STOCKHOLM SCHOOL OF ECONOMICS
Department of Accounting
BE351 Degree Project in Accounting and Financial Management
Fall 2025

Strengthening or Skewing? How benchmarking affects enabling or coercive use of the Balanced Scorecard

Harald Ivemark (26013) and Jesper Schlyter (26153)

Abstract

The aim of this study is to deepen understanding of how the Balanced Scorecard (BSC) operates when integrated with benchmarking, a combination that remains, to our knowledge, unexamined. Both systems carry the potential to be either enabling or coercive, yet we know little about what happens when their dual logics are brought together in daily managerial work. We address this gap by conducting a qualitative single-case study of "Hotelia", a large Nordic hotel chain that uses an integrated BSC-benchmarking system. Guided by Adler and Borys' (1996) distinction between enabling and coercive formalisation, and supplemented by insights on development processes and hierarchical interactions, we analyse how benchmarking shapes the BSC's usability. Our findings show that benchmarking shifted the BSC toward a more coercive use by amplifying financial indicators and weakening causal reasoning. Overall, the study contributes to research on integrated management control systems by showing that combining the BSC with benchmarking may undermine the balance the BSC is designed to achieve.

Keywords: Balanced Scorecard, BSC, Benchmarking, Enabling, Coercive

Supervisor: Kalle Kraus, Professor at the Department of Accounting, Stockholm School of Economics

Acknowledgements: We thank our supervisor, Kalle Kraus, for his structured and thoughtful guidance, as well as all the interviewees who shared their time and perspectives.

Table of Contents

1. Introduction 2

2. Literature and Theory 3

2.1 The Balanced Scorecard: Theory, Practice and Persistent Tensions 3
2.1.2 The Balanced Scorecard in Combination With Other Management Control Practices 6
2.1.3 Research Gap 9

2.2 Theory: Enabling and Coercive Control 9
2.2.1 Design Features 10
2.2.2 Development Process 12

3. Methodology 14

3.1 Research Approach 14

3.2 Data Collection 15

3.3 Data Analysis 17

4. Empirical Findings 18

4.1 Background and Context 18

4.2 Repair 20

4.3 Internal Transparency 23

4.4 Global Transparency 26

4.5 Flexibility 30

5. Discussion 32

5.1 Does the Scorecard Still Balance When Used With Benchmarking? 33

6. Conclusion 35

6.1 Limitations and Further Research 36

7. References 38

8. Appendix 43

Appendix 1. Interview Guide Templates 43

Appendix 2. Interview Table 45

Appendix 3. Organizational Chart 46

8.1 Use of Generative AI 47

1. Introduction

What matters most when organisations evaluate performance? Is it financial outcomes, non-financial drivers, or the interaction between them? With this in mind, Kaplan and Norton (1992) introduced the Balanced Scorecard (BSC), a framework intended to connect strategic objectives, non-financial performance drivers, and long-term financial results. Since then, the BSC has gained widespread adoption across industries (Huber, Kraus, and Meidell, 2025).

However, empirical studies reveal considerable variation in how the BSC functions in practice, with some finding enabling effects and others reporting more coercive uses. Research highlights persistent tensions around establishing clear causal linkages between the different BSC dimensions (Kraus and Lind, 2010; Nørreklit, 2000; 2003; Malmi, 2001), variation in how much weight managers place on financial versus non-financial indicators (Cardinaels and Van Veen-Dirks, 2010; Lipe and Salterio, 2000), and whether the framework remains forward-looking or instead becomes a static reporting tool once implemented (Naro and Travaillé, 2011). These tensions raise questions about whether the BSC succeeds in delivering the balanced and forward-looking control logic it promises.

At the same time, a recent study by Huber et al. (2025) shows that combining the BSC with another control tool, such as Enterprise Risk Management, may enhance managerial decision-making by adding perspectives that the BSC alone does not provide. Yet, the authors emphasise the need for more research on combinations of management control practices. Benchmarking is a particularly interesting candidate here. On the one hand, it might counter the BSC's inward-looking tendencies by bringing in external signals. On the other hand, prior work warns that benchmarking can trigger defensive responses and excessive focus on measures that are common, comparable and often financial (Chua, Graaf and Kraus, 2022; Knechel, Salterio and Kochetova-Kozloski, 2010). With this in mind, we set out to explore the following research question:

How does benchmarking affect whether the Balanced Scorecard functions as an enabling or coercive control tool?

To address this question, we conduct a qualitative single-case study of "Hotelia", a large Nordic hotel chain that uses a company-wide BSC alongside an extensive digital benchmarking infrastructure as its integrated control system. To analyse how this combination functions in practice, we draw on Adler and Borys' (1996) distinction between enabling and coercive formalisation. Guided by this framework, we examine four design features; repair, internal transparency, global transparency, and flexibility, to assess how benchmarking affects the usability of the BSC. We complement this lens with Wouters and Wilderom's (2008) perspective on the development process and we also incorporate insights on hierarchical interactions (Jordan and Messner, 2012; Cäker, Siverbo and Åkesson, 2022) to analyse whether the integrated system is experienced as enabling or coercive.

Our main contribution is to the literature on integrated management control practices. While Huber et al. (2025) find that integrating the BSC with Enterprise Risk Management can strengthen the BSC's strategic orientation, our study shows that combining the BSC with benchmarking can instead produce more coercive dynamics. By demonstrating that benchmarking amplifies financial dominance and weakens the BSC's causal reasoning, we give new perspectives on longstanding debates regarding the BSC's effectiveness. Additionally, we contribute to benchmarking research by answering Chua et al.'s (2022) call to examine how peer selection unfolds when benchmarking is embedded in everyday organization. Our study showed that comp set selection can be enabling or coercive depending on which hierarchical level is examined and who is included in the peer selection process.

2. Literature and Theory

2.1 The Balanced Scorecard: Theory, Practice and Persistent Tensions

In the late 1980s, scholars increasingly questioned the dominance of traditional accounting measures of performance, arguing that an exclusive focus on financial indicators provided only a short-term and backward-looking view of organizational performance (Hoque, 2014; Kaplan and Norton, 1992). In response to this, Kaplan and Norton (1992) introduced the BSC, which represents a shift from purely financial measurement toward also incorporating non-financial

measures. By connecting lagging and leading indicators, the BSC aims to align short-term actions with long-term strategic goals (Kaplan and Norton, 1993; 1996). Since its introduction, studies and surveys on the BSC suggest that it has been a widely adopted management control system (Huber et al., 2025). However, empirical studies reveal considerable variation in how the BSC functions in practice, with some finding enabling effects and others reporting more coercive uses.

The BSC organizes performance measures across four perspectives: learning and growth, internal business processes, customer, and financial (Kaplan and Norton, 1992; Hoque 2014). The learning and growth perspective captures the firm's ability to innovate and improve, providing the capabilities needed to provide future value creation (Kaplan and Norton, 1992). The internal business processes are the "critical internal operations that enable them [the company] to satisfy customer needs" (p.74). These could be measures like "cycle time, quality, and employee skills" (p.75). The customer perspective measures the factors that "really matter to customers" (p. 73), including time and quality. The financial perspective examines the extent to which an organization contributes to bottom-line improvement, serving as the ultimate indicator of a firm's "profitability, growth, and shareholder value" (p.75).

In later work, the BSC evolved from a measurement system into a strategic management system (Kaplan and Norton, 1996; Kraus and Lind, 2010). Strategy maps become central to this development by expressing causal relationships among strategic objectives. The logic is that non-financial performance drivers ultimately lead to financial outcomes (Kaplan and Norton, 2000; 2004; Kraus and Lind, 2010). Malina and Selto (2001) shows that if "the BSC factors are seen to be causally linked to each other" (p.75) it can contribute to managers reacting favourably to the BSC. In this way, clarifying the causal relationships can improve the BSC as a control tool.

However, as Humphreys et al. (2016) note, "field research in accounting has questioned the validity of these causal models for dynamic decision-making in practice" (p. 1443). Moreover, empirical studies show that organisations rarely articulate explicit cause-and-effect linkages, and even when strategy maps are adopted, managers frequently do not evaluate whether the assumed relationships actually hold (Ittner, Larcker and Meyer, 2003; Kraus and Lind, 2010; Malmi, 2001; Nørreklit, 2000; 2003).

Assuming that managers are expected to act on relationships among BSC measures, it can be observed that how these relationships play out in practice depends on which measures are included, how they are presented, and which ones managers view as most relevant for performance. Cardinaels and Van Veen-Dirks (2010) note that "managers' behavior is driven by weights placed on the performance measures" (p. 567). One consequence of these weightings can be that managers misinterpret or undervalue certain types of measures (Banker, Chang and Pizzini, 2004; Humphreys and Trotman, 2011; Malina and Selto 2004). In their experimental study, Lipe and Salterio (2000) identify a cognitive limitation in manager's BSC use: "unique measures in a business unit's BSC may be underweighted in performance evaluation" (p.293). The authors attribute this to a tendency among evaluators to prioritize common measures, typically financial, over non-financial strategy-specific indicators that are often unique to individual units. Consequently, the ways in which managers prioritise BSC measures raise broader questions about how consistently and coherently the framework is applied in practice.

Building on this, Malina and Selto (2001) found that when key BSC elements are ineffectively implemented, "focus shifts to more objective, short-term financial measures" (p.75) increasing "imbalance" in the scorecard and undermining its communication and control benefits. As Kaplan and Norton (1996) themselves acknowledge, while common measures facilitate internal comparison, non-financial and strategically differentiated indicators are essential for long-term value creation. Yet in practice, the former may crowd out the latter, limiting the BSC's potential as a nuanced control system.

