins as a reliable metric for the evaluation of financial prosperity. It is worth mentioning, that strategic decisions are occasionally being made by the CEO and the sales department which include projects with negative contribution margins. In order to remove the disadvantage for the project leader who takes over a project with a predefined loss, the company implements the rule that even a negative contribution margin of a finalized project might have a positive evaluation if its profitability exceeds the target margins of the sales department.

The improvement of contribution margins is considered with high importance and therefore was assigned a weighting of 1, which is the highest one in this experiment. It is defined as a quantitative measure of performance (see Appendix C). However, similar to other categories, the controlling department was not relying solely on the metrics collected from the database and the reports. During the individual meetings with the project leaders, the calculated contribution margins were presented to the responsible parties and their consent was obtained. If they were not to approve the calculated margin, substantiated modifications were made on the target calculation in order to ensure a fair comparison between target and current values. This highlights the idea that almost every metric classified as a qualitative measurement owns quantitative features, and vice versa.

If the project leader manages to increase the target contribution margin by over nine percentage points, they are assigned the best grade, which is one. For an improvement between six and nine percentage points, a grade of two is being assigned. The middle part of the model requires an increase of contribution margins between three and six percentage points and brings three points. If the project leader closes the project with an improvement of up to three percentage points, they get four points. For no improvement or a deterioration of contribution margins, the project manager receives the worst grade of five. The final grade of the current projects is compared to similar projects of the past. The project volume and the approximate wage-to-material ratio are utilized as a base of comparison (see Appendix A).

Adherence to invoice submission deadlines

Another way to ensure financial prosperity is the adherence to invoice submission deadlines. In other words, not only the amount on the invoice but also the date is important. One of the reasons is an efficient liquidity planning. The financial department needs to

create forecasts and budgets and manage the current liquidity of the company; therefore, it requires up-to-date information regarding the billing processes of the projects. Furthermore, it is essential not only to enter the invoice submission deadlines but to adhere to these as well.

The criterion regarding the adherence of payment plans is categorized as a quantitative measure of performance because the evaluation is conducted mostly objectively with the help of exact numbers. However, before the final grading, the project leaders have been asked if the payment plans need to be adjusted due to occurrences that lay beyond their control. For instance, such reason would be an additional request of the customer that was communicated at a later point. If this order addendum contributes to the delay of the invoice submission, it is allowed to expand the deadline in favor of the project manager.

The weighting of the adherence to billing plans is considered with low importance due to multiple reasons. This criterion is evaluated once a project has been finalized. Although it is crucial to promote financial literacy on organizational level, the CEO of the company recognizes that many of the influences on the billing process are beyond the control of the employees (see Appendix C). Furthermore, the decision for the low weighting of the category was made in order to avoid a drawback in quality of work. The company in this experiment places a higher value on quality of service, market positioning, customer retention, and long-term goals. Therefore, small delays in the billing process are being tolerated for the sake of more important aspects, mentioned above.

If a project manager maintains the scheduled invoice deadline on time or with a delay of up to one month, they are assigned a grade of one, which is the highest score. The next level of two is scored at a schedule maintained with a delay between one and two months. In order to get graded with three points, the project leader has delayed the billing process up to three months. For invoice submission between three and four months, the employee has four points in this category. Anything beyond four months delay is graded with a five (see Appendix A).

Controlling instruments

The controlling instruments used for the execution of the experiment vary in their purpose, characteristics and standard operators. The main tools utilized are the ERP system File Maker, the BI software Tableau, Microsoft Office and an Oracle data base, customer satisfaction assessment forms, behavior assessment scales.

ERP system

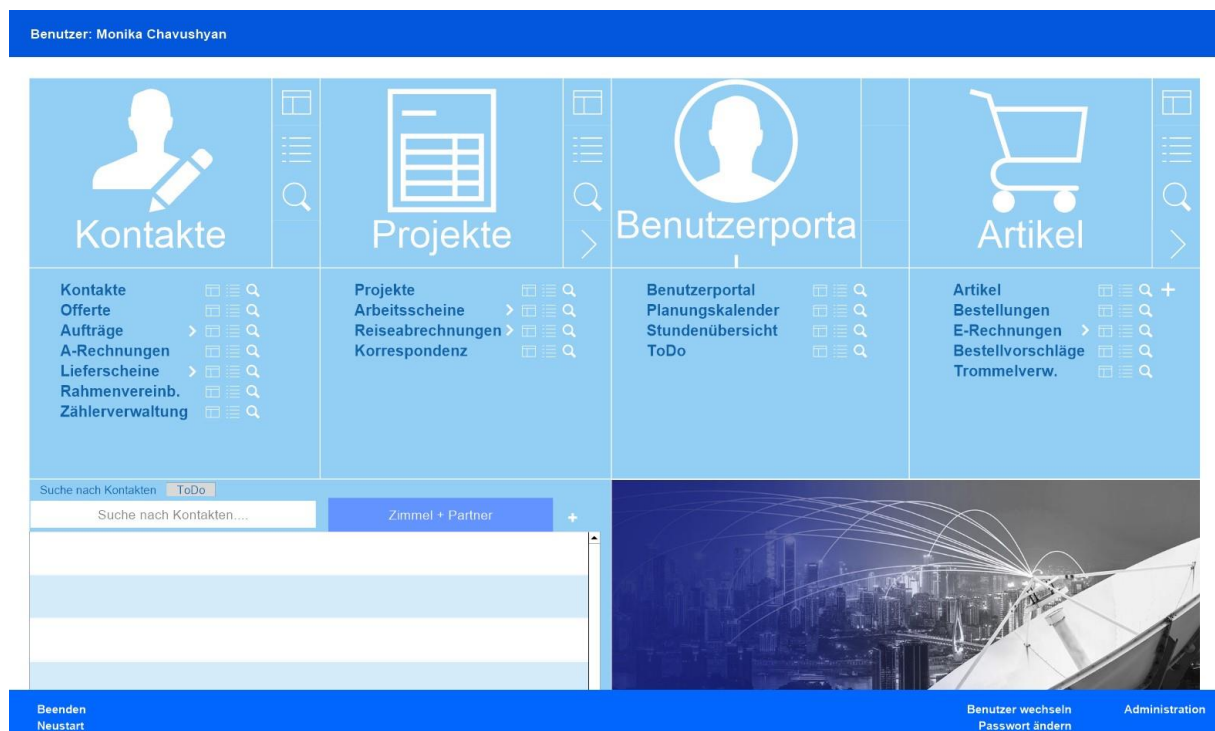
The ERP system consolidates multiple purposes. It supports the business process from the beginning to the end with small exceptions. Beginning with the custom relationship management (CRM), to writing quotations and contracts, all of those steps incorporate into project management and are crucial for the evaluation. Managing the product catalog, booking incoming invoices, and billing for services are further processes that need a regular monitoring which was managed by just monitoring the project management itself. The accounting of the company is performed in another software called BMD, which was not explicitly used for the purposes of this master's thesis.

For the experiment it is relevant and important to control the entries in the system, as they lay the basis for further analysis and are essential for other processes as well. Important entries for the experiment are the time record, the target calculations and the payment plans.

Another purpose that the ERP system serves within the organization is the data preparation and export. Every day between 19:00 and 02:00, the entire dataset is being extracted from the ERP system and stored on the company's server. There are 21 CSV files that are being exported every day, which makes up to 200 MB just from simple text. Those datasets are then used by the controlling department for further analysis.

What is also important and being measured for the experiment is the interim control of project data, which again takes place mainly in the ERP system. This is where the project leaders need to control whether all of the planned materials and external services have been deployed. Furthermore, this is where more simple evaluations of the target and current calculations, time records and payment plans are being executed by the controlling department.

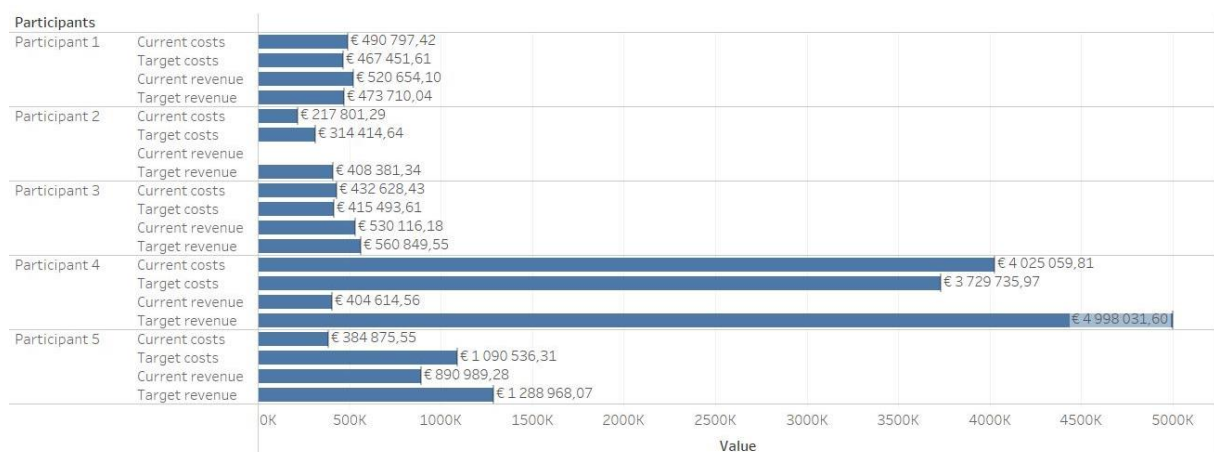
Figure 1: ERP-system Home page



Tableau

The datasets from the ERP system arrive at the controlling department and are most often being imported in the BI software Tableau. As the data is always up-to-date, once the report is being established in Tableau, one only needs to update the dataset and already has the old standard report with the new data. This is where the most important connection between data quality and monitoring lays. The ERP entries represent the data quality and the Tableau software represents a part of quantitative and qualitative monitoring.

Figure 2: Tableau report “Comparison between target and current calculations”



Oracle personnel database

As mentioned, the company is contracting mostly on public tenders. Therefore, there are multiple technical requirements that need to be met in order to obtain an order. The technical requirements that are related to the personnel refer to different trainings and certificates that the assembly operators need to acquire. Those certificates have different purposes, categories, providers, and validity periods. Some construction sites cannot be entered by the staff of the company if it does not have the necessary certificates for the execution of the work. As a result, the firm always has to keep an eye of those external trainings and make sure that none of them has expired and the employees have the required credentials for the planned projects in the future.

In order to optimize that process, the controlling department established an additional module in the human resources database with the cooperation of the head of energy department and an external software engineer. That module provides an overview of every training of every person that has been done until today. In addition, one can easily control which are the next pending certificates and plan the required resources accordingly.

However, as every other controlling instrument, this database also needs to be maintained and filled with data. Hence, the controlling department, together with the management and the head of energy department have decided to embed the aspect of database maintenance into the monitoring system. The evaluation is conducted on how frequently and how completely the assembly documentation is filled out.

Microsoft Office

The next tool is the Microsoft Office program Excel. It was used on multiple occasions for ad-hoc analyses and data preparation. Spreadsheets have been used for the primary definition of the model and the final grade assignment. Microsoft Word has been utilized for the creation of checklists while Microsoft Excel was used for their evaluation.

Responsible parties for evaluation

The people who were chosen for the evaluation of the single categories are either the ones who have the better observation on the measured activity or the required knowledge and experience with the certain controlling instrument. The controlling department is responsible for collecting all the data from the ERP system and importing, analyzing and interpreting it in Tableau. As mentioned above, this is something that cannot be done by anyone else due to the lacking experience with the BI software. However, the controlling division summarizes the data and presents it to the further responsible parties. During the

evaluation period there was a consistent communication between the controlling department and the project leaders. Sometimes aiming at clarification and others – to provide feedback.

The next responsible party is the head of energy department. The criteria definition was developed through a collaborative work, incorporating both inputs from the controlling department and the energy department manager. He¹ is responsible for the operative part of the work, knows his personnel and short- and middle term department goals. During the experiment he was responsible for the evaluation of the proper project takeover, management and documentation, collegiality, team cooperation and customer satisfaction. When necessary, the controlling department was maintaining the data collection for those categories as well.