While Kaplan and Norton (1996) emphasize that the BSC should be considered "a template, not a straight-jacket" (p.34) and argue that employees who understand the strategy "may innovate and find new and unexpected ways to achieve high-level strategic objectives" (p.315), empirical research suggest that such flexibility may not be realized in practice. Naro and Travaillé (2011) find that although the BSC can stimulate strategic discussion during its design phase, it often becomes more rigid once implemented. As their field study shows, the BSC "had progressively become a reporting and not a forward-looking tool," (p.229) meaning it no longer enabled questioning or updating the underlying strategy once implemented. Instead, the system settled

into a diagnostic and compliance-oriented role that struggled to accommodate strategic uncertainty. This illustrates a broader tension in the literature: while the BSC was intended to facilitate strategic learning, its use in practice may drift toward more coercive and less strategically focused forms of control.

With these tensions in mind, literature points that BSC's intended use does not always represent reality. This is similar to what Malina and Selto (2001) found in their empirical study where they outlined how the BSC can function as an effective strategic and control tool, but only under certain conditions. The authors note that "Managers react favorably to the BSC and heed its message" (p.75) when measures are seen as strategically aligned, causally linked, and accompanied by appropriate benchmarks that guide improvement. However, when "measures are inaccurate or subjective; communication about the BSC is one-way" (p.75) and benchmarks are inappropriate, the system generates tensions and perceptions of ineffectiveness. Thus, for the BSC to achieve its intended outcomes, organisations must strengthen the value-adding conditions in its design and implementation (Malina and Selto, 2001).

So, these different tensions suggest that the BSC may not fully resolve the challenges of management control on its own, raising the question of whether complementary control practices can support or strengthen its use in organizations.

2.1.2 The Balanced Scorecard in Combination With Other Management Control Practices

While the BSC is often discussed as a standalone framework, research acknowledges that management control systems (MCS) do not operate in isolation, but are often used in combination (Bedford, Malmi, and Sandelin, 2016; Malmi and Brown, 2008; Merchant and Otley, 2020). This reflects a growing understanding that control systems function as parts of broader packages, where their effects depend not only on their individual design but also on how they interact with other controls in use. As Malmi and Brown (2008) put it, "the challenge is to understand how all the systems in an MCS package operate as an inter-related whole" (p.288).

However, empirical studies exploring combinations with the BSC and other management control systems remain limited. A notable exception is Huber et al. (2025) who investigate the interplay between the BSC and Enterprise Risk Management (ERM). They find that integrating ERM into the BSC helps surface strategic risks and enriches managers' understanding of how risk exposure interacts with performance objectives which enhance decision-making. This shows that the BSC as a control tool can be enhanced when used with other control practices. Yet, Huber et al. (2025) state "much practical work goes into integrating management control practices [...] and such work deserves future scholarly attention" (p.13). This suggests that while integrating the BSC with other control practices can strengthen its use, there remains value in examining how it interacts with, and is complemented by, additional systems.

Malina and Selto (2001) emphasise that "BSC benchmarks are […] useful for guiding changes" (p.75) and that managers respond favourably to the scorecard when its benchmarks help them understand where improvements are needed. In light of this, benchmarking, with its emphasis on relative performance and the identification of best practice (Chua et al., 2022), offers one approach that may positively complement the BSC.

Benchmarking can take different forms depending on the source of comparison. External benchmarking involves comparing performance against industry peers, while internal benchmarking focuses on identifying high-performing practices within the organisation itself by comparing between similar units (Northcott and Llewellyn, 2003). There is some evidence that suggest that external benchmarking may enhance managerial decision making (Knechel et al., 2010), and Northcott and Llewellyn (2003) further state that it will "lead to performance enhancement if the external results-orientation leads to an internal examination of processes" (p. 52). However, this may, in turn, involve certain risks. Chua et al. (2022) explains that "benchmarking has induced "gaming" behaviors when organizations are being compared to (and often ranked against) others" (p.226). The authors report that earlier research shows that organizations reportedly respond to benchmarking initiatives by becoming increasingly defensive, meaning that instead of being forward-looking and learning-oriented, they become more concerned about protecting their existing position. This contrasts with the BSC's aim to

support long-term thinking and learning, raising questions about how defensive responses to benchmarking may interact with the BSC framework's intended role.

Given concerns that benchmarking may risk reinforcing results-oriented behaviour, it can become relevant for organisations to consider forms that emphasise process learning, most notably cooperative benchmarking. This is where several firms "voluntarily share information about selected areas of their operations in order to identify and learn from the best practices" (Elnathan and Kim, 1995, p.346). Thus, the choice of appropriate peers becomes relevant. As Chua (2022) explains, peer selection shapes the very outcomes of benchmarking, since the choice of peer groups affects the "identification of aspirational performance and learning" (p.226). Put differently, benchmarking and the selection of peer groups can shape the criteria managers use to assess performance.

In addition to selecting appropriate peers, organisations must also determine which performance measures are suitable for benchmarking. Limited availability of reliable external data means that some indicators can be benchmarked while others cannot, constraining the scope of comparison (Knechel et al., 2010). This difficulty in finding directly comparable information can also shift attention toward common measures, as Knechel et al. (2010) note that "performance measures that are common across units have been shown to be more influential […] because unit-specific measures are not directly comparable across units." (p. 318). This tendency can be noteworthy in the context of the BSC, where empirical studies have shown that financial indicators may attract disproportionate attention in practice (Cardinaels and Van Veen-Dirks, 2010; Lipe and Salterio, 2000).

At the same time, Chua et al. (2022) provide a counterexample to the concern that benchmarking merely encourages convergence toward average performance, citing Siverbo and Johansson (2006) that relative evaluations can "reduce uncertainty regarding an organization's quality, productivity and efficiency" (p. 226) and thereby support more targeted improvement efforts. This suggests that benchmarking may either increase reliance on common measures or, under certain conditions, help reduce uncertainty in areas such as quality and productivity, metrics often used in the BSC.

2.1.3 Research Gap

The review above shows that BSC as a control tool may be strengthened when complemented by other practices, as shown by Huber et al. (2025). However, we find that research often studies the BSC in isolation, and empirical work on its integration with other controls remains scarce. There is therefore value in exploring how the BSC interacts with complementary practices.

Under certain conditions, the BSC enables learning and strategic reflection, while in others it becomes more coercive, reinforcing internal focus and financial dominance. Literature reveals that benchmarking carries similar dual possibilities: it can broaden attention, stimulate learning, and introduce external perspectives, but it can also intensify pressure and standardisation.

So, working with the two control practices together seems to create a conceptual tension that remains unexplored: does benchmarking complement the BSC by enhancing strategic adaptation and learning, or does it risk reinforce the very tendencies, financial dominance, short-termism and loss of strategic nuance, that the BSC seeks to overcome? To investigate this tension, we draw on Adler and Borys (1996), whose framework offers a lens for analysing when practices function in enabling ways and when they take on more coercive qualities.

2.2 Theory: Enabling and Coercive Control

To analyse whether benchmarking complements or constrains the BSC as a control tool, we adopt a theoretical lens that explains the conditions under which systems become supportive or restrictive. Adler and Borys (1996) offer such a perspective by distinguishing between enabling and coercive forms of formalization. The authors argue that coercive formalization seeks to secure employee compliance which in turn reduces job satisfaction and innovation, whereas enabling formalization is designed so that employees experience rules and systems as supportive of their work. According to Adler and Borys (1996), these forms of formalization can be differentiated along three dimensions: 1) the design features, 2) the process through which the system is designed, and 3) the implementation of the system.

Ahrens and Chapman (2004) were the first to apply Adler and Borys' (1996) framework to MCS. Their study of a restaurant chain illustrated that "processes of coercive formalization existed side by side with processes of enabling formalization" (p.20). Later research highlights the hierarchical elements on whether a control system becomes enabling or coercive. Jordan and Messner (2012) show that enabling use can emerge when superordinate managers engage in continuous interplay with subordinates. Building on this interpersonal perspective, Cäker et al. (2022) analyse how hierarchical accountability structures influence enabling and coercive uses of PMS. As the authors observe, "foreground interactions allow superior managers to learn about local aspects […] within a contextualised understanding" (p.871), while background accountability processes "keeps opportunistic, inappropriate, and substandard behaviour under control" (p.871). This illustrates that control systems can be important for superior managers' control while still being experienced as enabling by subordinate managers.

While these studies concentrated primarily on Adler and Borys' (1996) design features, work by Wouters and Wilderom (2008) puts attention toward the design and implementation process. The authors describe that "design and implementation are hard to distinguish and we prefer to combine them and use the phrase "development process" (p.490). The reason is that these activities happen at the same time in practice: as people start using the system, they learn from it and then adjust the design, which in turn leads to more changes in how it is used (Wouters and Wilderom, 2008).

In what follows, we focus first on the four design features and then the development process.

2.2.1 Design Features

The usability of formal systems can be assessed in terms of four features, namely: repair, internal transparency, global transparency, and flexibility (Adler and Borys, 1996). These can take either enabling or coercive forms. Ahrens and Chapman (2004) show that these features are "directly relevant to the study of management control systems" (p. 3). The design features can therefore assist in exploring whether benchmarking makes the BSC more enabling or coercive. In the following, we describe each feature in turn.

Repair concerns the extent to which users can identify and correct breakdowns in formal systems (Adler and Borys 1996). In coercive systems, any "deviation from standard procedure is seen as suspect" (p.71). In contrast, the enabling logic emphasises procedures that can respond to real work situations, encouraging employees to raise issues with rules and standards so that breakdowns become opportunities for learning rather than failures (Adler and Borys, 1996; Ahrens and Chapman, 2004). In the context of control systems, this can for example mean that "managers have the permission and ability to modify the definition and measurement of performance indicators, if deemed appropriate" (Jordan and Messner, 2012, p.546). Against this background, Ahrens and Chapman (2004) suggest that repair can help prevent control systems from becoming misaligned with operations.

The second enabling feature is internal transparency which concerns the extent to which users can understand the internal logic of the system and the processes they work with (Adler and Borys 1996). In coercive systems, procedures are formulated as "lists of flat assertions of duties" (p.72) that offer little insight into why rules, indicators, or targets look the way they do, thereby serving managers more than users. In contrast, enabling procedures provide users with "visibility into the processes they regulate by explicating its key components and by codifying best-practices" (Adler and Borys, 1996, p.72). Employees are also provided with feedback on their performance by "providing metrics that help users assess their performance against historical standards" (p.72). In a management control context, studies show that internal transparency supports knowledge-sharing and enables managers to understand how performance is produced, diagnose issues, and use measures for improvement rather than mere compliance (Ahrens and Chapman, 2004; Jordan and Messner, 2012; Cäker et al., 2022).

Global transparency differs from internal transparency by concerning the extent to which users can understand the broader organizational context in which their specific tasks are embedded (Adler and Borys, 1996; Ahrens and Chapman, 2004). In coercive systems, access to organization-wide information is restricted and asymmetrical and employees who look beyond their assigned task may be told "that's not your job" (Adler and Borys, 1996, p.73). In contrast, enabling systems provide users with wider organizational information to help them understand how their work contributes to broader organizational goals (Jordan and Messner, 2012). By

sharing key targets and aggregated measures across departments, coordination is supported rather than limited to senior management (Ahrens and Chapman 2004; Cäker et al., 2022).

Flexibility refers to how easily users can adapt formal procedures and control information to the specific situation they face (Adler and Borys, 1996; Ahrens and Chapman, 2004). In coercive systems, procedures prescribe a fixed sequence of steps and require supervisory approval for any deviation. The underlying assumption is that "the manual prescribes, the employee implements" (Adler and Borys, 1996, p.74). Enabling systems instead recognise that not all contingencies can be anticipated and therefore designs procedures that users can adjust depending on the circumstances. Ahrens and Chapman (2004) show that flexibility in management control can be supported through customizable reporting tools and tailored performance information and Cäker et al. (2022) further states that superior managers can "influence flexibility by allowing access to the systems" (p.868).

Viewed through the lens of design features, the use of the BSC and benchmarking can be interpreted as shaping how transparency, repair, and flexibility are experienced in practice, either in enabling or coercive ways (Adler and Borys, 1996). Ahrens and Chapman (2004) helps put these design features in a control context. For example, benchmarking may create or reinforce the BSC's flexibility so that managers either use comparative insights to search for new performance measures, or they may become constrained by a narrow set of standardised ones.

We will extend Ahrens and Chapman (2004), by also analyzing the interaction between managers at different hierarchical levels and its implications on the four design features (Jordan and Messner, 2012; Cäker et al., 2022). This allows us to examine whether benchmarking encourages dialogue and negotiation that enables managers to use the BSC and benchmarking in a more dynamic way, or if it will push the system to be more competitive and less dynamic.

2.2.2 Development Process

Adler and Borys (1996) argue that whether a formal system becomes enabling or coercive does not depend on its features alone, but also on how it is designed and implemented. Wouters and Wilderom (2008) propose "three characteristics of a development process that is likely to result

in an enabling PMS" (p. 492). These are: experience-based development process, experimentation, and professionalism.

Experience-based development processes involve identification and appreciation of existing local knowledge. Wouters and Wilderom (2008) explain that in practice, "organizational change processes that utilize local knowledge are more likely to lead to [...] improvements" (p.493). Experimentation refers to the iterative creation and refinement of performance measures until it fits local conditions (Wouters and Wilderom, 2008). This can include testing and refining a measure's definition, data requirements, and presentation in several rounds together with employees. Professionalism refers to "an orientation toward learning for the purpose of improving work practices," (p.493) which reflects employees' commitment to continuously upgrading the quality of their work through ongoing development and improvement efforts. Together, these elements support a development process that produces an enabling rather than coercive PMS.

Wouters and Wilderom (2008) explain that PMS is typically designed from the perspective of top management. However, they note that these initiatives are less likely to be successful if, for example, the top managers overrule the local unit managers' experience-based knowledge. This suggests that the dynamics between superordinate and subordinate managers are relevant to whether a PMS is experienced as enabling or coercive.

The development process offers us a lens to not only look at how benchmarking is intended to complement the BSC, but also how it is actually used over time. Because benchmarking can introduce new comparisons, targets, and interpretations over time, its influence on the BSC may not fully be understood from design features alone. Wouters and Wilderom (2008) helps us analyse whether these developments are shaped by local experience, experimentation, and professionalism, or whether they are imposed in more top-down, coercive ways. Thus, by analysing the development process, we can see whether benchmarking becomes a driver for learning and refinement, or whether it makes the BSC increasingly rigid over time through standardisation, target fixation, and competitive pressure.

So, the four system features (Adler and Borys, 1996; Ahrens and Chapman, 2004; Jordan and Messner, 2012; Cäker et al., 2022) provide a lens for analysing how benchmarking influences the way the BSC is used in practice. By supplementing this with the development-process perspective (Wouters and Wilderom, 2008), we also capture elements of how BSC-Benchmarking practice evolves over time. Together, this allows for a comprehensive assessment of how and whether benchmarking makes the BSC more enabling or coercive.

3. Methodology

3.1 Research Approach

The aim of this study is to deepen understanding of how the BSC operates when integrated with benchmarking, a combination that remains, to our knowledge, unexamined. As outlined under section 2.1, both systems carry the potential to be either enabling or coercive, yet we know little about what happens when their dual logics are brought together in daily managerial work. By exploring this integration in practice, we seek to illuminate how benchmarking shapes the way managers experience and use the BSC. To investigate this, we address the following research question:

How does benchmarking affect whether the Balanced Scorecard functions as an enabling or coercive control tool?

To answer this, we adopt a qualitative research approach. The motivation is that qualitative methods are appropriate when the aim is to understand how management controls are interpreted and enacted in everyday organisational life, rather than to measure predefined variables (Vaivio, 2008). Since our study seeks to capture managers' experiences of how the BSC and benchmarking interact in practice, a qualitative approach is necessary to assess the meanings, tensions, and judgments that arise around these systems. The primary level of analysis is the organizational level, focusing on the design and use of benchmarking systems within Hotelia, complemented by individual-level insights from BPMs and hotel managers."

We employ a single-case study design as it provides the depth required to analyse managerial practices and interactions across organisational levels. Case studies are particularly suitable for exploring contemporary phenomena within their real-life context and are recommended when existing theory is limited or underdeveloped (Eisenhardt, 1989). Additionally, a single-case design is appropriate when the case provides unusually strong opportunities to observe the phenomenon of interest (Eisenhardt and Graebner, 2007). Hotelia offers such conditions as it uses both a BSC and a benchmarking infrastructure, enabling a detailed examination of their interplay. A single-case design is further compatible with our method theory, as Adler and Borys' (1996) framework requires close, contextual analysis of how systems are enacted and used in practice to assess whether they function in enabling or coercive ways.

3.2 Data Collection

Data for this study was collected through semi-structured interviews and internal documents provided by Hotelia. Semi-structured interviews are well suited for in qualitative research as they allow for exploring predefined themes while remaining open to ask follow-up questions when unanticipated insights emerge (Dumay and Qu, 2011). This openness was particularly important in our study, as our understanding of the interplay between the BSC and benchmarking evolved throughout the early stages of data collection. While we entered the field with a general focus on performance measurement practices, we did not yet know which aspects managers themselves would perceive as most relevant. The semi-structured format therefore enabled us to follow interviewees' reasoning, probe emerging issues, and adjust the conversation in response to what they emphasized as central in their daily work, turning out to be the BSC and benchmarking. Our interview guide mirrored this evolving understanding: we began with a broad set of questions on management control, but after the first interviewee made it clear that the BSC and benchmarking formed the core of Hotelia's control system, we refined the guide to explore this dynamic more in depth (see Appendix 1).

In total, nine semi-structured interviews lasting 40-60 minutes each were conducted during the autumn of 2025. Participants were selected using a combination of purposive and snowball sampling to ensure access to managers with relevant expertise while also allowing additional key actors to be identified through referrals as our understanding of the case developed. Initially, we

reached out to Hotelia's Business Performance Managers (BPMs) and business controllers, as their roles are directly tied to performance measurement. After the first interviews, additional informants were identified through snowball sampling by asking interviewees whom they believed we should speak to next. In line with Atkinson and Flint (2001), this approach proved effective in widening access: once one employee opened the door to the organization, others became increasingly willing to participate and offer support.

However, snowball sampling can introduce selection bias through network-based referrals (Atkinson and Flint, 2001). To mitigate this, we explicitly requested referrals to individuals beyond the BPM function, for instance hotel managers, to broaden the perspectives included. In line with this, we also actively sought access to different hierarchical levels to avoid restricting our insights to a single division or managerial layer. Through this combination of purposive and snowball sampling, we gained access to a wide range of roles, from Vice Business Performance managers to hotel managers (see Appendix 2 and 3). All interviewees are anonymised through city-based pseudonyms.

The interviews were conducted online via video call. This format was chosen based on interviewees' preferences and geographical locations, as several participants were based in Norway. While online interviews reduce non-verbal cues, we judged that the increased feasibility and broader access mitigated this limitation. All interviews were recorded with participants' consent and transcripts were generated through Microsoft Word. One interview was conducted in English whilst the others were conducted in either Swedish or Norwegian, depending on the interviewee's preference, and then later translated into English for analysis. Although translation may introduce minor interpretive shifts, allowing participants to speak in their native language was prioritised to enable more authentic and detailed reflections.