The final accountable party is the manager of the company. He has a supporting function and has the final decision if the grading varies between two levels. There is no category which is evaluated only by him, but there are some where he collaboratively assigns his grades with the other two parties. The management is involved in the evaluation of collegiality, team cooperation and customer satisfaction, including the head of department. The analysis of the improvement of contribution margins and the adherence to invoice submission deadlines was conducted by the controlling, under the continuous communication with the management.

Weighting

The weighting in the experiment represents the perceived importance for the company and the ability of fair evaluation. For a very important criterion, that can be measured very accurate, the full weighting of one has been assigned. On the other hand, if the category cannot be easily targeted by the project leader in the short term, is hard to evaluate or has a moderate importance for the company, the lowest weighting of 0.5 has been assigned. There are three tiers of assessment in total: 1, 0.8 and 0.5.

The categories with a weighting of one are: the collaborative interfaces, project takeover and -management as a qualitative criterion and the improvement of contribution margins as a quantitative performance indicator. The data quality together with collegiality and team collaboration are considered to have the second high priority and therefore a weighting of 0.8. Both of the categories are evaluated qualitatively. However, the objective is to make a quantitative analysis of data quality in the future for the sake of better resource allocation.

¹ Gender-neutral writing was omitted on multiple parts in this thesis in the interest of maintaining simplicity

The last level weights 0.5 and includes the adherence to invoice submission deadlines and customer satisfaction. It is worth mentioning that customer satisfaction is one of the most important indicators for the company of the experiment, but as it cannot be measured with a high accuracy, the controlling department and the head of energy department assigned the lowest weighting. The main reason for the problematic measurement of customer satisfaction is the low participation rate of the customers in the customer feedback acquisition. The adherence to invoice submission deadlines falls into the category with the lowest priority as well, because the management decides to focus more on quality of work and customer relations than on a rushed billing process.

Definition of the grades

The grades were designed in accordance with the short- and long-term objectives of the company. The initial idea was born in 2021 when the first pay-for-performance model in the company has been established. It was introduced in another department, had a different target group, controlling instruments, and measurements. However, the goal was the same – enhance motivation among employees. In this experiment, we explore the effect of monitoring on performance without introducing a further motivator in the form of incentive plans. Nevertheless, the management of the company together with the controlling department use the trial as an experiment to compare to what extent monitoring by itself manages employees' motivation. A further step in the plan would be introducing a monetary reward in return of a good score resulted from the observation, but this is not a matter that will be considered in this thesis.

Aiming at a better data quality in terms of time tracking, target calculations, current calculations, and payment plans, ensures a better controlling-project management interface. This, on the other hand, points at an improved quality of work, transparency and motivation. Furthermore, with an enhanced data quality, the company manages to expand monitoring mechanisms which support transparency and motivation as well.

A proper project takeover and management represent the capacity of the project leaders to do their jobs correctly. The incorporation of that aspect into the monitoring system is essential to demonstrate employees that the quality of their work matters and it is being monitored. Furthermore, aiming at the improvement of project management itself is likely to carry additional improvement with itself.

For instance, if a project leader is able to run their projects in a manner that satisfies all the criteria of this model, they will also be able to achieve a higher customer satisfaction. Customer satisfaction is not only a category in the monitoring process, but is also crucial

for many business processes as well. It was built in the grading system with the strive for a better service quality and customer retention, which on the other hand ensures a long-term success by providing higher sales volumes and a better market position.

The financial aspect is represented in another criterion as well – the improvement of contribution margins. The grading was set very tightly, aiming at an enhanced motivation. The establishment of a precise measurement system with little room for leniency has proved to improve employees' attention to performance and henceforth – their outcomes (Saez-Marti & Sjögren, 2007). The history of the company shows a tendency for employees neglecting financial success, therefore the monitoring system together with the strict standards are expected to increase profitability by several percentage points.

The billing process also refers to the financial part of the business process. Although that category is not directly linked to the amount of money that is being invoiced, liquidity and forecasting also have the highest priority for the firm. However, the management wants to evaluate their employees based on their performance and matters that can be directly influenced by themselves. Hence, the grades in the category of “adherence to invoice submission deadlines” were defined with a more lenient and less strict approach.

The last category is the collegiality and team cooperation. This is an entirely subjective category based on the direct supervision of the management and the head of department. While it is very important to provide a safe and cooperative environment for your employees, companies often fail to address that part of the work by neglecting its importance. Moreover, it can be hard to influence that behavior, as it refers only to the soft skills of a person and has often very less to do with their qualifications. For instance, a very smart and successful person might be lacking basic social skills. The company in this experiment recognizes the importance of collegiality and team cooperation and has developed an exact definition of the desired soft-skill profile. That profile was then gradually changed from one to five as a description of each level of the scale.

Participants of the experiment

There are five participants in total, designated as Participant 1 through Participant 5. In the next paragraph each one will be introduced based on their age, gender, time in the company, and education level.

Participant 1 is male, 34 years old and works for the company for approximately one and a half years. His highest level of education is a master's degree in engineering. He is known for his attention for detail, therefore is expected to be good in things like documentation

and data quality, but open to improvement in categories like customer satisfaction and overall project takeover.

Participant 2 is male, 49 years old and works for the company for approximately 20 years. The highest level of education that he has is a higher technical vocational school diploma with an engineering title. He used to be head of department for around five years, before a colleague of his took over the position. Ever since, he has not shown much motivation to develop and grow. As he has a long-term experience in the project management of the company, it is expected for him to score good at the project takeover aspect. However, his current data shows that he is often lagging behind in things like time tracking and customer satisfaction.

Participant 3 is male, 52 years old, and works for the company for approximately 24 years. However, only in the last five years he was executing the job of a project leader. Before that he was a mechanic worker on the constructions site. He is the most competent employee in terms of technical work. This means that he is expected to be good with the project takeover and management. His higher education is a secondary school diploma. Nevertheless, the late entry into the project management field might be an obstacle on his way due to lack of experience in several aspects such as email management or software skills. Therefore, it is expected for him to be falling behind on the part of data quality and contribution margins.

Participant 4 is male, 29 years old and works for the company for approximately 8 years. He holds a master's degree in engineering and economics as his highest level of education. Due to his economical background, an enhanced interest and understanding of categories like the improvement of contribution margins and the adherence to invoice submission deadlines are expected from him. However, due to his young age it is assumed that he might be lagging behind on customer satisfaction.

Participant 5 is male, 48 years old and works for the company for approximately 8 years. The highest level of education he has is a master's degree with an engineering title. He is responsible for the most technical specifications of the company and is a highly qualified employee. The quality of his data is expected to be good. On the contrary, there might be some room for improvement in the category of "collegiality and team cooperation".

The summary on the type, description, frequency of measurement, controlling instrument, responsible parties, weightings and points of each category is represented in Appendix C.

Data collection

The data was collected separately on each category. While the evaluation was in the responsibility of the management, the head of energy department and the controlling department, the data collection on the separate criteria was conducted entirely by the controlling department.

The collection of information on data quality took place regularly since 2022 in the Business Intelligence software Tableau. Different reports have been prepared and updated when needed in order to collect information on that how fast, how accurate, how often, and how complete the data has been entered in the ERP system. Regarding the project takeover and management, several meetings and emails have been used between the controlling department, the management, and the head of department. A blank form for the evaluation of the project management has then been sent to the CEO and the head of energy department. It was then returned to the controlling department with a filled evaluation for 2022 and 2023. This data was then analyzed by the controlling department and filled in the final form.

The data collection on collegiality and team cooperation was conducted mostly per email. The grades have been directly assigned by the management and the head of department in the final table for evaluation, including the years 2022 and 2023. The customer data was gathered in two different ways. The first one is the customer satisfaction form, filled by the customers with a reference to the concerned employee. However, those were often not sufficient for the evaluation and were therefore supported by the subjective opinion of the head of department and the management of the company. The data was collected once in 2022 and once in 2023.

Whether the project leader has managed to improve their target contribution margins or to adhere to the invoice submission deadlines, was being observed in the BI software Tableau entirely by the controlling department. The data was collected after a project has been completed, again comparing projects from 2022 to such from 2023.

Data analysis method

Several data analysis methods have been used for the evaluation and interpretation of the gathered information. On the one hand, a comparison was made between the data of each category in 2022 and 2023. On the other hand, the results of each participant were compared for 2022 and 2023 as well. The statistical metrics used in that part of evaluation are variance, (weighted) average, and the first, second (median), and third quartile of the

data set. Further measures used for graphical representations are the minimum and maximum values, upper and lower outliers, and upper and lower whiskers. When possible, boxplots have been used for the graphical analysis of the data. However, some of the results were not suitable for illustration and were reviewed only according to the numerical data.

A histogram was used to display the frequency distribution of the grades in both years 2022 and 2023. Moreover, the correlation between all of the variables was calculated for the two periods, but not much relevance for the research questions was to be found in the results. The significance of the findings was determined by a Wilcoxon signed-rank test. Further metrics include the number of observations, sum, rank, test-statistic, and critical values. Due to the small size of the data set, the analysis was conducted entirely in Excel and it was abstained from more complex programs like SPSS.

Simple mathematical methods were often used in the single categories for evaluation of the data. The data quality of time tracking was measured by the average delay of the time entry. The average delay was calculated by comparing the date of entry with the date of the work assignment. For instance, if until September 2022 the project manager has not entered his working time of June 2022, they will have a delay of three months, which classifies for the grade of two if the entries were complete. The delay was calculated and rounded in months. Then the delay was multiplied by the duration of the working time, as the quantity of hours entered in the system represents another important side of this criterion. This multiplication results in a score that is used for the weighting of the delay. For example, if a project manager has entered 50 hours of his working time with a delay of 4 months and 200 hours – with a delay of one month, his average delay would be 1,6 months, rounded to 2.

A delta between the accomplished and the planned contribution margins was calculated in order to evaluate the employees' financial performance. If the difference was positive, it means that the project leader managed to improve the target calculation and if it was negative, it means that they finalized the project with worse profitability than planned. The deltas of each participant for each year have been compared in order to gather information regarding a potential improvement due to the newly implemented monitoring system.

For the total grading of the participants, a weighted assessment system was utilized. The sum product of all the grades of the single categories and their weightings has been calculated per participant and then divided by the sum of the total weightings. In that way,

an average of every category was obtained considering the assigned importance. This method was used again for both data sets of 2022 and 2023.

Ethics and data privacy

The company in this experiment has a privacy policy in a written form. It states that the personal data of the employees is confidential and cannot be shared with third parties. In order to ensure the conformity of that clause within the employment contract, there were six consent forms that have been distributed among the participants, including the CEO of the company. On the one hand, the informed consent form of the manager of the company states that no names of employees, customers, projects or the company itself will be disclosed in this thesis (see Appendix D). On the other hand, the participants are informed by a slightly different document, that their data will not be used only anonymously and are required to sign a consent that their participation is voluntarily (see Appendix E).

Experiment execution

Participant recruitment

The participants that were selected for the experiment have all been in the department for a year or longer. All of them are executing a project management function and have project data from a year ago or older to compare with the new data. There was no restriction on the age, gender or education level of the employees. However, as there are no female executors of this function at the company, only male participants have been included. The reason for restricting the department are the differences both telecommunication and energy department have in terms of billing processes, technical specifications, duration of projects, and more. Furthermore, the telecom division already has an implemented monitoring system and the energy division – does not.

Informed Consent

An informed consent was obtained from the participants in November 2023. A document where the most important aspects regarding the data of the participants and the objectives of the study was printed out for each project manager that participated in the experiment. Firstly, the background and the study objectives of the experiment are introduced, followed by the voluntary nature of involvement. Then it is described how the study objectives are being achieved by presenting the study procedures. The next point states that the collected data is treated confidentially and anonymously. The last two points pertain to contact information – firstly the participant data such as age, highest level of education and duration of employment with the company, and secondly - the contact details of the person

responsible for the execution of the experiment. Finally, a signature and a date were placed at the bottom of the document, stating that the participation in the study is entirely voluntarily and all of the questions have been answered to a full satisfaction.

Experimental Procedure

The first step towards the experiment was the decision of the management to invest more resources in monitoring performance and enhancing motivation. The decision has been taken in December 2022 in a small circle of three people: the management of the company, the controlling department and the head of human resources. There were several consequent meetings between the head of energy department and the controlling. In those meetings, the main goals and their measurements have been defined.

That step was followed by meetings in person between the controlling department and each project leader individually. The idea of monitoring project success was introduced together with the expectations for enhanced motivation and better performance. Those meetings took place in February 2023. The output of the sessions was then subsequently summarized in an email to each participant. The email stated the most important parts of the experiment such as the fact that project data needs to be provided in a proper form in order to continue with the evaluation, the fact that the project needs to be finished until the end of June 2023 and that employees' behavior is now being monitored. The two example projects have also been selected during the meeting and mentioned in the email as a reminder to update the data.

The project leaders have been supported continuously by the controlling department by organizing training sessions, writing emails with descriptions, setting up new functions in the system and sending regular reminders. That phase took place between March 2023 and September 2023. It was the longest part of the experiment due to two main reasons. Firstly, some of the project leaders did not manage to close their projects in the time frame that has been set in the beginning. In some cases, this was attributed to factors beyond the influence of the participants. Secondly, the personnel needed a considerable amount of training hours, which outstripped the original time horizon of this phase. In the meantime, all the necessary reports and additional instruments in Tableau, Excel and Word have been set up.

In October 2023, the final evaluation of the criteria "project takeover and management", "collegiality" and "customer satisfaction" took place. An empty form was sent to the manager of the firm and the section leader of the energy department regarding the rating of the quality of project management. By asking whether points 1-17 are "insufficient",

“average” or “excellent”, the responsible parties for the assessment have given their subjective opinion on topics that cover almost the whole project management process. The questions in the form refer to essential parts of the job of a project leader – from the interferences with the sales departments, folder structure and personnel planning, through procurement of goods and services and supervision of the site, to the interface with the order processing and invoicing. The filled form came back per email in the middle of November 2023.

The behavior scale for the evaluation of cooperation and collegiality was entered directly in the final evaluation table by the CEO and the head of energy department. This happened again, via email in the middle of November 2023. The evaluation of the customer satisfaction was conducted similarly to the behavior scale for team cooperation in the middle of November 2023. The only difference here was the fact that in some cases for some employees there was personal feedback from customers. This is something that the company was already monitoring and collecting every year due to several regulations including those of the ISO quality management standards.

After all the evaluations have been gathered, the final scoring table was formatted and evaluated as well. By considering the weighting of each category, a grade was assigned to each participant. The same method was used for the assessment of the 2022 and 2023 grades. After every category and every participant was graded, an analysis and interpretation of the results was carried out by the controlling department.

Results

Data Presentation

The detailed outcomes of the experiment will be presented in the following chapter. Each category will be reviewed separately. The grades for each participant will be provided and compared with the baseline before monitoring. If a more in-depth analysis of the criterion was conducted, it will be discussed in the following chapter as well.

Data quality: Time tracking

The category of data quality was evaluated qualitatively. Nevertheless, similar to most of the measurements, there was numerical data which required revision before proceeding to the interpretation. The information gathered in the category of time tracking is summarized in Table 1.

In 2023 Participant 1 has entered 99% of his working time in the system with an average delay of zero months. According to the assessment scale, this applies to a grading of one. The same employee has put 96% of his working hours into the system, again with no delay. To sum up, Participant 1 was assigned a one for both settings – with and without monitoring.

Participant 2 managed to record only 35% of his project work with an average delay of three months. These results have brought him a grade of five in 2023. In 2022 he entered 21% of his working time in the system with an average delay of one month. Although the delay was minimal, he does not classify for a better grading, as the minimum number of hours entered in the system has not been achieved. Therefore, Participant 2 was assigned a grade of five in 2022.

Participant 3 documented 77% of his project work with an average delay of three months in 2023. This corresponds to a grade of 2. In 2022 the same employee logged 79% of their working time in the system with an average delay of four months. Consequently, Participant 3 is given a grade of four for the setting without monitoring in 2022.

Participant 4 entered 59% of his working hours in the system for 2023. His average delay was then three months, which leads to the grade of four. In 2022, 75% of his working hours were recorded in the ERP system with an average delay of one month. Therefore, Participant 4 is assigned a grade of one for 2022.

Participant 5 has entered only 3% of his project working hours in 2023. The delay here is one month. In that case the average delay is irrelevant, because the minimum of 50% working time entries has not been achieved. The setting without monitoring from 2022 has brought him a grade of five as well, as he entered only 7% of his working time.

Data quality: Target calculations

The evaluation of the quality of target calculations was performed entirely through qualitative means. A main part of the assessment took place in person or via email. Figure 2 shows the Tableau report used for the final analysis and control. This instrument was used for both quality of target calculations and improvement of contribution margins. For 2022 everyone is assigned a grade of five, because in the past, employees have never provided their target calculations. However, in the year, where the monitoring system was established, there was a considerable improvement in this area.

Participant 1 was assigned a grade of one, as they delivered the required data on time and adapted it if necessary. They were cooperative and their data was accurate. Participant 2 was assigned a grade of three. Despite the precision of the target calculation, their data was provided with a delay. Participant 3 is given a grade of five, because they never provided any target calculations. Participant 4 received a grade of one. Their data was reliable, on time, and they were helpful and supportive when needed. Participant 5 was assigned a grade of two. They provided their target calculations but did not pursue the task actively and needed to be repeatedly reminded of deadlines and updates (see Table 1).

Data quality: Current calculations

In 2022, before the monitoring system was introduced, three out of five participants were assigned a grade of five (see Table 1). Those are Participant 2, 3, and 5. The grading of five was assigned due to various duplicate material entries and delayed authorization of work orders (more than one month). Participants 1 and 4 were the only ones who received a slightly better grading. Participant 1 is given a grade of four due to multiple inconsistencies in the material listing of the projects and delayed authorization of work orders (up to one month). Participant 4 has a grade of two which implies to only a few duplicate material entries and a punctual authorization of work orders.

In 2023, Participant 1 was assigned a grade of two for only few duplicate material entries and delayed authorization of work orders (up to one month). The same grade was assigned to Participants 2, 3, and 5. Only Participant 4 got a grade of one for having no duplicate material entries and no delay in work orders authorization.

Data quality: Payment plans

Similar to the target calculations, the quality of data of all participants regarding the payment plans of the projects was rated with a grade of five. In 2022, before the monitoring on data quality was introduced, payment plans were only known by the few people involved in the project. Furthermore, this information was not systematically entered anywhere, so it remained in the correspondence emails of singular employees. The grades were assigned to all of the participants for the setting without monitoring, based on this information and experience.

After introducing the model, Participant 1 has been assigned a grade of one. This implies to a satisfactory quality of payment plans data. The person has entered the payment plans and has updated them when changes occurred. The same score was assigned to

Participants 2, 4, and 5. In 2023, Participant 3 got the grade of three, which means that they entered the payment plans, but on a later point in time (see Table 1).

Collaborative interfaces, project takeover and management

In the category of project management, a thorough analysis was executed beyond the grading scale between one and five. The results from the form for each participant and each year were first evaluated by counting the points that the person got on each question. “Insufficient” brings one point, “average” - two and “excellent” – three points. The higher the score – the better was the final grading for the category.

Participant 1 achieved 40 points in the setting without monitoring. This qualifies for the grade of 2. In 2023, with monitoring, a total of 42 points was collected by the same employee, which has brought them again a grade of two. Participant 2 had 32 points in 2022, which means a grade of three. The same grade was assigned in 2023, but for the lower score of 28 points. Participant 3 achieved 37 points before the monitoring system was introduced and got a grade of two. In 2023, they as well got the grade of two, but for 36 points. Participant 4 collected 45 points before knowing that their ability to lead a project was being monitored. This implies to a grade of one. One year later, they achieved a higher score of 48 points, which has brought them again the grade of one. Participant 5 had 38 points in 2022 and 37 points in 2023. Those results correspond both to a grade of two. The results of this category are summarized in Table 1.

Collegiality and team cooperation

The performance of the participants in the category of “collegiality and team cooperation” was evaluated through the custom-defined behavior assessment scale by the CEO and the head of energy department.

Participant 1, 3, and 5 received a grade of one in both years, corresponding to the description of a highly helpful and supportive colleague with optimal availability. In 2022, the year before monitoring, Participant 2 was assigned a grade of two, corresponding to the second-best level of the scale. According to the assigned grade in the following year, their collegiality and team cooperation declined after monitoring was introduced. Participant 4 managed to improve the perceived collegiality and teamwork after monitoring resulting in a grade of two in 2022 and a grade of one in 2023 (see Table 1).

Customer satisfaction

The evaluation of the customer management is based on the subjective opinions of the CEO of the company and the head of energy department. There was no numerical data to analyze, as the grades have been assigned according to the pre-defined behavior assessment scale for this category. The summary of the grades can be found in Table 1.

In 2022, before the monitoring system was introduced, Participant 1 was assigned a grade of two. This implies to a high quality of work perception of the customers, reliability and mostly a prompt reaction to customer requests. His grade has not changed in 2023. Participant 2 got a grade of three in 2022. His customer support was classified as satisfactory but not exceptional. Occasional delays have been reported. After the monitoring system was implemented in 2023, his grading on customer care has declined. The support on the side of the customer was perceived as inconsistent and multiple delays have been reported. Participant 3 got the grade of three in both 2022 and 2023. Participant 4 received the best grade of one for both years 2022 and 2023. The grade of one was assigned when the customer service was considered excellent, quick and efficient solutions were proposed to customer concerns. Participant 5 accomplished a grade of two for both settings with and without monitoring.

Improvement of contribution margins

Participant 1 has achieved a delta of -12,01%-points in 2022. This result refers to a grade of five. On the contrary, after introducing the monitoring system in the department, Participant 1 managed to close a project with an improvement of 4,23%-points. This has brought them a grade of three. Participant 2 has a calculated delta of -58,92%-points for 2022, which classifies them for the lowest score. In 2023 they achieved an improvement of 6,02%-points and were therefore assigned a grade of two. Participant 3 finalized their observed projects with a negative delta of -14,59%-points, bringing them the grade of five in 2022. In 2023 they have shown an improvement compared to the previous year. However, this was not enough for a positive score. Participant 4 managed to close their project with a positive change in the contribution margins before introducing the monitoring system. The improvement was calculated at 6,08%-points and corresponds to the grade of two. In 2023, the same employee managed to finalize their projects with higher margins than expected. However, the calculated delta of 4,30%-points equates to the grade of three. Participant 5 had a delta of -9,90%-points in the year without monitoring and was assigned a grade of five. For 2023 they managed to exceed the target calculations by 3,30%-points, which is equal to a grade of three. The findings of this category are summarized in Table 1.

Adherence to invoice submission deadlines

For each of the projects in the previous category there was a payment plan that needed to be entered in the system. Using that data, the controlling department has calculated an average delay for each participant once for 2022 – the setting without monitoring and once for 2023 – the setting with monitoring. The summarized data is shown in Table 1. The average delay was calculated in Tableau by subtracting the today's date from the planned date of the billing plan. The result was then divided by 30 and rounded to zero decimal places. This calculation was performed on each step of the billing process of each project and was then summarized for the individual participants in the experiment, comparing 2022 and 2023.

Participant 1 had an average delay of half a month in both years. This score resulted in a grade of one before, as well as after the monitoring system was implemented. Participant 2 had a calculated delay of approximately eight months in 2022. According to the grading scale, a delay longer than four months qualifies for a grade of five. In 2023 they followed the payment plan with a delay of almost three months and therefore a grade of three was allocated to them. The average delay of Participant 3 was equal to zero in both settings, which led to a grade of one. Participant 4 was given a grade of one for both years as well. His average delay was 0,5 and one month for 2022 and 2023 respectively. Participant 5 did not have any delay in 2022. On the contrary, they managed to close their projects one month earlier and were therefore assigned a grade of one. After introducing performance monitoring, Participant 5 adhered to the invoice submission deadlines with no delay. This has brought them the grade of one.