In addition to interviews, we obtained internal documents related to Hotelia's BSC, organizational structure, and benchmarking database. These documents served to contextualize interview insights and ensure consistency in our understanding of how BSC indicators and benchmarking data are formally defined, used, and communicated within the organization.

3.3 Data Analysis

Our analytical process followed an abductive approach in which theory and empirical material were developed in parallel. As Ahrens and Chapman (2006) note, "problem, theory, and data influence each other throughout the research process. The process is one of iteratively seeking to generate a plausible fit between problem, theory, and data" (p. 836). This accurately reflects how our analysis unfolded. We did not enter the field with a fixed research question; instead, our understanding of the interplay between the BSC and benchmarking evolved as interviews progressed. Early conversations revealed that these two systems were more interdependent than anticipated, which prompted us to refine both our interview guide and analytical focus. This in turn led us to refine our theoretical lens and adopt Adler and Borys (1996) framework of enabling and coercive formalization. Iteratively moving back-and-forth in this way resembles Dubois and Gadde's (2002) notion of systematic combining, where empirical insights continuously inform theoretical framing and vice versa.

When analysing the data, we employed a thematic approach to identify, interpret, and organize patterns within the data (Braun and Clarke, 2006). This approach is particularly suited for exploratory research, as it allows for identification of themes that capture important aspects of the data in relation to our research question. Following Braun and Clarke's (2006) guidelines, we began with familiarization through repeated readings of transcripts, making short memos to capture early impressions. These initial codes were intentionally broad and descriptive, reflecting our aim to capture the full range of experiences and issues raised by the interviewees. Once all transcripts had been coded, we organised the codes under Adler and Borys' (1996) four design features, which served as our overarching themes. The broad initial impressions were therefore consolidated into these four categories, and within each we reviewed the coded segments to identify recurring sub-patterns. This process enabled us to capture how benchmarking shaped the enabling or coercive tendencies of the BSC across different managerial roles. In the following section, these themes structure our empirical analysis and form the foundation for interpreting our findings.

4. Empirical Findings

4.1 Background and Context

Hotelia is a private Nordic hotel operator running more than 200 hotels, having 18,000 employees and generating a revenue of approximately SEK 18 billion. The organization is divided into four major divisions: City Services, Core Services, Signature Services, and Destination Services. Each of these divisions have its own business model, guest segment, and operational logic, resulting in different revenue drivers, cost structures and local operating conditions. As Helsinki explained:

> "City and Core Services cover our more traditional hotel operations, where some properties focus mainly on room revenue and others rely more on food, beverage, and conference business. [...] Signature and Destination Services represent our more high-end and experience-driven concepts, with these hotels varying a lot among themselves."

Each division has its own business performance team where each BPM is responsible for approximately 15 hotels. A BPM's work includes "setting targets and forecasts, building comp sets and helping the hotels to analyse their results" (Helsinki). As part of their work, BPMs conduct monthly performance meetings, with the respective hotel manager where they discuss historical performance, projections, and operational concerns. More sporadically, they hold workshops to discuss the underlying reasons behind these numbers and translate the insights into their strategic work. The BPMs also report divisional performance directly to the CFO on a monthly basis.

As part of the BPMs performance work, they use two management control systems, namely the BSC and benchmarking. In the interviews, it was evident that Hotelia used both practices as an integrated control system. Oslo explained that "We start with the BSC because that's the structure, and then benchmarking is part of helping us make sense of whether our performance is actually good or not". Consequently, when referring to "Hotelia's system" or "control system" henceforth, it denotes the use of these two practices together.

Looking at Hotelia's BSC, it is structured around five strategic areas, each comprising specific measures and associated targets (see figure 1). Helsinki explained that "the BSC is according to our guidelines used to balance financial and non-financial indicators". Hotelia incorporates two financial indicators in its BSC: the Revenue Generating Index (RGI), "calculated as revenue per available room (RevPAR) divided by the aggregated RevPAR of a competitive set" (Oslo), and Gross Operating Profit (GOP). In addition, the BSC includes three non-financial indicators: the Net Promoter Score (NPS), the Employee Net Promoter Score (eNPS), and Hotelia's proprietary sustainability index. The BPMs and hotel managers use these "dimensions to pinpoint what is driving our results and what we need to adjust going forward and to discuss current performance" (Helsinki). Performance is visualised in a way where measures that fall below target or expected trend are marked in red, while those that meet or exceed expectations are shown in green.

Strategic Area:	**Sustainability**	**Guest Experience**	**Market Performance**	**Profitability**	**People**
Measured as:	Hotelia's own Sustainability Index	Net Promoter Score (NPS)	Revenue Generating Index (RGI)	Gross Operating Profit (GOP)	Employee Net Promoter Score (eNPS)
Operations target:	Target	Target	Target	Target	Target
Trend:	Current performance vs. Target	Current performance vs. Target	Current performance vs. Target	Current performance vs. Target	Current performance vs. Target

Table 1: Hotelia's Balanced Scorecard (Anonymized)

As part of this system, Hotelia complements its performance monitoring with benchmarking, which forms a recurring input in managerial discussions. As Oslo noted, "we are [...] using benchmarking to end up with our BSC measures, both internal, like between our hotels, and

externally against competitors and the market," while Helsinki observed that "benchmarking and the BSC are closely related" when explaining how comparative information is embedded in everyday control. Internally, hotels are compared across all BSC dimensions using shared performance data to assess how units perform relative to one another. Externally, Hotelia is part of the "Benchmarking Alliance", a Nordic cooperative data-sharing platform that "provides daily comparative information on key financial indicators, such as RevPAR and revenue per operational labor unit" (Bergen). This data is derived from comp sets that BPMs construct themselves, typically by "selecting hotels with similar characteristics, market positions, and customer segments" (Oslo). Such comprehensive, competitor-level transparency that Benchmarking Alliance provides was explained to be "unusual outside the hotel industry" (Kalmar).

We will now analyse the empirics using the four design features as the main structure, complemented by insights from the development process.

4.2 Repair

During the interviews, repair work emerged as a key feature of how the integrated control system is used in daily practice at Hotelia. Rather than treating the BSC and Benchmarking framework as fixed, BPMs actively adjust and reshape it to improve it. Their repair practices unfold in three ways: first, by breaking down and modifying standardized BSC measures when they feel these do not support their work; second, by experimenting with new ways of combining and visualising BSC and benchmarking information to create more actionable insights; and third, by revising the benchmarking tool itself through ongoing adjustments to the comp sets.

Hotelia's BSC framework appeared uniform across all divisions and hotels. This standardization applies not only to the five dimensions but also to the specific measures and indexes included in the BSC. However, given the heterogeneous nature of the hotels, the standardized framework is intended to provide a broad representation of performance that can apply across this diversity. For the BPMs, this means that the strategic areas and corresponding measures "capture a lot of performance, but not in a way that reflects the different units' unique circumstances or allows for meaningful local interpretation" (Sundsvall).

However, instead of creating frustration and seeing the BSC as inherently problematic, BPMs emphasize that the system gives them room to interpret and adapt the measures so they become usable in their own contexts. Bergen explains that they are not expected to "pay close attention to the indicator itself" but rather to adjust and refine it when needed. Lund built on this saying, "headquarters sets the structure, but we decide how to make it usable". This means that incompleteness does not automatically create pressure or rigidity; it becomes a starting point for local problem-solving rather than a constraint. Oslo describes this as a normal part of her work:

> "The measures we get centrally, such as GOP, are a good starting point, but they're too broad. We always need to break them down into smaller pieces so they make sense for the day-to-day work in the hotels in my comp set."

This repair work means that the standardized and sometimes incomplete BSC can still be experienced as enabling. BPMs do not feel restricted in their ability to repair the measures, nor do they encounter the kind of reinforced emphasis on standardized BSC indicators that might otherwise produce a more coercive experience (Jordan and Messner 2012).

During the interviews, it appears that the BPMs experiment a bit with the system when they feel like the company-wide measures do not produce desired outcomes. Hotel managers often describe "difficulty understanding which measures to prioritise within the BSC" (Sundsvall). To address this, Sundsvall and Stockholm began experimenting together, developing a visualisation tool that today helps them to weigh the measures against one another and highlights which areas require the most attention by integrating benchmarking data. Instead of simply showing red or green boxes, the tool makes the BSC more actionable by using benchmarking data to clearly indicate where improvement efforts would have the greatest impact on operations:

> "We have developed and are still refining a type of spider chart that visualizes how the measures relate to each other, not just which ones are red or green. [...] When you look at it, you immediately see what measures you're strong on like maybe GOP and RGI, but

relatively weak on NPS, and then you understand which measures you need to work more with."

This illustrates how BPMs are experimenting with more useful ways of visualising and prioritising performance. Their willingness to draw on their own operational experience and adjust the tool iteratively reflects Wouters' (2008) argument that enabling systems emerge through both experience-based development and ongoing experimentation.

Furthermore, several BPMs in other divisions describe how unusual spikes in red or green results in the BSC often signal not an operational issue but a mismatch between the hotel and the peer group it is compared against. In these moments, the repair work involves reassessing which hotels belong in the comp set and whether their performance reflects comparable, local conditions. Helsinki recalled a case where "a hotel's BSC figures suddenly dropped relative to its peers", not because the hotel had performed worse, but because seasonality effects had boosted the results of only a few hotels in the comp set. To correct for this distortion, Helsinki initiated repair work by adjusting the composition of the comparison group:

"[...] from one quarter to the next I decided that some hotels no longer made sense in the comp set, and replaced them with some others instead. [...] You constantly need to engage in that kind of work to better reflect what is really going on."

Moreover, Helsinki gives an example of a nearby hotel benefiting from a major concert announcement, explaining that "suddenly all rooms sell out on a certain date a year from now […] and you get five times the revenue with the same cost base". When such distortions occur, BPMs remove outlier hotels from the comp set and add others that they believe better reflect local market conditions. This can make the measures in the BSC updated to better reflect reality.