Table 1: Grades per participant for 2022 and 2023

	Time tracking	Target calculations	Current calculations	Payment plans	Project management	Collegiality and team cooperation	Customer satisfaction	Improvement of contribution margins
				2022 (2023)				
Participant 1	1 (1)	5 (1)	4 (3)	5 (1)	2 (2)	1 (1)	2 (2)	5 (3)
Participant 2	3 (2)	5 (3)	5 (3)	5 (1)	3 (3)	2 (3)	3 (4)	5 (2)
Participant 3	5 (5)	5 (5)	5 (3)	5 (3)	2 (2)	1 (1)	3 (3)	5 (5)
Participant 4	1 (4)	5 (1)	2 (1)	5 (1)	1 (1)	2 (1)	1 (1)	2 (3)
Participant 5	5 (5)	5 (2)	5 (3)	5 (1)	2 (2)	1 (1)	2 (2)	5 (3)
Average	3,00 (3,40)	5,00 (2,40)	4,20 (2,60)	5,00 (1,40)	2,00 (2,00)	1,40 (1,40)	2,20 (2,40)	4,40 (3,20)

Final grades

The final grades were obtained per participant and per year and are summarized in Table 4. Each participant has gathered a certain score which was translated into grades. The lower the score, the better was the assigned grade.

Participant 1 has achieved 21,30 points over all criteria in 2022. This corresponds to an average grade of three. After monitoring, Participant 1 has managed to reduce their score to 12,10 which translates into a grade of two. Participant 2 had 28,00 points in 2022 and was therefore assigned a grade of four. In 2023, they scored down to 18,10 points, which corresponds to a grade of three. Participant 3 was assigned an average grade of four before the monitoring system was introduced, as the total count of their points was 25,80. After monitoring they had a total of 22,60, which has brought them a grade of three. Participant 4 had the lowest score in 2022, which was 16,00 points. According to the grading scale, 16,00 points correspond to a grade of two. In the next year, Participant 4 was again assigned a grade of two, for the slightly lower score of 11,40. Participant 5 had an average grade of four for the 25,30 points collected in 2022. After introducing the monitoring concept, they managed to go down to 16,10 points which translates into a grade of two.

Data Interpretation

Inter-category comparison

The first method used for the analysis of the final results was an inter-category comparison. In the context of this experiment, inter-category is to be understood as the comparison between the final grades for each category for each year. The basis for the calculations was the transposed table with the final grades of each participant in each category for both years (see Table 2). Based on this data, the minimum and maximum values, first-, second- (median), and third quartile per category have been calculated. Furthermore, the interquartile range, the upper and lower outliers, as well as the upper and lower whiskers have been determined to support the illustration of the data. Not the entire dataset was suitable for a graphical representation, because in some categories the minimum, maximum and the median values were the same.

The comparison between the single categories was based on the average values for each category and each year. The average values were calculated with a simple Excel formula without considering any weighing across the participants, as all the participants are considered to contribute to the single categories to equal extent and importance.

The final comparison between the setting before and after monitoring is displayed in Table 3, where the average values of each category have been compared for both years 2022 and 2023. It can be observed that for the categories data quality of target calculations, current calculations and payment plans, as well as for the improvement of contribution margins and the adherence to invoice submission deadlines, there was a substantial improvement in the average grades. The average grade for the data quality of target calculations in 2022 was 5,00 and in 2023 – 2,60. The quality of current calculations was evaluated at 4,20 for 2022 and 2,60 for 2023. The highest improvement was achieved in the subcategory of data quality – payment plans. The average grade for that measure in 2022 was 5,00 and in 2023 – 1,40. The improvement of contribution margins changed from 4,40 in 2022 to 3,20 in 2023. A slight improvement can be noticed in the category “adherence to invoice submission deadlines”. The average grade for 2022 was 1,80 and for 2023 – 1,40.

Table 2: Statistical data, grades per category for 2022 and 2023

Statistics	Time tracking	Target calculations	Current calculations	Payment plans	Project management	Collegiality and team cooperation	Customer satisfaction	Contribution margins	Adherence to payment plans
2022 (2023)									
Min	1 (1)	5 (1)	2 (1)	5 (1)	1 (1)	1 (1)	1 (1)	2 (2)	1 (1)
Q ₁	1 (2)	5 (1)	4 (3)	5 (1)	2 (2)	1 (1)	2 (2)	5 (3)	1 (1)
Median	3 (4)	5 (2)	5 (3)	5 (1)	2 (2)	1 (1)	2 (2)	5 (3)	1 (1)
Q ₃	5 (5)	5 (3)	5 (3)	5 (1)	2 (2)	2 (1)	3 (3)	5 (3)	1 (1)
Max	5 (5)	5 (5)	5 (3)	5 (3)	3 (3)	2 (3)	3 (4)	5 (5)	5 (3)
IQR	4 (3)	0 (2)	1 (0)	0 (0)	0 (0)	1 (0)	1 (1)	0 (0)	0 (0)
Upper Outliers	0 (0)	0 (0)	0 (0)	0 (1)	1 (1)	0 (1)	0 (0)	0 (1)	1 (1)
Lower Outliers	0 (0)	0 (0)	1 (1)	0 (0)	1 (1)	0 (0)	0 (0)	1 (1)	0 (0)
Variance	4,00 (3,30)	0 (2,80)	1,70 (0,80)	0 (0,80)	0,50 (0,50)	0,30 (0,80)	0,70 (1,30)	1,80 (1,20)	3,20 (0,80)
Average	3,00 (3,40)	5,00 (2,40)	4,20 (2,60)	5,00 (1,40)	2,00 (2,00)	1,40 (1,40)	2,20 (2,40)	4,40 (3,20)	1,80 (1,40)
Data Table									
Participant 1	1 (1)	5 (1)	4 (3)	5 (1)	2 (2)	1 (1)	2 (2)	5 (3)	1 (1)
Participant 2	3 (2)	5 (3)	5 (3)	5 (1)	3 (3)	2 (3)	3 (4)	5 (2)	5 (3)
Participant 3	5 (5)	5 (5)	5 (3)	5 (3)	2 (2)	1 (1)	3 (3)	5 (5)	1 (1)
Participant 4	1 (4)	5 (1)	2 (1)	5 (1)	1 (1)	2 (1)	1 (1)	2 (3)	1 (1)
Participant 5	5 (5)	5 (2)	5 (3)	5 (1)	2 (2)	1 (1)	2 (2)	5 (3)	1 (1)

Table 3: Inter-category comparison

<i>Statistics</i>	<i>2022</i>	<i>2023</i>	<i>Delta absolute</i>	<i>Delta relative</i>
Median	3,00	2,40	-0,60	-20%
Variance	2,07	0,58	-1,50	-72%
Average	3,22	2,24	-0,98	-30%
Weighted average	3,30	2,30	-1,00	-30%
<i>Data Table</i>				
Data quality: time tracking	3,00	3,40	0,40	13%
Data quality: target calculations	5,00	2,40	-2,60	-52%
Data quality: current calculations	4,20	2,60	-1,60	-38%
Data quality: payment plans	5,00	1,40	-3,60	-72%
Collaborative interfaces, project takeover and -management	2,00	2,00	0,00	0%
Collegiality and team cooperation	1,40	1,40	0,00	0%
Customer satisfaction	2,20	2,40	0,20	9%
Improvement of contribution margins	4,40	3,20	-1,20	-27%
Adherence to invoice submission deadlines	1,80	1,40	-0,40	-22%

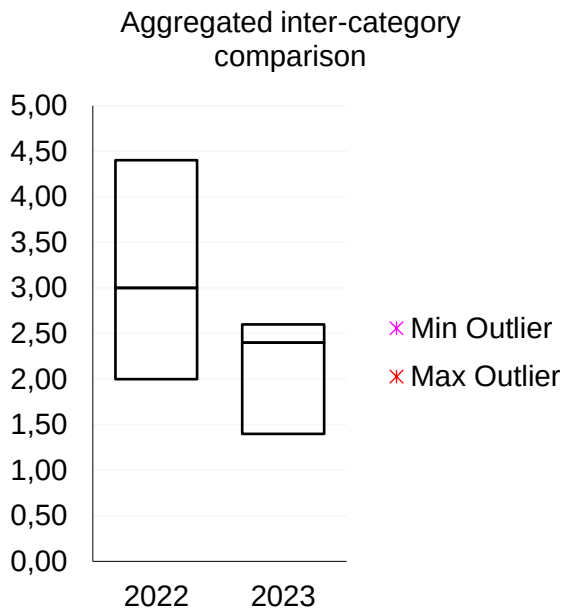
It is worth mentioning that the effort required varies for each category. For instance, to improve the data quality in the system regarding the payment plans involves a comparably low effort at the beginning of the project and if necessary – small adjustments during the projects. The time needed for that task can be estimated at around 30 minutes to an hour per project. On the other hand, improving the target contribution margins requires an ongoing work throughout the project, self-initiative, and several soft skills of the executing party. Therefore, the weighting does not only represent the importance of the criterion, but also ensures a certain balance among the categories considering the effort involved.

There were two criteria that had slightly lower grades after introducing the monitoring system. The data quality of time tracking went from 3,00 in 2022 and 3,40 in 2023. Customer satisfaction was also averaged at 2,20 in 2022 and 2,40 in 2023. For the categories of project management and collegiality there was no change in the grades observed.

The median for the year without monitoring was calculated at 3,00. This value was improved in the next year by 0,60 resulting in a median of 2,40 for 2023 (see Figure 3). However, as mentioned above, the median of the data sets does not have enough significance in terms of effort involved in the single categories. Therefore, a weighted

average value was calculated for both years, considering both the weighting and the grades of each criterion. The weighted average for 2022 was 3,30 and for 2023 – 2,30, which results in a higher improvement than the one of the median values.

Figure 3: Boxplot inter-category comparison



To sum up, considering the predominant number of categories demonstrating improvement from period one to period two, together with the enhancement of the weighted average values, it can be inferred that introducing the monitoring system had a positive effect on employees' overall performance. The absence of further incentives, aside from the monitoring itself, suggests that the positive change in behavior and outcome might be linked to the Hawthorne effect.

Inter-participant comparison

Another method used for the analysis of the data is the comparison of the two settings for each participant. In the context of this experiment, inter-participant is to be understood as the comparison between the final grades for each participant for each year. The detailed findings are shown in Table 4. The same values like in the inter-category comparison were here calculated per participant and represented graphically if possible. In Table 5 the total grades for all categories of the participants are being compared for 2022 and 2023. The extended analysis has been made at this point to refine the results and their differences between 2022 and 2023.

Table 4: Statistical data, grades per participant for 2022 and 2023

Statistics	Participant 1	Participant 2	Participant 3	Participant 4	Participant 5
2022 (2023)					
Min	1 (1)	2 (1)	1 (1)	1 (1)	1 (1)
Q ₁	1 (1)	3 (2)	2 (2)	1 (1)	2 (1)
Median	2 (1)	5 (3)	5 (3)	2 (1)	5 (2)
Q ₃	5 (2)	5 (3)	5 (5)	2 (1)	5 (3)
Max	5 (3)	5 (4)	5 (5)	5 (4)	5 (5)
IQR	4 (1)	2 (1)	3 (3)	1 (0)	3 (2)
Upper Outliers	0 (0)	0 (0)	0 (0)	2 (2)	0 (0)
Lower Outliers	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Variance	3,36 (0,75)	1,50 (0,75)	3,28 (2,61)	2,69 (1,28)	3,53 (1,69)
Average	2,89 (1,67)	4,00 (2,67)	3,56 (3,11)	2,22 (1,56)	3,44 (2,22)
Data Table					
Data quality: time tracking	1 (1)	3 (2)	5 (5)	1 (4)	5 (5)
Data quality: target calculations	5 (1)	5 (3)	5 (5)	5 (1)	5 (2)
Data quality: current calculations	4 (3)	5 (3)	5 (3)	2 (1)	5 (3)
Data quality: payment plans	5 (1)	5 (1)	5 (3)	5 (1)	5 (1)
Project management	2 (2)	3 (3)	2 (2)	1 (1)	2 (2)
Collegiality and team cooperation	1 (1)	2 (3)	1 (1)	2 (1)	1 (1)
Customer satisfaction	2 (2)	3 (4)	3 (3)	1 (1)	2 (2)
Improvement of contribution margins	5 (3)	5 (2)	5 (5)	2 (3)	5 (3)
Adherence to invoice submission deadlines	1 (1)	5 (3)	1 (1)	1 (1)	1 (1)

The first comparison between the participants was based on the final grades of each employee for each year. The final grades were calculated considering the grade for each category and its weighting. The summarized data is shown in Table 5. It can be observed that everyone, except for one participant, has improved their overall performance since the monitoring system was introduced. Participant 1 went from a grade of three for 2022 to a grade of two for 2023. Participant 2 has improved his overall performance by achieving a final grade of three for the setting after monitoring, compared to the grade of four without monitoring. Participant 3 had the exact same success in their final grading as Participant 2. Participant 5 was the only one who managed to enhance their grade from a four to a two. However, some project managers have more projects per year than others. This brings a higher possibility to score lower. For example, a person who manages only five projects per year is expected to have more resources for entering their data on time or adhering to invoice submission deadlines than a person who manages 50 projects per year. This limitation will be discussed in a further chapter.

Table 5: Inter-participant comparison (grades)

<i>Statistics</i>	<i>2022</i>	<i>2023</i>	<i>Delta absolute</i>	<i>Delta relative</i>
Median	4,00	2,00	-2,00	-50%
Variance	0,80	0,30	-0,50	-63%
Average	3,40	2,40	-1,00	-29%
<i>Data Table</i>				
Participant 1	3,00	2,00	-1,00	-33%
Participant 2	4,00	3,00	-1,00	-25%
Participant 3	4,00	3,00	-1,00	-25%
Participant 4	2,00	2,00	0,00	0%
Participant 5	4,00	2,00	-2,00	-50%

The performance of Participant 4 remained the same over the two periods. It is worth mentioning, that even though their performance has not changed in 2023, it does not mean that their performance was unsatisfactory. On the contrary, a fundamentally good performance can be more challenging to improve than a poor performance. Participant 4 had the best score in both settings.

The second comparison was based on the total points of each participant, ensuring a more detailed analysis than the final grades. Table 6 summarizes those findings. It becomes evident that Participants 1, 2 and 5 had almost the same change in the final score, ranging between 9,20 and 9,90. Therefore, it can be concluded that Participant 5 did not have a more positive development than the other two employees. Their score just happened to be

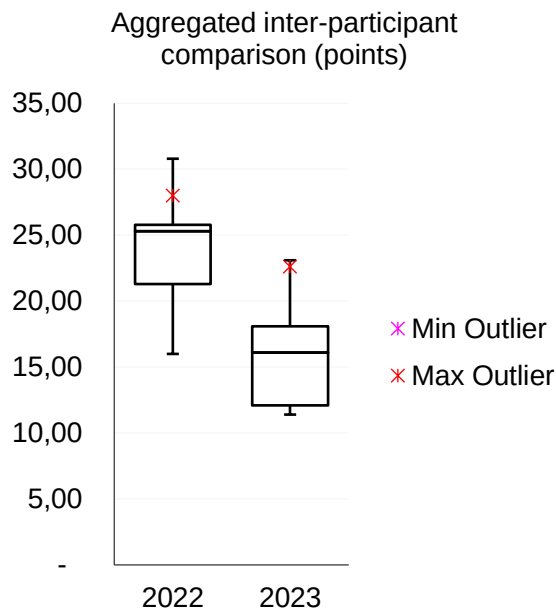
distributed in a way that their final grades seem to be further apart. Similarly, the progress of Participant 3 was calculated at merely 3,20 points and yet they successfully improved their final grade from four to three.

Table 6: Inter-participant comparison (points)

<i>Statistics</i>	<i>2022</i>	<i>2023</i>	<i>Delta absolute</i>	<i>Delta relative</i>
Median	25,30	16,10	-9,20	-36%
Variance	22,41	21,08	-1,32	-6%
Average	23,28	16,06	-7,22	-31%
<i>Data Table</i>				
Participant 1	21,30	12,10	-9,20	-43%
Participant 2	28,00	18,10	-9,90	-35%
Participant 3	25,80	22,60	-3,20	-12%
Participant 4	16,00	11,40	-4,60	-29%
Participant 5	25,30	16,10	-9,20	-36%

The median of the grades was calculated at four for 2022 and two for 2023. The median of the total score for each participant was 25,30 for 2022 and 16,10 for 2023. The second method has provided a better graphical representation of the boxplots for the inter-participant comparison than the one with the final grades (see Figure 4). To summarize, the inter-participant comparison provides further evidence for the enhancement of the overall performance after implementing the monitoring system. The results support the hypothesis that a positive change in performance might be caused by implementing a monitoring system without a monetary incentive plan.

Figure 4: Boxplot inter-participant comparison



Statistical Analysis

Besides the descriptive statistics introduced above, a Shapiro-Wilk test has been used to test the normality of the data. The representation of the findings has been done with a Q-Q (quantile-quantile) plot and histograms. The Shapiro-Wilk test is used to determine whether a data set is normally distributed. This is a prerequisite for small data samples before a t-Test can be conducted. The t-Test proves the significance of the findings.

Four data sets have been tested for normality: the average scores for each category in 2022 and 2023 and the final grades for each participant in 2022 and 2023. The Shapiro-Wilk test has shown that only one data sample is normally distributed. Therefore, it was not possible to conduct a t-Test for significance (see Appendix F and Appendix G). An alternative for small data samples which are not normally distributed is the Wilcoxon signed-rank test. It was conducted on the inter-category comparison, as the number of samples (five) of the inter-participant comparison was not satisfying one of the conditions for the test.

The data from the inter-category comparison is paired, meaning that the two observations from 2022 and 2023 are taken on the same participants under different conditions. The difference in the conditions is the introduction of the monitoring system in 2023. As the hypotheses of the first research question are being tested, a two-tailed Wilcoxon signed-rank test has been utilized. The two-tailed test is conducted when the directionality of the results is unknown. In this experiment, we want to test whether there is a significant

difference between the findings, regardless of whether the change is expected to be positive or negative. The direction will be then determined by the descriptive statistics of this thesis.

The test was carried out in Excel and its results are summarized in Table 7. After computing the direction of the delta of the values in 2022 and 2023 and assigning them with a ranking from one to seven (two of the categories have not changed over time), the sum of the positive and the negative ranking has been calculated. The lower sum was then considered a test statistic value which was compared to the critical value (p-value). The test statistic is equal to three and p-value from the corresponding Wilcoxon signed-rank table is equal to five. To conclude, the test statistic is smaller than the critical value, therefore there is sufficient evidence for the significance of the difference and the Null Hypothesis can be rejected.

Table 7: Wilcoxon signed-rank significance test

	2022	2023	Dir	D(A-B)	Rd	R+	R-	p
Data quality: time tracking	3	3,4	-	0,4	2		2	
Data quality: target calculations	5	2,4	+	2,6	6	6		
Data quality: current calculations	4,2	2,6	+	1,6	5	5		
Data quality: payment plans	5	1,4	+	3,6	7	7		
Collaborative interfaces, project takeover and -management	2	2						
Collegiality and team cooperation	1,4	1,4						
Customer satisfaction	2,2	2,4	-	0,2	1		1	
Improvement of contribution margins	4,4	3,2	+	1,2	4	4		
Adherence to invoice submission deadlines	1,8	1,4	+	0,4	3	3		
n	9	9			7			
Sum	29	20,2			28	25	3	5

Discussion

Research questions

Considering the positive change in overall performance and the significance of the results in the inter-category comparison, the first research question of this thesis:

“Is it possible to improve employees’ overall performance by implementing a multi-variable monitoring system compared to a setting without monitoring?”,

can be answered by rejecting the Null Hypothesis (H0) and the Alternative Hypothesis (H2).

Null Hypothesis (H0): Multi-variable monitoring has no impact on employees’ overall performance.

Alternative Hypothesis 1 (H1): Multi-variable monitoring has a negative impact on employees’ overall performance.

Alternative Hypothesis 2 (H2): Monitoring has a negative impact on employees’ performance.

In order to answer the second research question, which reads as follows:

“Does the Hawthorne effect in employees’ monitoring have a bigger impact on qualitative criteria like customer satisfaction and collegiality or on quantitative criteria such as the improvement of contribution margins or the adherence to invoice submission deadlines?”,

the inter-category comparison can be used for a more detailed analysis of the separate categories. Table 3 demonstrates that three out of seven qualitative measures have been improved after the monitoring model was implemented. This corresponds to 43% of the qualitative criteria. Those are: the data quality of the target calculations with a delta of 2,60, the data quality of the current calculations with a delta of 1,60, and the data quality of the payment plans with a delta of 3,60. The average improvement over the whole set of qualitative categories, including those without an enhancement in performance and those showing negative effects, results in exactly 1,00 points.

On the contrary, there were only two quantitative measures involved in the whole model, both of which were improved over time. The average grade of improvement of contribution margins was calculated by 1,20 points better in the year after the monitoring system was introduced. The adherence to invoice submission deadlines has enhanced by 0,40 points

in 2023. The average improvement over the whole set of quantitative categories is 0,80 points.

Although the results suggest that the average improvement of the qualitative criteria was higher than the one of the quantitative categories, there are some limitations of the model that need to be considered. Firstly, as mentioned above, it is easier to improve a lower grade to a better one than to improve an already good grade. For instance, the data quality of the target calculations and the payment plans was equal to 5,00 in the first year, simply because these requirements did not exist before monitoring. Therefore, after announcing that employees' performance is being monitored and the data quality is an important part of the model, the project managers increased their attention to those aspects and scored significantly better in the next year. On the contrary, the adherence to invoice submission deadlines was already excellent before introducing the monitoring system. There was still a small improvement in the next year, but from a mathematical or statistical perspective, it was not as good as the change in the data quality category.

Another argument against drawing a final conclusion based on the simple improvement of the average grades are the differences in the effort needed for improvement across the categories. For example, the improvement of contribution margins requires an ongoing work throughout the whole project duration. On the contrary, the data quality of payment plans can be enhanced by a simple decision to put the required data into the system. The effort associated with that category is estimated to range between 30 minutes and an hour per project. Thus, the improvement of performance in these two categories can only then be directly compared when the weightings of the single categories represent solely the effort involved and not the importance of the criteria according to the CEO. This was tested with a robustness check.

A further argument against the direct comparison is the fact that not all the participants are managing an equivalent workload in terms of project count. The average number of projects per year per participant varies significantly. Participant 1 and 2 usually manage around five projects per year, Participant 3 – between 25 and 30, Participant 4 – between 10 and 15, and Participant 5 is usually responsible for up to two projects per year. Furthermore, the size and the complexity of the projects show pronounced differences. For these reasons, the participants cannot be considered with the same importance and therefore the improvement of the average grades has only a limited significance.

To conclude, the second research question could not be answered based on the findings of this experiment due to the complex nature of the projects and the employment relations.

In order to achieve exact results from this multi-variable model, the experiment has to include participants with very similar background as well as comparable projects. Further suggestions for improvement are discussed in chapter “Future Research”.

Integration with Literature

According to Holderness et al. (2022), employees that expect to be monitored are more likely to perform better. The results from this experiment have found further evidence that supports this study. The positive change in behavior due to monitoring was studied by numerous authors, mentioned in the literature review, most of which could be reinforced by the inter-category and inter-participant analysis. Regarding the positive aspect of cost saving, there was no evidence gathered except for the improvement of contribution margins. However, the real effect on cost optimization should be observed on the long run, which was not possible in this experiment due to time limitation.

The findings from the experiment also support the theory about efficiency and optimized resource allocation by Ukko et al. (2007). They suggest that monitoring can be used to improve working processes, which was found to be evident in the experiment when comparing the grades on data quality. Before introducing the monitoring system, there were no guidelines on how often or how exact specific data needs to be entered in the system. This resulted in a disorganized, inefficient workload. After monitoring, this category has improved substantially, leaving room and resources for further tasks, as the data was gathered and shared transparently with all parties involved.

Based on the findings from the literature review, suggesting on multiple occasions that too invasive techniques might have negative consequences, no excessive computer monitoring has been used during the experiment. The sense of autonomy in the participants has been a priority in the company and there were special efforts made to ensure they feel independent and that only work-related tasks are being monitored. In that way, the risk of encountering of negative consequences was mitigated, which can be observed in the positive results of both inter-category and inter-participant comparisons.

A study by Balcazar et al. (1985) was used in the definition phase of the model. The authors found that the combination of feedback and goal setting has proved to be the most effective on employee motivation. Therefore, a consistent feedback and communication between the participants, the CEO, the controlling department and the head of energy department was ensured during the experiment. The goal setting in the form of the grades that have been defined, has contributed to higher overall performance by promoting a sense of

accountability. The results of the experiment align with a Japanese study on monitoring approaches (Itoh et al., 2008).