At the same time, given that BPMs have responsibility in selecting and adjusting comp sets introduces a risk that repair practices may drift away from accurately reflecting local conditions. Helsinki, who oversees several BPMs in the City division and ultimately evaluates their reported performance relative to these self-defined peer groups, expresses concerns that not all repair work

serve the same purpose. As he put it, “I’m not always sure whether a comp-set change is about improving the data or just improving the numbers.” His concern is that adjustments framed as correcting misrepresentation could just as easily serve to present performance more favourably. In these cases, repair shifts from enhancing usability to distorting reality, creating a more coercive dynamic where BPMs feel pressure to “make the figures look right” (Helsinki) rather than ensure they are right.

Taken together, the interviews show that repair practices play a central role in how the BSC and benchmarking are made usable in Hotelia, and they reflect enabling tendencies while also raising questions about accuracy of the integrated control system. BPMs and hotel managers are not required to follow the standardized indicators rigidly; instead, they routinely break down broad measures, experiment with new visualisations, and adjust comp sets when the system fails to capture local conditions. These practices can help preserve local relevance and prevent the control system from becoming too detached from day-to-day work. However, the same type of repair practices that can support meaningful interpretation also introduces a coercive risk. Because BPMs define comp sets, repair can drift from correcting distortions to strategically improving numbers, raising questions around whether data are being improved or merely made more favourable. Thus, while repair enhances the usability of Hotelia’s control system, it also exposes a tension which was reflected in Helsinki who were cautious about whether the repair practices were for the better of the organisation.

4.3 Internal Transparency

During interviews, internal transparency emerged as an important part of how BSC and benchmarking is being used in daily work at Hotelia. Rather than simply reporting results, the system appears to provide different forms of “visibility” into how performance is produced and how managers understand the logic of the control tools. The role of the integrated control system becomes evident through three central themes: the monthly performance meetings, the joint workshops, and a difference in clarity between financial and non-financial indicators.

Across Hotelia, the divisions follow a broadly similar structure in how they review performance, evaluate past results, set goals, and project future outcomes. Each month, the BPMs and hotel

managers convene to examine performance. These meetings were consistently described as a forum for clarifying expectations, talking through monthly results, and establishing a shared understanding of how hotels perform relative to both internal targets and market movement. As Kalmar put it, “the meetings are fundamentally about forming a picture of whether the hotels are positioned as expected”.

For BPMs, the meetings appeared to create visibility into how hotel managers interpreted their own numbers and what they saw as driving performance. Several BPMs described these as occasions where they gained valuable day-to-day insight that cannot be retrieved from dashboards alone. As Bergen framed it, the meetings make it possible “to understand what has happened behind the numbers, not just the numbers themselves”. The hotel manager instead described the meetings in similar practical terms but emphasized different benefits. Trondheim explained:

> “I see the meetings as opportunities to explain why we might have red numbers on certain measures [...]. I think that the meetings provide me with clarity about what is expected of me and also actually which indicators mattered most for the BPM when we talk through the BSC.”

BPMs frequently described how benchmarking shaped these conversations by providing a reference point for interpreting the BSC measures. The divisions can benchmark all BSC measures internally, but only the financial ones externally through the benchmarking alliance. Stockholm noted that “the numbers don’t mean much on their own, benchmarking tells you whether you’re actually behind or ahead”, showing that benchmarking gives the BPM a clearer basis for assessing performance. At the same time, the hotel managers describe that they get a better sense of what constitutes good or poor results, noting, “A measure that looks bad compared to our goal can still be good compared to competitors” (Lund). In this way, benchmarking appears to enable both parties’ understanding of performance itself, indicating that it, together with the BSC, is interpreted as beneficial for both superordinate (BPMs) and subordinates (hotel managers) (Cäker et al., 2022).

However, the monthly performance meetings are not the only occasions where BPMs engage with the hotel managers. At regular intervals, BPMs visit hotels to run workshops aimed at understanding the drivers behind performance outcomes. As Stockholm described, these workshops are designed to translate numbers into understanding:

> "I go out and hold workshops at the hotels and first identify problems or challenges we can see through our benchmarking figures [...] then try to take it all the way out to some kind of action at the hotels."

Here, the focus is not only performance but the reasoning behind it. Lund emphasizes the value of these sessions explaining how "the BPMs bring the numbers, I [hotel manager] bring the context, and then we together try to figure out what actually happened".

While the workshops can help in illuminating how different BSC dimensions connect, several interviewees highlight that this internal transparency does not extend to all parts of the system. In particular, many describe persistent difficulties in understanding how the non-financial indicators are constructed. These measures, although central to Hotelia's strategic ambitions, often lack a clear internal logic that BPMs can follow. As Helsinki explained, even though NPS appears as a standardized indicator internally, "no one really knows how the NPS score is calculated", and therefore they may not even take a look at relative performance on that measure. The same confusion applied to Hotelia's other non-financial measures:

> "I can quite confidently say that no BPM you will talk to can explain the logic behind our sustainability index. Sometimes the non-financial targets just feel fictional […] someone puts their finger in the air and says 'this is good'." (Sundsvall)

Because the underlying calculation remains unclear, these indicators risk becoming detached from the operational work they were meant to reflect. Instead of helping BPMs understand the reasons behind non-financial performance, they create uncertainty about how their units are expected to influence them. As Stockholm noted, "of course benchmarking helps us see how we're doing, but it's not useful when the [non-financial] measure itself doesn't make sense."

In contrast, financial indicators such as GOP and RevPAR are consistently described as more intuitive, not because the causal linkages in the BSC make them clearer, but because other drivers are better understood by the BPMs. Bergen explains that "with the financial measures in the BSC, you always know what drives them […] you can break them down and see where it went wrong." This clarity makes the financial metrics far easier to use in both workshops and monthly meetings, and several interviewees note that these measures tend to dominate discussions. As Stockholm put it, "it's simply easier to talk about what everyone understands." In effect, the lack of internal transparency around the non-financial indicators reinforces a focus toward the financial side of the BSC.

Taken together, the performance meetings and workshops show that benchmarking enhances internal transparency mainly for the financial indicators. By placing results in a comparative context, how a hotel performs relative to other units and competitors, benchmarking makes it easier for both BPMs and hotel managers to interpret financial outcomes. These measures were widely viewed as intuitive, objective, and up to date, since they could be broken down into familiar drivers and verified through external comparisons. In contrast, the non-financial indicators generated more friction: their underlying calculations were more unclear, they lacked external benchmarks, and their meaning felt more subjective. As a result, they were rarely used to explain performance, and discussions tended to focus on the financial side of the BSC. So, benchmarking made financial measures more transparent and actionable, but it seems to do little to clarify the non-financial ones, creating more coercive elements and weakening their perceived usefulness.

4.4 Global Transparency

Empirics highlight three themes where global transparency becomes especially noticeable: how managers use organisation-wide comparisons to interpret local results, how shared performance information can shift from supporting learning to fostering competition, and how top management's emphasis on benchmarked indicators shapes the targets experienced at lower levels.

Given Hotelia's multi-division structure and its use of a company-wide BSC with strategically important and globally applied measures, BPMs must continually update their understanding of performance across the organisation. Combined with the use of internal and external benchmarking, this requires BPMs to maintain a holistic view rather than focusing solely on their own units. This was evident in all divisions. Rather than evaluating results in isolation, the BPMs routinely step back to interpret local deviations in relation to the organization's goals:

> "The thinking is we're up in the helicopter asking: okay, red or green? Are the hotels performing in line with the organisation's goals or not? If we are, great move on. If not, we need a deep dive and then you start looking, which geographies are growing less? Which hotels are laggards? Then you go down and try to understand it." (Helsinki)

This illustrates a core aspect of global transparency: the ability to situate local performance within a wider organisational frame (Adler and Borys, 1996; Ahrens and Chapman, 2004). What enables this helicopter perspective appears to be Hotelia's benchmarking platform, which allows managers to access the performance of any hotel, comp set, or division at any time. Lund described this visibility as follows:

> "You can choose whatever Hotelia hotel you want and compare whatever KPIs you want. This gives a picture of not just our hotel and comp set, but everyone at the same time. [...] You can of course also compare some financial indicators to competitors through the benchmarking alliance tool."

On top of this, hotel managers and BPMs receive daily reports with internal metrics like "housekeeping productivity and breakfast revenue" (Trondheim), paired with external data on "how many rooms competitors sold and at what price" (Lund). The purpose of these reports is comparative: they allow hotel managers to see, each day, how their hotel performs relative to both internal peers and the market. By making wider performance transparent, benchmarking enables managers to interpret whether a deviation reflects a unit issue, a divisional trend, or a broader market movement.

However, several interviewees also describe how a consequence of these shared reports and databases is a more competitive environment where the daily group-wide report become a leaderboard:

> "The hotel managers love receiving the report that shows how all the hotels in their group performed each day. They check it every morning to see who is ranked first, second, third, and so on. [...] They really enjoy competing with one another." (Sundsvall)

What first resembles an enabling mechanism for sharing organisational information slips into a "game-like" behavior where hotel managers start working toward outperforming peers rather than understanding the strategic priorities that underlay the reports. While this may drive the hotel managers to work harder, it instead seems to have been at the expense of sharing operational insights:

> "With the reports being shared so widely now, I feel like my colleagues are a bit more careful about what they put forward. It's not that we don't want to help each other, but the comparisons are right there for everyone to see, so you naturally start thinking about how you're positioned. [...] I guess it changes the way we talk about what we do." (Trondheim)

In this way, benchmarking's organisation-wide visibility drifts into coerciveness, shaping behaviour in ways that reduce willingness to share performance insights openly.

Moreover, several interviewees highlight that Hotelia's targets are largely shaped by what in the BSC the headquarters and top management choose to focus on. This emphasis, in turn, is influenced by the benchmarking figures that signal what "success" looks like in the market. As Helsinki explained, "Our CEO looks at competitors' financials, and may see they have a 10% EBITDA. So he wants it too [...] This goal is then shared in the organisation." As a result, hotel managers can feel that important local priorities fall outside senior management's (and BPMs') strategic focus, follow-ups, and performance implications.

> "I would have wished that the result goals that are set reflect more of the non-financial parts of the BSC. I mean, the human aspect and the cultural work, the things that matter locally. [...] We spend a lot of time building culture here at this hotel, and that's also a big part of what makes Hotelia what it is, but I often feel that those priorities get lost in the result expectations." (Trondheim)

Moreover, because top management concentrates on the financial indicators that can be externally benchmarked, these goals become selectively visible and can overshadow other important aspects of performance. Trondheim emphasized that cultural work is not only intrinsically important but also affects financial outcomes in the long run, stating "if you push people too hard they become sick-listed […] the costs for sick absence are sky-high. And that could have been avoided if instead you gave a little more slack and had even more focus on the culture." In her view, the organisation's focus fails to reflect how soft dimensions such as staff well-being (eNPS), shape the very financial results that headquarters prioritises. This disconnect illustrates how the top management's global goals become incomplete and coercive, limiting hotel managers' ability to articulate how their local initiatives contribute to long-term organisational goals (Jordan and Messner, 2012).

Overall, our findings show that global transparency shapes whether the Hotelia's control system becomes enabling or coercive. On the enabling side, having access to performance data from all hotels helps BPMs and hotel managers place local results in a broader context and understand how their units contribute to Hotelia's goals. However, this transparency also brings challenges. The daily benchmarking reports often function as leaderboards, which seem to shift attention from learning to competing with peers. In addition, because external benchmarking mainly covers financial indicators, top management tends to emphasise financial targets, even though the BSC formally highlights several non-financial strategic areas. This emphasis trickles down through the organisation and makes it difficult for hotel managers to prioritise non-financial work, such as culture and staff well-being, even when these are understood to influence long-term results. In this way, global transparency moves the system toward more coercive outcomes by reinforcing a narrow focus on financial performance and encouraging behaviour aimed at "looking good" rather than supporting a balanced view.

4.5 Flexibility

Flexibility in Hotelia's control system emerges in two distinct ways. First, the divisions adapt their use of Hotelia's control system to fit the different operational realities, resulting in divergent benchmarking practices across Core Services and Signature Services. Second, BPMs exercise discretion in adjusting comp sets when comparisons no longer reflect local market conditions. The following sections examine how these two forms of flexibility are enacted in practice and how they shape whether the integrated system is experienced as enabling or coercive.

The way in which BPMs use benchmarking and select which BSC measures to focus on takes different forms depending on which of Hotelia's divisions is examined. Much of this variation reflects differences in business models, physical layouts, and service concepts. Looking at the Core Services division, benchmarking follows a relatively standardized procedure which makes the employees experience direct comparisons meaningful. This is shown in Kalmar stating "we [Core Services] benchmark our hotels' BSC measures with other [internal] hotels and competitors almost the same way every time". She added that the reason for this is because the hotels operate so similarly making it "straightforward to compare them". This uniformity means that BPMs in Core Services feel little need to adapt or modify the measures or procedures, instead they can rely on a consistent benchmarking approach that applies similarly across hotels.

Looking at the Signature Services division instead, Stockholm expresses the difficulties in using benchmarking the same way across all hotels:

> "Our hotels are exceptionally hard to benchmark. Rooms are different sizes, restaurants have entirely different concepts and public spaces are very different. So, I can't really say straight up, 'You at Hotel X should have performed the same as Hotel Y on this ratio.'"

Given the difficulty of comparing the Signature Services hotels one-to-one, as is done in Core Service, Stockholm's division instead relies primarily on trend-based benchmarking. Rather than comparing a hotel to a small group of very different peers, they look at broader geographical trends to capture general tendencies in the market. As Sundsvall explains it, "we look at the

bigger picture, if the whole Norwegian portfolio is improving over time, but one hotel is moving in the opposite direction, then that tells us something important [...]".

So, building on the need to adapt the benchmarking practice, Hotelia's control system enables such flexibility by letting the BPM's in the respective divisions use benchmarking practices as they deem fit their specific situation the best.

As part of these different ways of using benchmarking in relation to the BSC, it does not appear that these adjustments are developed in dialogue with the hotel managers. When asked whether they were part of this process, Lund noted: "We have no real influence over who ends up in our comp set or how the comparisons are made. [...] it's all decided elsewhere", further expressing that this "sometimes feels unfair."

This reveals a tension in the development process. The adjustments to the comp sets relied mainly on the BPMs' own knowledge and prior experience, with limited incorporation of the hotel managers' local knowledge (Wouters and Wilderom, 2008). From the BPMs' perspective, this is experienced as an enabling outcome of flexibility for control purposes. Yet from the hotel managers' point of view, the same adjustments take place outside the kind of foreground dialogue that helps subordinate managers contextualise and influence performance measures in a way that is experienced as enabling (Cäker et al., 2022). Without access to these discussions, the hotel managers experience the benchmarking adjustments as more coercive, useful for BPMs' control purposes but less meaningful for those being evaluated.

Taken together, these patterns show that the flexibility built into Hotelia's benchmarking and BSC system is experienced very differently across hierarchical levels, producing both enabling and coercive effects. For BPMs, the ability to tailor benchmarking, whether through standardized comparisons in Core Services or trend-based approaches in Signature Services, seem to create autonomy that supports meaningful interpretation of performance and thus feels enabling. Yet this flexibility is not shared with hotel managers. Because comp set adjustments and benchmarking adaptations seem to occur without dialogue or the inclusion of local knowledge,

hotel managers are left to follow decisions they neither influence nor fully understand, making the system feel more coercive than enabling.

5. Discussion

This study set out to explore in what ways benchmarking affects whether the BSC functions as an enabling or coercive control tool through a single-case study of a large hotel chain. By examining how BPMs and hotel managers at Hotelia work with the BSC in combination with benchmarking, we mainly contribute to the literature on management control practices used in combination, and in particular to the recent work of Huber et al. (2025). Their study shows that when the BSC is combined with another control practice, in their case ERM, the integration produces positive effects by offering complementary features that enhance managerial decision-making. Specifically, the integration strengthened the organisation's BSC by adding a forward-looking, risk-oriented perspective that the original scorecard had lacked. In their study, Huber et al. (2025) call for future research to examine other integrations of management control practices to investigate empirically how such integration unfolds in practice.

Our study answers this call. Whereas Huber et al. (2025) found that integrating the BSC with another control tool produced positive complementary effects, our findings reveal a different dynamic. At Hotelia, the combination of the BSC and benchmarking did not strengthen the BSC's intended balance across performance areas. Across the four features of enabling control (Adler and Borys, 1996; Ahrens and Chapman, 2004), our findings instead show that the integration of the BSC and benchmarking at Hotelia resulted mainly in coercive rather than enabling outcomes. Internal transparency was skewed because benchmarking made financial indicators easier to interpret than non-financial ones, reinforcing a narrow financial focus. Global transparency gave everyone visibility into each unit's results, but this often led to competitive "leaderboard" behaviour rather than shared learning. The focus on externally benchmarked financial targets also pushed managers toward short-term financial wins over longer-term initiatives. Flexibility was formally allowed, but in practice it was mainly BPMs, rather than BPMs and hotel managers together, who adjusted the measures. As a result, it did not create the two-way improvement in work practices associated with enabling use (Cäker et al., 2022). Repair

showed more enabling elements, as BPMs refined comp sets and clarified measures to better fit local contexts, though this also carried risks of "gamification". Overall, while dimensions of usability and enabling elements existed, the integrated BSC-benchmarking system was predominantly coercive, with a short-term and financial focus consequently limiting the balanced performance view the BSC intends to create. This

Altogether, while Huber et al. (2025) show that integrating the BSC with ERM can strengthen the scorecard's intended balance and forward-looking orientation, our study demonstrates a different outcome when the BSC is paired with benchmarking. This contrast constitutes our main contribution, showing that the effects of control-system combinations are contingent, and that integration can just as readily weaken the BSC's intended purpose as reinforce it.

As part of contributing to research on combined management control systems, the BSC-benchmarking has to our knowledge not been previously explored through a single-case study. By addressing this gap, our analysis extends understanding of how the two systems function together and sheds new light on the conceptual tensions outlined under section 2.1. Drawing on Hotelia's case, we now revisit these tensions by analyzing how the interplay between the two practices plays out in practice and its implications for the two streams of literature.

5.1 Does the Scorecard Still Balance When Used With Benchmarking?

As outlined under section 2.1, one tension in the BSC literature concerns whether the framework succeeds in balancing financial and non-financial indicators (Cardinaels and Van Veen-Dirks, 2010, Kraus and Lind, 2010; Lipe and Salterio, 2000; Malina and Selto, 2001; 2004). Our study contributes to this stream of research by showing that integrating the BSC with benchmarking made Hotelia's system more coercive, reinforcing a shift toward financial measures.

Benchmarking offered clear, comparable, and externally validated signals for the financial measures, making these indicators the most interpretable and actionable part of Hotelia's system. The non-financial indicators, by contrast, lacked transparency, could not be benchmarked externally, and were widely viewed as ambiguous or even "fictional". This asymmetry meant that when managers discussed performance or attempted to diagnose problems, the financial measures

consistently became the default reference point (Kraus and Lind, 2010). This imbalance was left uncorrected since the repair work largely focused on breaking down and refining financial measures and adjusting comp sets to improve financial comparisons. These practices enhanced the usability of the financial indicators but left the non-financial ones unrepaired, and therefore reinforcing the imbalance already created by the lack of internal transparency (Malina and Selto, 2001). Our findings thus contribute by revealing that combining the BSC with benchmarking amplifies financial dominance and undermines the intended balance of the scorecard.