Both categories – improvement of contribution margins and adherence to invoice submission deadlines, were associated with goal and deadline setting and have shown an improvement over time. This supports the findings from Waller et al. (2002) regarding the positive influence of monitoring. In the meantime, there were several studies claiming that setting goals and deadlines might have a negative impact on employees' performance, for which no evidence was found in this experiment. One possible explanation for the absence of negative effects on performance and motivation could be that the deadlines were not defined very rigidly. The company in the experiment operates in a high-risk sector, where safety-aspects need to be satisfied with higher priority than completing the project on time. However, it may be worth considering whether the goals related to the improvement of contribution margins were not set too leniently.

The study of Porter et al. (2010) demonstrates that monitoring performance might have a positive effect on team cooperation. The results from the inter-category comparison could not find evidence for any improvement in the second year. One of the reasons for that might be the fact, that team cooperation was already one of the best scoring criteria in the experiment. Therefore, improving an already good score might be harder than improving a poor performance.

There was a slightly lower average grade in the category of customer satisfaction when compared to the setting without monitoring. Frank and Enkawa (2009) suggest that if managers prioritize financial outcomes over quality of work, customer satisfaction might undergo negative changes. The change between the results for 2022 and 2023 in this experiment has been calculated at only -0,20. Thus, the findings about the negative impact of monitoring on customer satisfaction could not be supported or rejected.

The most notable improvement was found in the category of data quality. The experiment by Miyazaki (2013) has demonstrated that people tend to work more slowly but accurately when they know that they are being watched. The findings of the experiment can partly support this theory, because data quality was the category where an average improvement of 1,85 was detected.

Regarding the improvement of contribution margins there are numerous opposing theories. On the one hand, the original Hawthorne effect has provided a significant prove that observation can enhance efficiency in terms of quantities produced. On the contrary, Fisher (1990) has found other approaches like tighter financial control and product rationalization

to be more effective on the quantitative side. In this experiment, the data shows that an improvement of contribution margins is possible due to monitoring. However, it is worth mentioning that the contribution margins of the projects included in this thesis vary strongly and are therefore only occasionally comparable.

Lied and Kazandjian (1998) suggest that the Hawthorne effect usually fades away in a short period of time. This theory could not be tested in the experiment due to the limited time available. A further exploring of the duration of the positive change in performance might provide valuable insights for further review.

Another part of the literature review that could not be tested is the difference in monitoring of qualitative and quantitative measures. Aiello (1993) puts forward the idea that quality of work, measured by customer satisfaction, might be poorly affected by performance monitoring. Since the second research question regarding the differences between qualitative and quantitative means could not be answered, this part of the thesis remains unresolved. In order to accomplish significant results from this multi-variable model, the experiment has to include participants with very similar background as well as comparable projects.

Practical and Theoretical Implications

Beside the three limitations addressed when answering the second research question, there were several practical limitations that have occurred when designing the experiment, during its execution, and during the data analysis. The initial phase of the experiment, where the employees were informed about the new model and its requirements, has shown that the necessary training time was underestimated by approximately one month. The more employees and the more categories are included in the monitoring system, the more resources are needed for the initial training. The delay of one month during the training phase has contributed to a delay of the final results, as it was not possible to proceed with further steps of the plan.

Another limitation of the experiment is the influence of external factors on project turnover. External factors might have influenced the old as well as the new projects and therefore might have delivered inaccurate results. For instance, if the contribution margins in the past were very good and, in the present - not anymore, it doesn't necessary mean that the performance has gotten worse. Firstly, a bigger data set would help reduce those uncertainties. Secondly, a numerical correction would be necessary after a conversation with the employee if a deviation has been detected. The deviation refers to the delta

between the original target calculation and the corrected target calculation and not between target and current ones.

A further limitation of the projects in this experiment was that the company often submits bids with very low or even negative contribution margins in order to retain its position on the market. Moreover, the projects included can rarely be compared to each other due to multiple factors such as: size, duration, customer, complexity, and wage-material ratio. The projects of different customers that the company is working with often deliver different contribution margins. For instance, customer A usually employs projects with a margin between 20% and 30%, while customer B is known to demand lower prices or discounts, which leads to a lower contribution margin around 10%.

The wage-material ratio describes the amount of labor costs and material costs in a single project. This metric varies strongly in the company as well and has an influence on the effort involved and eventually on the profitability of the project. For example, a project that has 90% labor costs and only 10% material costs would be highly time consuming due to the coordination of personnel on the construction site, technical requirements that need to be met, and documentation. On the contrary, a project with 90% material costs and 10% labor costs represents a low-effort project, because it does not involve high effort besides the arrangements for the delivery and pickup from the back-office personnel. These two types of projects cannot be compared fairly due to the differences in the effort and responsibility involved. A possible solution for that problem was proposed in the chapter "Future research".

Furthermore, the grade assignment often did not offer a great variability in the results as it represents only an aggregated form of the achieved performance. There might have been some changes in the behavior in 2022 and 2023 but not enough to assign a higher or lower grade. A partial solution was looking into the points and parameters of a category when possible. This has been done in the category of project management.

During the data analysis there was another limitation with the representation of the results. Not every category or participant was suitable for graphical representation with a boxplot chart, because the minimum and the maximum values, as well as the interquartile ranges were often overlapping. Therefore, only the aggregated version of the findings was graphically represented in the inter-category and the inter-participant comparison.

The last drawback of the model that was detected during the elaboration of the experiment refers to the accuracy of the collected data. The evaluation of the contribution margins was partially based on trusting the employees that they have entered all of their working time

in the project. If they have not, the profitability of the project would appear better than it was in reality.

Robustness Check

The robustness check aims to test the sensitivity of the findings from the experiment by adjusting the weightings used in the model. It also offers a basis for further research on the comparability between the qualitative and quantitative criteria. The weightings in the experiment were defined by the CEO according to the perceived importance the single categories had for the company. For instance, the category of data quality was considered with medium importance as it was providing the basis for further analysis of information. During the robustness check, the weightings have been adapted to the effort needed by the participants to fulfill the requirements of the single criteria.

All four categories of data quality: time tracking, target calculations, current calculations and payment plans, have been assigned a low weighting because the estimated time and therefore, effort involved are considerably low compared to the other categories. The category “Collaborative interfaces, project takeover and -management” was weighted again with the highest priority, as it was during the original experiment. This category is not only crucial for company performance but requires an ongoing effort from the beginning of the project until the end and therefore is considered to be a high-effort part of the model. While collegiality and team cooperation refer to the soft skills of the participants in the company, customer satisfaction refers to proper use of interpersonal skills outside of the company. Both of those categories have been assigned a medium weighting of 0,8 because they are associated with an ongoing work throughout the whole process. However, while the category of project management addresses nine subcategories of high importance and effort, the categories of teamwork and customer satisfaction consider a smaller range of requirements in the evaluation.

The improvement of contribution margins is weighted with one. This is the highest score, because the improvement of contribution margins requires an excellent project management throughout the project duration as well as financial knowledge that needs to be applied. The last criterion is the adherence to invoice submission deadlines. It was assigned the medium weighting, as it entails for a good time management and a feeling for a prioritization.

As the weighting does not have any influence on the grades within the category itself, the only change that can be observed during the robustness check is the alteration in the findings of the inter-participant comparison. Table 8 shows the resulted delta between the

two periods. It can be observed that the positive change in the grades is still evident after the robustness check.

Table 8: Inter-participant comparison (robustness check)

<i>Statistics</i>	<i>2022</i>	<i>2023</i>	<i>Delta absolute</i>	<i>Delta relative</i>
Median	20,20	3,70	-6,50	-32%
Variance	20,58	15,41	-5,16	-25%
Average	19,32	14,26	-5,06	-26%
<i>Data Table</i>				
Participant 1	17,70	11,20	-6,50	-37%
Participant 2	25,00	17,50	-7,50	-30%
Participant 3	21,00	19,00	-2,00	-10%
Participant 4	12,70	9,90	-2,80	-22%
Participant 5	20,20	13,70	-6,50	-32%

When adapting the weighting, the linear grade distribution changes as follows:

- 1: between 6,4 and 9,5
- 2: between 9,6 and 15,9
- 3: between 16,0 and 22,3
- 4: between 22,4 and 28,7
- 5: between 28,8 and 32.

Although the categories have not been individually impacted by the change in their weightings, the weighted average across the whole data set has changed after the adaptation. The weighted average of 2022 was calculated at 3,02, compared to the 3,30 from the original experiment. The weighted average of 2023 from the robustness check was 2,2 and from the original experiment - 2,3. To conclude, although the relative delta of overall performance dropped from 30% to 26%, a considerable improvement could be observed during the robustness check as well, which confirms the stability of the model and the delivered results. The significance of the robustness check findings could not be tested with any statistical analyses due to the small sample size of the inter-participant comparison.

Future Research

In order to answer the second research question and to compensate the limitations mentioned before, several adaptations of the model would be needed for further research. Firstly, a more frequent measurement of the criteria might bring more accurate insights on

employees' performance and its development over time. Secondly, the analysis of the improvement of contribution margins should be conducted over a longer period of time to gain real insights and to notice the real effect of monitoring. However, the literature suggests that the Hawthorne effect is most likely to detect in the early stages of a new implementation, but in that case, it was not possible to test whether the positive impact of the observation bias "wears off".

To eliminate the uncertainty of trusting employees that they have entered their working time completely, a further controlling mechanism might be implemented. A mechanism, that distributes the differences between the time clock and the time entries in the ERP system among the projects. A potential allocation key for that calculation would be according to the sales volume of the projects. For instance, if an employee has worked 200 hours according to the time clock and has entered only 150 hours in the ERP system, the difference of 50 hours should be allocated proportionally based on the sales volume of the active projects at the time.

A further suggestion regarding the base of comparison of different projects would be another type of weighting or categorization among the projects. For example, a project with a duration longer than a year, wage-material ratio around 50/50 and generally considered more complex than others, should be weighed higher than smaller, more simple projects. Another alteration that might be needed for future research is repeating the experiment with an equal number of qualitative and quantitative categories. In that way, the second research question regarding the strength of the Hawthorne effect on both types could be answered.

Lastly, in the experiment of this thesis, the improvement of contribution margins was assigned with a positive grade only when the calculated delta between target and current calculations was positive. A variation of this rule would be a more lenient approach where not only the improvement of contribution margins is being graded, but also the improvement of previous performance. For instance, Participant 3 had a delta of -15% before monitoring and -7% after monitoring. According to the grading scale, they would receive a grade of five in both periods. But if the management of the company decides to encourage an improvement of previous performance even if it is still negative, Participant 3 would receive a positive grade in 2023. This could contribute to a feeling of appreciation among the employees and eventually improve the employee retention rates in the company.

Conclusion

The objective of this thesis was to identify the connection between employees' monitoring and overall performance. Considering the literature on the impact of performance monitoring on employees' behavior and the Hawthorne effect, it can be concluded that organizations can achieve an enhancement in effectiveness and efficiency by implementing a monitoring system. This was supported by the findings of the empirical experiment of this thesis. The implementation of a multi-variable performance measuring model has proved to influence the overall quality of work of the five participants in the experiment. Therefore, a positive impact on employees' overall performance was demonstrated and proven to be statistically significant.

The results from the inter-category- and the inter-participant comparison provide an amplification of existing literature on the topic with a more refined perspective. While previous researches focus mostly on managerial performance, the experiment in this thesis has provided an alternative review of performance located lower in the hierarchy. Furthermore, the selected combination of the criteria used in the multi-variable model contributes to the literature of the modern Hawthorne effect in the economical context.

Although the second research question could not be answered within the scope of this thesis, several suggestions and alterations for further analysis have been proposed in the chapter "Future Research". Taking into consideration the few aspects mentioned in the previous chapter, a similar model can be used by companies as a universal performance measuring instrument by adjusting only a few parameters. For instance, the adjustment of the weightings, mentioned in the chapter "Robustness check", offers a certain flexibility according to the needs and requirements of different companies and situations.