This imbalance also had implications for the causal relationships, another stream of BSC literature (Humphreys et al., 2016; Ittner et al., 2003; Kraus and Lind, 2010; Malmi, 2001; Nørreklit, 2000; 2003). As benchmarking made the financial measures the most credible and actionable part of the system, the assumed causal logic of Hotelia's BSC was weakened in practice. We found that top management responded to external financial benchmarks, such as competitors' EBITDA levels. This made the financial organizational goals more globally visible than the non-financial targets. This trickled down in the organization and removed focus from the non-financial measures that are supposed to drive long-term financial performance in the BSC (Ittner et al., 2003). However, one hotel manager at Hotelia did understand the causal relationships between measures, which was reflected in the recognition of how local culture shaped long-term performance, but were stalled in acting on this due to top management's priorities (Kraus and Lind, 2010; Malmi, 2001). This highlights how benchmarking, reinforced by top-down financial expectations, can crowd out causal reasoning and limit managers' ability to act on the BSC's strategic drivers.

Additionally, we also contribute to the benchmarking stream of research involving peer selection (Knechel et al., 2010; Chua et al., 2022). More specifically, we do this in terms of answering Chua et al. (2022) call for research on how "peer selection unfolds when benchmarking is coercively mandated throughout an organization, particularly when benchmarking outcomes become a basis for […] the exercise of accountability" (p.245). In Hotelia, the benchmarking tools enabled both internal and external comparisons, and the flexibility for BPMs to decide which hotels should be included in their comp sets, gave them considerable discretion over how performance was represented.

Our empirics show that this discretion was exercised not only when setting comp sets but also through ongoing repair practices, where BPMs revised peer groups to (according to them) better reflect reality. While these adjustments helped BPMs tailor the system to their needs, they largely excluded hotel managers, whose local knowledge did not influence how peers were chosen. This asymmetry meant that peer selection became a continual managerial practice shaped by BPMs' interpretations of comparability, sometimes correcting distortions, but at other times overlooking the comp sets that hotel managers viewed as most appropriate for their local context. Because benchmarking outcomes fed directly into daily reports and performance meetings, these peer choices had consequences for evaluation and accountability.

6. Conclusion

We set out to explore in what ways benchmarking affects whether the BSC functions as an enabling or coercive control tool. What we found is that, in Hotelia, benchmarking pushed the BSC toward more coercive use, moving it away from its promised balance. Across Hotelia, benchmarking consistently amplified financial focus, weakened the causal reasoning, and shaped a system in which comparative pressures and externally anchored targets overshadowed the BSC's non-financial and long-term ambitions.

In doing so, our study mainly contributes to the literature on management control practices used in combination. Whereas Huber et al. (2025) show that integrating the BSC with ERM can strengthen its strategic balance, our findings demonstrate that combining the BSC with benchmarking does not necessarily yield complementary benefits. This highlights that the outcomes of control-system combinations are contingent: integration can weaken, rather than reinforce, the BSC's intended purpose.

Given that the specific interplay between the BSC and benchmarking has, to our knowledge, not previously been examined through a single-case study, our findings offer new perspectives on both the BSC and benchmarking literatures. First, within the BSC stream, we show that benchmarking reinforced financial dominance and weakened the causal reasoning associated with the BSC. Second, in the benchmarking literature, we respond to Chua et al.'s (2022) call by

illustrating how peer selection unfolds when benchmarking is embedded in everyday work. Our study found that BPMs experienced peer selection as an enabling practice, as they could adjust comparison groups to better represent performance. However, hotel managers experienced the same process as coercive because they were excluded from shaping comp sets, leaving their local knowledge unaccounted for.

6.1 Limitations and Further Research

This study is not without limitations. First, the empirical material was collected within a limited time frame. Although several interviewees described how Hotelia's control system had evolved, we were not able to observe these developments directly. As a result, our analysis captures managers' retrospective interpretations rather than the unfolding day-to-day dynamics through which enabling or coercive tendencies might shift over time. A longitudinal study could therefore provide deeper insight into whether the coercive patterns we identified persist, increase, or decrease as organisational priorities and BSC-benchmarking practices evolve.

Our findings rely solely on interviews, which illuminate how managers interpret the integrated BSC-benchmarking system but cannot fully capture the informal negotiations that occur during workshops, performance meetings, and everyday interactions. Several interviewees hinted at sensitive issues such as comp set adjustments made to "improve the numbers", that may be better understood through direct observation. On-site access to monthly meetings or workshops could therefore reveal how performance information is emphasised in social dynamics, and how this shapes the emergence of enabling or coercive use.

Moreover, we were not granted access to senior executives, even though interviewees frequently referenced their influence on target setting, financial prioritisation, and the strategic interpretation of benchmarking signals. Access to these actors could offer valuable complementary insights, especially regarding how benchmarking is used at the corporate level with the BSC and how such usage trickles down into divisional and hotel-level practices.

Future research could extend our work in several ways. Firstly, one direction is to examine whether the dynamics identified at Hotelia are specific to hotel chains, an industry found to have

unusually rich access to external benchmarking, or whether similar patterns emerge where external comparators are less readily available. Similar studies in other sectors could therefore expand on whether benchmarking consistently pushes the BSC toward financial dominance, or whether its effects depend on the informational environment in which benchmarking is embedded. Secondly, further research could study peer-selection practices directly, for instance by following how comp sets are constructed and revised over time, and investigating how different hierarchical levels negotiate (or fail to negotiate) these decisions. Lastly, further research could investigate the BSC alongside other control practices to assess whether alternative combinations may produce enabling effects.

7. References

Adler, P. S., and Borys, B. (1996). Two Types of Bureaucracy: Enabling and Coercive. *Administrative Science Quarterly*, 41, 61-89.

Ahrens, T. and Chapman, C. (2004). Accounting for Flexibility and Efficiency: A Field Study of Management Control Systems in a Restaurant Chain. *Contemporary Accounting Research.* 21. 271-301.

Ahrens, T., and Chapman, C. (2006). Doing qualitative field research in management accounting: Positioning data to contribute to theory. *Accounting, Organizations and Society,* 31(8), 819-841.

Atkinson, R., and Flint, J. (2001). Accessing hidden and hard-to-reach populations: Snowball research strategies. *Social Research Update, 33*, 1-4.

Banker, R. D., Chang, H., and Pizzini, M. J. (2004). The balanced scorecard: Judgmental effects of performance measures linked to strategy. *The Accounting Review, 79*(1), 1-23.

Bedford, D. S., Malmi, T., and Sandelin, M. (2016). Management control effectiveness and strategy: An empirical analysis of packages and systems. *Accounting, Organizations and Society, 51*, 12-28.

Braun, V., and Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77-101.

Cardinaels, E., and van Veen-Dirks, P. M. G. (2010). Financial versus non-financial information: The impact of information organization and presentation in a balanced scorecard. *Accounting, Organizations and Society, 35*(6), 565-578.

Chua, W. F., Graaf, J., and Kraus, K. (2022). Mapping and contesting peer selection in digitalized public sector benchmarking. *Financial Accountability & Management, 38*(2), 223-251.

Cäker, M., Siverbo, S., and Åkesson, J. (2022). Performance measurement systems, hierarchical accountability and enabling control. *Accounting and Business Research, 52*(7), 865-889.

Dubois, A., and Gadde, L.-E. (2002). Systematic combining: An abductive approach to case research. *Journal of Business Research, 55*(7), 553-560.

Eisenhardt, K. M. (1989). Building theories from case study research. *Academy of Management Review*, 14(4), 532-550.

Eisenhardt, K. M., and Graebner, M. E. (2007). Theory building from cases: Opportunities and challenges. *Academy of Management Journal, 50*(1), 25-32.

Elnathan, D., and Kim, O. (1995). Partner selection and group formation in cooperative benchmarking. *Journal of Accounting and Economics, 19*(2-3), 345-364.

Hoque, Z. (2014). Twenty years of studies on the balanced scorecard: Trends, accomplishments, gaps and opportunities for future research. *The British Accounting Review, 46*(1), 33-59.

Huber, C., Kraus, K., and Meidell, A. (2025). Integrating the balanced scorecard and enterprise risk management: Exploring the dynamics between management control anchor practices and subsidiary practices. *Management Accounting Research*, 66, 100924.

Humphreys, K. A., Gary, M. S., and Trotman, K. T. (2016). Dynamic decision making using the balanced scorecard framework. *The Accounting Review, 91*(5), 144-1465.

Humphreys, K. A., and Trotman, K. T. (2011). The balanced scorecard: The effect of strategy information on performance evaluation judgments. *Journal of Management Accounting Research, 23*(1), 81-98.

Ittner, C. D., Larcker, D. F., and Meyer, M. W. (2003). Subjectivity and the weighting of performance measures: Evidence from a balanced scorecard. *The Accounting Review*, 78(3), 725-758.

Jordan, S., and Messner, M. (2012). Enabling control and the problem of incomplete performance indicators. *Accounting, Organizations and Society, 37*(8), 544-564.

Kaplan, R.S. and Norton, D.P. (1992) 'The Balanced Scorecard - Measures That Drive Performance' Harvard Business Review 70(1), pp. 71-79.

Kaplan, R. S., and Norton, D. P. (1993). Putting the balanced scorecard to work. *Harvard Business Review*, 71(5), 134-147.

Kaplan, R. S., and Norton, D. P. (1996). The balanced scorecard: Translating strategy into action. *Harvard Business School Press*.