Literature

- Aiello, J. R. (1993). *Computer-Based Work Monitoring: Electronic Surveillance and Its Effects*. 499–507.
- Alder, G. S., & Tompkins, P. K. (1997). Electronic Performance Monitoring: An Organizational Justice and Concertive Control Perspective. *Management Communication Quarterly*, 10(3), 259–288. <https://doi.org/10.1177/0893318997010003001>
- Alge, B. J. (2001). *Effects of computer surveillance on perceptions of privacy and procedural justice*. 797–804.
- Anderson, D. M., & Stritch, J. M. (2016). Goal Clarity, Task Significance, and Performance: Evidence From a Laboratory Experiment. *Journal of Public Administration Research and Theory*, 26(2), 211–225. <https://doi.org/10.1093/jopart/muv019>
- Asch, S. E. (1955). Opinions and social pressure. *Scientific American*, 31–35.
- Balcazar, F., Hopkins, B. L., & Suarez, Y. (1985). A Critical, Objective Review of Performance Feedback. *Journal of Organizational Behavior Management*, 7(3–4), 65–89. https://doi.org/10.1300/J075v07n03_05
- Bartik, T. J. (1995). *Using Performance Indicators to Improve the Effectiveness of Welfare-to-Work Programs*. W.E. Upjohn Institute. <https://doi.org/10.17848/wp95-36>
- Berman, E. M. (1998). *Productivity in public and nonprofit organizations: Strategies and techniques*. Sage Publications.
- Bessane, S. F., Camara, M. S., Fall, I., & Bah, A. (2018). Causal model of performance measurement systems by combining qualitative and quantitative models for robust results. *2018 International Conference on Intelligent Systems and Computer Vision (ISCV)*, 1–7. <https://doi.org/10.1109/ISACV.2018.8354076>
- Bylinski, G. (1991). How companies spy on employees. *Fortune*, 131–140.
- Chen, L. F., Vander Weg, M. W., Hofmann, D. A., & Reisinger, H. S. (2015). The Hawthorne Effect in Infection Prevention and Epidemiology. *Infection Control & Hospital Epidemiology*, 36(12), 1444–1450. <https://doi.org/10.1017/ice.2015.216>
- Christensen, C. M., Horn, M. B., & Johnson, C. W. (2008). *Disrupting class: How disruptive innovation will change the way the world learns*.
- Dishneau, D. (1922). *Sears Admits Mistakes, Takes Workers Out of Commission*. The Associated Press.
- Fenner, D. B., Lerch, F. J., & Kulik, C. T. (1993). The Impact of Computerized Performance Monitoring and Prior Performance Knowledge on Performance Evaluation ¹. *Journal of Applied Social Psychology*, 23(7), 573–601. <https://doi.org/10.1111/j.1559-1816.1993.tb01104.x>
- Fernald, D. H., Coombs, L., DeAlleaume, L., West, D., & Parnes, B. (2012). An Assessment of the Hawthorne Effect in Practice-based Research. *The Journal of the American Board of Family Medicine*, 25(1), 83–86. <https://doi.org/10.3122/jabfm.2012.01.110019>

Fickel, L. (1991). Don't look now, but...: A new crop of network products is forcing a debate over workplace monitoring. *Infoworld*, 50–55.

Fisher, T. J. (1990). Quality Management and Productivity—A Preliminary Study. *Australian Journal of Management*, 15(1), 107–127. <https://doi.org/10.1177/031289629001500105>

Gale, E. A. M. (2004). The Hawthorne studies—A fable for our times? *QJM*, 97(7), 439–449. <https://doi.org/10.1093/qjmed/hch070>

Gathani, S., Gomez, M. P., Sabates, R., & Stoelinga, D. (2015). The Effect of Monitoring: How Data Collection Type and Frequency Boosts Participation and the Adoption of Best Practices in a Coffee Agronomy Training Program in Rwanda. *Evaluation Review*, 39(6), 555–586. <https://doi.org/10.1177/0193841X16633584>

Grant, R., Higgins, C. A., & Irving, R. H. (1988). Computerized Performance Monitors, Are They Costing You Customers? *Sloan Management Review*, 39–45.

Gupta, N., & Sharma, V. (2016). *Exploring Employee Engagement—A Way to Better Business Performance*. *Global Business Review*, 45–63.

Guthrie, J. P. (2000). Alternative Pay Practices and Employee Turnover: An Organization Economics Perspective. *Group & Organization Management*, 25(4), 419–439. <https://doi.org/10.1177/1059601100254006>

Hatry, H. P., Alexander, M., & Financial Accounting Foundation (Hrsg.). (1990). *Service efforts and accomplishments reporting: Its time has come: an overview*.

Hershey, P., & Blanchard, K. H. (1969). Life cycle theory of leadership: Is there a best style of leadership? *Training and Development Journal*, 26–34.

Heyman, G. D., & Dweck, C. S. (1992). Achievement goals and intrinsic motivation: Their relation and their role in adaptive motivation. *Motivation and Emotion*, 16(3), 231–247. <https://doi.org/10.1007/BF00991653>

Holland, P. J., Cooper, B., & Hecker, R. (2015). Electronic monitoring and surveillance in the workplace: The effects on trust in management, and the moderating role of occupational type. *Personnel Review*, 44(1), 161–175. <https://doi.org/10.1108/PR-11-2013-0211>

Ilgen, D. R., Fisher, C. D., & Taylor, M. S. (1979). Consequences of individual feedback on behavior in organizations. *Journal of Applied Psychology*, 64(4), 349–371. <https://doi.org/10.1037/0021-9010.64.4.349>

Itoh, H., Kikutani, T., & Hayashida, O. (2008). Complementarities among authority, accountability, and monitoring: Evidence from Japanese business groups. *Journal of the Japanese and International Economies*, 22(2), 207–228. <https://doi.org/10.1016/j.jjie.2008.03.003>

Jensen, J. M., & Raver, J. L. (2012). *When self-management and surveillance collide: Consequences for employees' trust, autonomy, and discretionary behaviors*. 37, 308–346.

Jones, S. R. G. (1990). Worker Interdependence and Output: The Hawthorne Studies Reevaluated. *American Sociological Review*, 55(2), 176. <https://doi.org/10.2307/2095625>

- Liao, E. Y., & Chun, H. (2016). Supervisor monitoring and subordinate innovation. *Journal of Organizational Behavior*, 37(2), 168–192. <https://doi.org/10.1002/job.2035>
- Lied, T. R., & Kazandjian, V. A. (1998). A Hawthorne strategy: Implications for performance measurement and improvement. *Clinical performance and quality health care*, 6, 201–204.
- London, M., & Smither, J. W. (1995). CAN MULTI-SOURCE FEEDBACK CHANGE PERCEPTIONS OF GOAL ACCOMPLISHMENT, SELF-EVALUATIONS, AND PERFORMANCE-RELATED OUTCOMES? THEORY-BASED APPLICATIONS AND DIRECTIONS FOR RESEARCH. *Personnel Psychology*, 48(4), 803–839. <https://doi.org/10.1111/j.1744-6570.1995.tb01782.x>
- Marks, M. A., & Panzer, F. J. (2004). The Influence of Team Monitoring on Team Processes and Performance. *Human Performance*, 17(1), 25–41. https://doi.org/10.1207/S15327043HUP1701_2
- Martin, J., Cormican, K., Sampaio, S. C. B., & Wu, Q. (2018). Shared leadership and team performance: An analysis of moderating factors. *Procedia Computer Science*, 138, 671–679. <https://doi.org/10.1016/j.procs.2018.10.089>
- McCarney, R., Warner, J., Iliffe, S., Van Haselen, R., Griffin, M., & Fisher, P. (2007). The Hawthorne Effect: A randomised, controlled trial. *BMC Medical Research Methodology*, 7(1), 30. <https://doi.org/10.1186/1471-2288-7-30>
- Meredith, J. R. (1992). The management of operations: A conceptual emphasis. New York: John Wiley & Sons.
- Miyazaki, Y. (2013). Increasing Visual Search Accuracy by Being Watched. *PLoS ONE*, 8(1), e53500. <https://doi.org/10.1371/journal.pone.0053500>
- Mulgan, G. J. (1991). *Communication and Control: Networks and the New Economies of Communication*, New York: Guilford.
- Okhuysen, G. A., & Waller, M. J. (2002). Focusing on midpoint transitions: An analysis of boundary conditions. *Academy of Management Journal*, 45, 1056–1065.
- Ordonez, L. D., Schweitzer, M. E., Galinsky, A. D., & Bazerman, M. H. (2009). Goals Gone Wild: The Systematic Side Effects of Over-Prescribing Goal Setting. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.1332071>
- Parkan, C., & Wu, M.-L. (1999). Measurement of the performance of an investment bank using the operational competitiveness rating procedure. *Omega*, 27(2), 201–217. [https://doi.org/10.1016/S0305-0483\(98\)00041-3](https://doi.org/10.1016/S0305-0483(98)00041-3)
- Porter, C. O. L. H., Itir Gogus, C., & Yu, R. C.-F. (2010). When Does Teamwork Translate Into Improved Team Performance? A Resource Allocation Perspective. *Small Group Research*, 41(2), 221–248. <https://doi.org/10.1177/1046496409356319>
- Ross, S. (1992). Big brother in workplace growing bigger every day. *Reuter Business Report*, 11–12.
- Saez-Marti, M., & Sjögren, A. (2007). Deadlines and Distractions. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.1076184>
- Sahay, B. S. (2005). Multi-factor productivity measurement model for service organization. *International Journal of Productivity and Performance Management*, 7–22.

Smet, A. D., Mugayar-Baldocchi, M., Reich, A., & Schaninger, B. (2023). Some employees are destroying value. Others are building it. Do you know the difference? *McKinsey & Company*.

Smither, J. W., London, M., & Reilly, R. R. (2005). DOES PERFORMANCE IMPROVE FOLLOWING MULTISOURCE FEEDBACK? A THEORETICAL MODEL, META-ANALYSIS, AND REVIEW OF EMPIRICAL FINDINGS. *Personnel Psychology*, 58(1), 33–66. https://doi.org/10.1111/j.1744-6570.2005.514_1.x

Soman, D., & Cheema, A. (2004). When Goals Are Counterproductive: The Effects of Violation of a Behavioral Goal on Subsequent Performance. *Journal of Consumer Research*, 31(1), 52–62. <https://doi.org/10.1086/383423>

Tucker, T. (1992). Keeping tabs: Call accounting saves money and more, a lot more. *Keeping tabs: Call accounting saves money and more, a lot more*.

Ukko, J., Tenhunen, J., & Rantanen, H. (2007). Performance measurement impacts on management and leadership: Perspectives of management and employees. *International Journal of Production Economics*, 110(1–2), 39–51. <https://doi.org/10.1016/j.ijpe.2007.02.008>

Waller, M. J., Zellmer-Bruhm, M. E., & Giambatista, R. C. (2002). Watching The Clock: Group Pacing Behavior Under Dynamic Deadlines. *Academy of Management Journal*, 45, 1046–1055.