Kaplan, R. S., and Norton, D. P. (2000). The strategy-focused organization: How balanced scorecard companies thrive in the new business environment. *Harvard Business School Press*.

Kaplan, R. S., and Norton, D. P. (2004). The strategy map: Guide to aligning intangible assets. *Strategy & Leadership*, 32(5), 10-17.

Knechel, W. R., Salterio, S. E., and Kochetova-Kozloski, N. (2010). The effect of benchmarked performance measures and strategic analysis on auditors' risk assessments and mental models. *Accounting, Organizations and Society*, 35(3), 316-333.

Kraus, K., and Lind, J. (2010). The impact of the corporate balanced scorecard on corporate control: A research note. *Management Accounting Research*, 21(4), 265-277.

Lipe, M. G., and Salterio, S. E. (2000). The balanced scorecard: Judgmental effects of common and unique performance measures. *The Accounting Review*, 75(3), 283-298.

Malina, M. A., and Selto, F. H. (2001). Communicating and controlling strategy: An empirical study of the effectiveness of the balanced scorecard. *Journal of Management Accounting Research*, 13(1), 47-90.

Malina, M. A., and Selto, F. H. (2004). Choice and change of measures in performance measurement models. *Management Accounting Research*, 15(4), 441-469.

Malmi, T. (2001). Balanced scorecards in Finnish companies: A research note. *Management Accounting Research*, 12(2), 207-220.

Malmi, T., and Brown, D. A. (2008). Management control systems as a package - Opportunities, challenges and research directions. *Management Accounting Research*, 19(4), 287-300.

Merchant, K. A., and Otley, D. (2020). Beyond the systems versus package debate. *Accounting, Organizations and Society*, 86, 101185.

Mundy, J., Levant, Y., de La Villarmois, O., Naro, G., and Travaillé, D. (2011). The role of the balanced scorecard in the formulation and control of strategic processes. *Journal of Applied Accounting Research*, 12(3), 212-233.

Northcott, D., and Llewellyn, S. (2003). The 'ladder of success' in healthcare: The UK national reference costing index. *Management Accounting Research*, 14(1), 51-66.

Nørreklit, H. (2000). The balance on the balanced scorecard: A critical analysis of some of its assumptions. *Management Accounting Research*, 11(1), 65-88.

Nørreklit, H. (2003). The balanced scorecard: What is the score? A rhetorical analysis. *Accounting, Organizations and Society*, 28(6), 591-619.

Qu, S. Q., and Dumay, J. (2011). The qualitative research interview. *Qualitative Research in Accounting & Management*, 8(3), 238-264.

Vaivio, J. (2008). Qualitative management accounting research: Rationale, pitfalls and potential. *Qualitative Research in Accounting & Management*, 5(1), 64-86.

Wouters, M., and Wilderom, C. (2008). Developing performance-measurement systems as enabling formalization: A longitudinal field study of a logistics department. *Accounting, Organizations and Society*, 33(4-5), 488-516.

8. Appendix

Appendix 1. Interview Guide Templates

This structure served as the basis for our interviews and was tailored prior to each to suit the interviewee.

Interview guide template for the first interview:

- Introduction of ourselves and our research project
- Background and role
    - Could you tell us a bit about your background and your role at Hotelia?
    - Which aspects of your work do you see as most central for managing performance in the organisation?
- Performance Measurement
    - How are performance targets generally set here? Who is involved, and how much influence do you have in the process?
    - Which systems, routines, or tools do you rely on when assessing performance?
- Tensions and Challenges
    - Can you recall situations where managing performance felt particularly challenging?
    - Are there aspects of the current control practices that you find especially helpful (or constraining) in your work?
- Ending
    - Is there anything we haven't asked that you think is important for us to know about your work?

Interview guide template for the rest of the interviews:

- Introduction of ourselves and our research project
- Background and role
    - Could you tell us a bit about your background and your role at Hotelia?
- The BSC
    - How would you describe the process of setting and updating KPIs in the BSC?

    - To what extent can measures be adapted to local hotel conditions? What does such a change require in practice?
    - Have there been situations where measures needed to be adjusted, clarified, or replaced? How was that handled?
    - In your experience, does the BSC function more as support for understanding performance or as a tool for ensuring compliance?
- Benchmarking
    - How is benchmarking used in your daily work?
    - What kinds of discussions typically arise when benchmarking results are presented?
    - Which types of measures become most visible in benchmarking reports, and how does this influence what receives attention?
- The control system as a whole
    - When the BSC and benchmarking are used together, how do they shape managers' focus?
    - Does benchmarking help you understand performance across hotels, or does it mainly drive comparison?
- Ending
    - Looking ahead, what aspects of Hotelia's control system would you most like to develop or improve?
    - What is something related to your work that keeps you up at night?

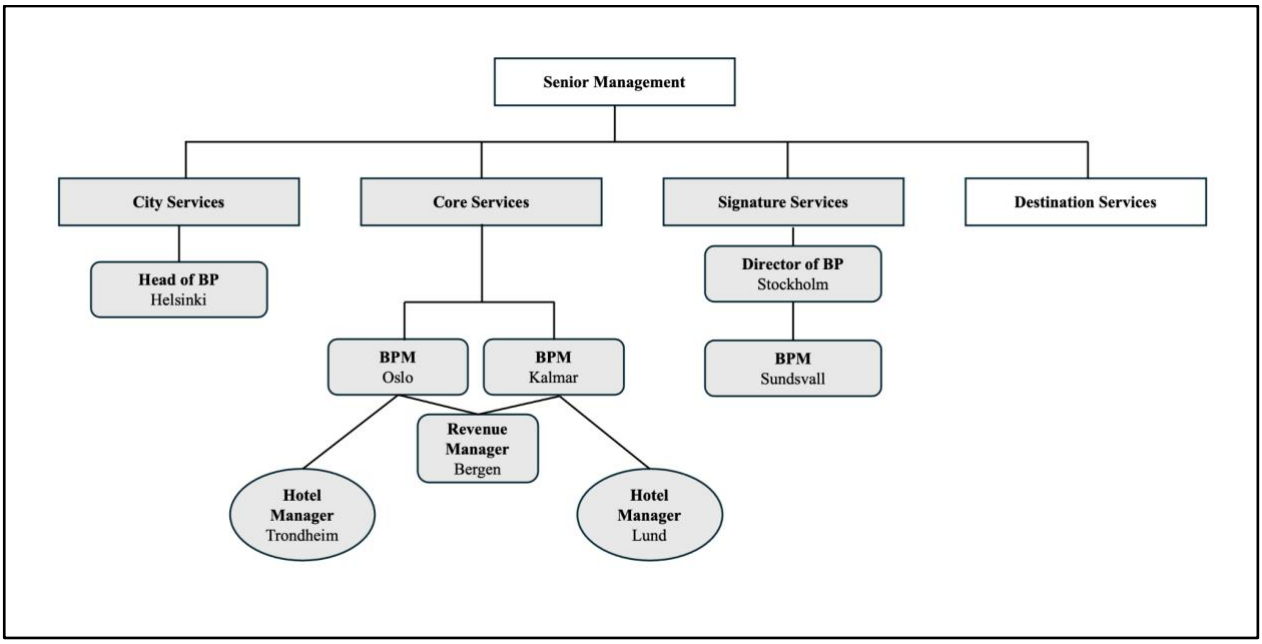
Appendix 2. Interview Table

Table overview of interviewees'.

Title	Pseudonym	Division	Date and Duration
Head of Business Performance	Helsinki	City Services	2025-09-12 41 minutes
Head of Business Performance	Helsinki	City Services	2025-10-14 58 minutes
BPM	Oslo	Core Services	2025-10-23 42 minutes
Director of Business Performance	Stockholm	Signature Services	2025-10-30 53 minutes
Revenue Manager	Bergen	Signature Services	2025-10-31 57 minutes
BPM	Sundsvall	Signature Services	2025-11-05 42 minutes
Hotel Manager	Trondheim	Hotel within Core Services	2025-11-07 41 minutes
BPM	Kalmar	Core Services	2025-11-12 57 minutes
Hotel Manager	Lund	Hotel within Core Services	2025-11-22 40 minutes

Appendix 3. Organizational Chart

Grey-shaded areas indicate the parts of the organization our study mainly focuses on.



8.1 Use of Generative AI

In line with SSE's guidelines, our use of generative AI was limited and carefully controlled. We occasionally consulted OpenAI's ChatGPT (versions 5/5.1, https://chatgpt.com) as a tool for inspiration, particularly when exploring alternative phrasing, improving clarity, or refining the flow of our writing. In later stages, we also used AI sparingly to check grammar and overall readability.

Importantly, no AI-generated text was included without thorough critical review and modification. All suggestions were assessed and reworked to ensure that the thesis reflects our own reasoning, analytical work, and academic judgement. AI tools were not used for empirical analysis, coding, or interpretation at any point in the research process.

Overview of occasions where generative AI (ChatGPT 5/5.1) was consulted.

Primarily for brainstorming (15 September 2025)
Primarily for improving clarity (18 September 2025)
Primarily for alternative wording (23 September 2025)
Primarily for improving clarity (27 September 2025)
Primarily for alternative wording (3 October 2025)
Primarily for improving clarity (7 October 2025)
Primarily for alternative wording (14 October 2025)
Primarily for improving clarity (21 October 2025)
Primarily for refining flow (27 October 2025)
Primarily for alternative wording (3 November 2025)
Primarily for improving clarity (10 November 2025)
Primarily for alternative wording (20 November 2025)
Primarily for improving clarity (25 November 2025)
Primarily for refining flow (29 November 2025)
Primarily for alternative wording (2 December 2025)
Primarily for improving clarity (7 December 2025)