Wells, D. L., Moorman, R. H., & Werner, J. M. (2007). The impact of the perceived purpose of electronic performance monitoring on an array of attitudinal variables. *Human Resource Development Quarterly*, 18(1), 121–138. <https://doi.org/10.1002/hrdq.1194>

Wholey, J. S., & Hatry, H. P. (1992). The Case for Performance Monitoring. *Public Administration Review*, 52(6), 604. <https://doi.org/10.2307/977173>

Yost, A. B., Behrend, T. S., Howardson, G., Badger Darrow, J., & Jensen, J. M. (2019). Reactance to Electronic Surveillance: A Test of Antecedents and Outcomes. *Journal of Business and Psychology*, 34(1), 71–86. <https://doi.org/10.1007/s10869-018-9532-2>

Zeydan, M., & Çolpan, C. (2009). A new decision support system for performance measurement using combined fuzzy TOPSIS/DEA approach. *International Journal of Production Research*, 47(15), 4327–4349. <https://doi.org/10.1080/00207540802662870>

Zhang, J. (2016). Deadlines in Product Development. *Management Science*, 62(11), 3310–3326. <https://doi.org/10.1287/mnsc.2015.2300>

Appendix A – Grade Definition

<i>Criterion</i>	5	4	3	2	1
Data quality: Time tracking	Less than 50% of the time clock hours are entered, or later than six months regardless of the quantity	50-70% of the time clock hours are entered in the first six subsequent months	70-100% of the time clock hours are entered in the first six subsequent months	70-100% of the time clock hours are entered in the first three subsequent months	70-100% of the time clock hours are entered in the subsequent month
Data quality: Target calculations	Target calculation not entered	Target calculation entered incorrectly	Target calculation entered at a later point	Target calculation entered right after project creation, but not updated if necessary	Target calculation entered right after project creation and updated if necessary
Data quality: Current calculations	Various duplicate material entries, work orders are authorized with delay (more than one month)	Various duplicate material entries, work orders are authorized with delay (up to one month)	Few duplicate material entries, work orders are authorized with delay (up to one month)	Few duplicate material entries, work orders are authorized on schedule	No duplicate material entries, work orders are authorized on schedule
Data quality: Payment plans	Payment plans not entered	Payment plans entered incorrectly	Payment plans entered at a later point in time	Payment plans entered but not updated when changes occurred	Payment plans entered and updated when changes occurred
Collaborative interfaces, project takeover and -management	0-20 points from the form	21-25 points from the form	26-35 points from the form	36-45 points from the form	46-51 points from the form
Collegiality and team cooperation	The person is hardly ever available by phone and neglects to answer emails. Face-to-face availability is virtually non-existent. Willingness to help and support with internal challenges is not evident and they make little or no contribution to achieving departmental and company goals.	The person is occasionally available by telephone and responds to e-mails with delays. Face-to-face availability is irregular. Helpfulness and support with internal challenges are rare and involvement in achieving departmental goals is limited.	The person is mostly reachable by phone and responds to emails in acceptable time. Face-to-face availability is given, although not always at optimal intervals. They are occasionally helpful and provide support with internal problems. Their involvement in achieving departmental goals is satisfactory, but not always proactive.	The person is generally available by phone and responds to emails promptly, with response times meeting most requirements. Face-to-face availability is sufficient and occurs at appropriate intervals. They show a certain willingness to help and provide support in solving internal challenges. Even if they do not always act proactively, they usually work constructively towards the department's goals.	The person is always available by phone within and, if necessary, outside their working hours. They respond promptly to emails and their response time meets requirements. The person can be reached face-to-face at appropriate intervals. They are helpful and actively support the resolution of internal challenges. In addition, they proactively engage with the achievement of departmental and company goals without losing sight of the bigger picture.
Customer satisfaction	Customers are dissatisfied with their support. Requests are often neglected and the work is perceived as unreliable and unsatisfactory. An urgent revision is required.	Customers are partially satisfied. The processing of requests is irregular and customer support is not always consistent. There is room for improvement.	Customers are satisfied with their support. Their concerns are generally dealt with appropriately, although there may be occasional delays. The work performed is perceived as acceptable.	Customers are very satisfied with their support. Most of their requests are dealt with promptly and satisfactorily. The work performed is perceived as high quality and reliable.	Customers are extremely satisfied with their support. They receive a quick and efficient solution to their concerns. The work performed is considered excellent.
Improvement of contribution margins	Contribution margin unchanged or worse	Contribution margins increased by up to 3%-points	Contribution margins increased between 3%- and 6%-points	Contribution margins increased between 6%- and 9%-points	Contribution margins increased by over 9%-points
Adherence to invoice submission deadlines	Schedule met with a delay of more than 4 months	Schedule met with a delay of up to 4 months	Schedule met with a delay of up to 3 months	Schedule met with a delay of up to 2 months	Schedule met with a delay of up to one month

Appendix B – Checklist Collaborative interfaces, project takeover and -management

1. Takeover of the project from the sales department with all rights and obligations:
 - 1.1. Was information deliberately obtained?
 - 1.2. Has responsibility been successfully assumed?
 - 1.3. Was initiative shown?
2. Folder structure:
 - 2.1. Were the documents filed appropriately?
3. Planning of personnel resources:
 - 3.1. Were the personnel resources planned on time?
 - 3.2. Was there successful coordination with the resource allocation management?
4. Timely procurement of materials and external services:
 - 4.1. Was the required material procured on time, e.g. cable purchases with quotations?
 - 4.2. Were the necessary external services ordered or utilized on time?
5. Work preparation for assembly:
 - 5.1. Was the documentation, e.g. the material removal list, successfully prepared?
 - 5.2. Were the plans made available to the employees?
 - 5.3. Were tools and other work equipment prepared on time?
6. Site supervision during execution:
 - 6.1. Was the construction site appropriately supervised during the execution of the project?
7. Intern project documentation and for the client, if required:
 - 7.1. Was the installation documentation, including the personal data bank and storage on the server, successfully created?
8. Order processing interface:
 - 8.1. Was there successful coordination and communication with the order processing department?
9. Invoicing of the project with the client:
 - 9.1. Have all necessary steps been considered for the invoicing process?
 - 9.2. Was the invoice complete?
 - 9.3. Was the best interest of the company served, if possible?

Appendix C – Criteria Information

<i>Criterion</i>	<i>Type</i>	<i>Description</i>	<i>How was it measured?</i>	<i>Controlling instrument</i>	<i>Responsible party</i>	<i>Weighting</i>	<i>Weighting</i>	<i>Points</i>
Data quality: Time tracking	qualitative	Time tracking	monthly in Tableau	Tableau report "GK-Angestellte"	Controlling	middle	0.8	1-5
Data quality: Target calculations	qualitative	Target calculations	weekly in Tableau	Tableau report "SOLL-IST Vergleich"	Controlling	middle	0.8	1-5
Data quality: Current calculations	qualitative	Current calculations	weekly in Tableau	Tableau report "SOLL-IST Vergleich"	Controlling	middle	0.8	1-5
Data quality: Payment plans	qualitative	Payment plans	weekly in Tableau	Tableau report "Zahlungspläne"	Controlling	middle	0.8	1-5
Collaborative interfaces, project takeover and management	qualitative	Most important aspects of Project Management function	Evaluation of the CEO and the head of department	Checklist "Project takeover"	CEO, Head of department	very high	1	1-5
Collegiality and team cooperation	qualitative	Availability, willingness to assist and support (new) colleagues	Evaluation of the CEO and the head of department	Behavior Scale on collegiality	CEO, Head of department	middle	0.8	1-5
Customer satisfaction	qualitative	Professional and reliable customer care	Evaluation of the CEO and the head of department; Feedback on the customer side throughout the year	Behavior Scale on customer satisfaction	CEO, Head of department	low	0.5	1-5
Improvement of contribution margins	quantitative	Positive difference between target and current values	Controlling evaluation after project closing	Tableau report "SOLL-IST Vergleich"	Controlling	very high	1	1-5
Adherence to invoice submission deadlines	quantitative	Invoicing must be done according to the payment schedule; any delay must be justified by the project manager	Controlling evaluation after project closing	Tableau report "Zahlungspläne"	Controlling	low	0.5	1-5

Appendix D – Declaration of Consent (CEO)

Consent to Participate in a Study

Title of the study: The Observation Bias in Performance Monitoring

*To the Managing Director of ***.*

I, M***** C*****, an employee of ***, would like to obtain your consent for my master thesis, which includes an experiment in our company. The aim of this experiment is to investigate the effects of monitoring on employee motivation and performance.

All information collected during this experiment will be kept strictly confidential. No names of employees, the company, the clients or the projects involved will be published in my work. Instead, a few numbers and general results will be used to maintain anonymity.

Your consent to participate in this experiment is of great importance as it will help to gain valuable insights and data. Please note that your participation is voluntary, and you can decline at any time without negative consequences.

If you agree, please sign below:

Name: _____ Signature: _____

Date: _____

If you have any questions or concerns, please do not hesitate to contact me.

Thank you for your support with this project.

Yours sincerely,
M***** C*****, BSc
E: m.c*****@***.at
T: +436608239531

Appendix E – Declaration of Consent (Participants)

Consent to Participate in a Study

Study Title: The Hawthorne Effect in Employees' Monitoring

Note: This document contains important information about the study. Please read it carefully and ensure that you understand all sections and information before proceeding. If you have any questions, do not hesitate to contact the principal investigator.

1. Background and Study Objectives:

The aim of this study is to examine the effects of monitoring on employee motivation and performance. You have been selected to participate in this study because you have been employed in the Energy Department of etn Energy and Telecom Networks GmbH for over a year.

2. Voluntary Participation:

Your participation in this study is voluntary. You can decide at any time not to participate in the study or to withdraw at a later date without any negative consequences.

3. Study Procedures:

Several aspects of your work will be measured using various methods. These aspects include data quality, project takeover, management, and documentation, collegiality, customer satisfaction, the financial performance of individual projects, and compliance with payment plans for individual projects. The applied methods for evaluation include checklists, behavioral assessment scales, control evaluations, and customer reviews.

4. Data Privacy and Confidentiality:

Your data will be treated confidentially. Your personal information will not be disclosed to third parties, and your results will remain anonymous.

5. Collection of Additional Information:

In this study, we also collect additional information about the following aspects:

Age: _____

(Please provide your current age.)

Highest Level of Education: _____

(Please specify your highest completed level of education.)

Length of Time with the Company: _____

(Please indicate how long you have been working at this company.)

6. Contact Information:

If you have questions about the study or need further information, please contact M***** C***** (m.c*****@***.at, +436608239531).

7. Consent:

I, _____, have read and understood the information sheet. All my questions have been satisfactorily answered. I voluntarily consent to participate in the study mentioned above.

Participant's Signature: _____

Date: _____

Appendix F – Shapiro-Wilk test of normality

<i>Data Table 2022</i>		$(x - \text{Mean})^2$	a_i	$a_i \cdot x_i$
Collegiality and team cooperation	1,40	3,32	0,59	0,82
Adherence to invoice submission deadlines	1,80	2,02	0,32	0,58
Collaborative interfaces, project takeover and -management	2,00	1,49	0,20	0,40
Customer satisfaction	2,20	1,04	0,09	0,21
Data quality: time tracking	3,00	0,05	0,00	0,00
Data quality: current calculations	4,20	0,96	-0,09	-0,40
Improvement of contribution margins	4,40	1,39	-0,20	-0,87
Data quality: target calculations	5,00	3,16	-0,32	-1,62
Data quality: payment plans	5,00	3,16	-0,59	-2,94
Mean/Sum	3,22	16,60		-3,82

<i>2022 inter-category</i>	<i>Values</i>
W numerator	14,60
W denominator	16,60
W	0,88
reference value	0,83
alpha	0,05

<i>Data Table 2023</i>		$(x - \text{Mean})^2$	a_i	$a_i \cdot x_i$
Data quality: time tracking	3,40	1,34	0,59	2,00
Data quality: target calculations	2,40	0,02	0,32	0,78
Data quality: current calculations	2,60	0,13	0,20	0,51
Data quality: payment plans	1,40	0,71	0,09	0,13
Collaborative interfaces, project takeover and -management	2,00	0,06	0,00	0,00
Collegiality and team cooperation	1,40	0,71	-0,09	-0,13
Customer satisfaction	2,40	0,02	-0,20	-0,47
Improvement of contribution margins	3,20	0,91	-0,32	-1,04
Adherence to invoice submission deadlines	1,40	0,71	-0,59	-0,82
Mean/Sum	2,24	4,62		0,96

<i>2023 inter-category</i>	<i>Values</i>
W numerator	0,92
W denominator	4,62
W	0,20
reference value	0,83
alpha	0,05

<i>Data Table 2022</i>		<i>(x-Mean) ^2</i>	<i>ai</i>	<i>ai*xi</i>
Participant 1	3,00	0,16	0,66	1,99
Participant 2	4,00	0,36	0,24	0,97
Participant 3	4,00	0,36	0,00	0,00
Participant 4	2,00	1,96	-0,24	-0,48
Participant 5	4,00	0,36	-0,66	-2,66
Mean/Sum	3,40	3,20		-0,18

<i>2022 inter-participant</i>	<i>Values</i>
W numerator	0,03
W denominator	3,20
W	0,01
reference value	0,76
alpha	0,05

<i>Data Table 2023</i>		<i>(x-Mean) ^2</i>	<i>ai</i>	<i>ai*xi</i>
Participant 1	2,00	0,16	0,66	1,33
Participant 2	3,00	0,36	0,24	0,72
Participant 3	3,00	0,36	0,00	0,00
Participant 4	2,00	0,16	-0,24	-0,48
Participant 5	2,00	0,16	-0,66	-1,33
Mean/Sum	2,40	1,20		0,24

<i>2023 inter-participant</i>	<i>Values</i>
W numerator	0,06
W denominator	1,20
W	0,05
reference value	0,76
alpha	0,05

Appendix G – Quantile-Quantile Plots

