Success only in the Numbers? A Critical Look at Outcome Metrics in Social Impact Bonds

Tim Hölting

Dissertation written under the supervision of Dr. Reinhard Millner

Dissertation submitted in partial fulfilment of requirements for the MSc in International Management with a specialization in Strategy & Consulting, at Universidade Católica Portuguesa and for the MSc in Strategy, Innovation and Management Control at WU University of Economics and Business, 14th of August 2025.

Abstract

My thesis critically examines the definition and measurement of contractual outcome metrics (COMs) and payment structures in Social Impact Bonds (SIBs). I also investigate the relationship between these factors and SIBs' success. Additionally, I consider the importance of non-contractual factors, which potentially impact success. Drawing on a mixed-methods approach, I combine a quantitative database analysis of completed SIBs in the UK with a qualitative content analysis of their evaluation reports. Agency theory is used to explore how SIBs aim to align stakeholder motivations.

I find that all investigated SIBs employed payment-per-outcome models, but COMs were often unrealistic, leading to mid-project adjustments and financial strain on providers. Furthermore, while quantitative targets were not always met, stakeholders frequently perceived qualitative outcomes as significant, though payment-relevant metrics rarely captured such outcomes. My research also identifies key structural challenges, including perverse incentives, mission drift, and persistent agency problems arising from diverging stakeholder interests. Finally, informal aspects like adaptability, stakeholder collaboration, and trust are equally key to success.

I argue for more realistic and context-sensitive COMs that reflect both measurable quantitative outcomes and broader social value. Recommendations include designing well-defined but flexible SIBs that safeguard against unintended consequences and using more robust evaluation methods to attribute results to interventions.

My thesis contributes to the research on SIBs by advancing the understanding of them from an agency theoretical perspective and highlighting differences or commonalities between the literature and practice. I also offer practical guidance for practitioners in designing more effective and equitable outcome-based financing mechanisms.

Keywords: Social Impact Bonds, Contractual outcome metrics, Payment-per-outcome, Impact measurement, Financial success, Qualitative outcome, Perverse incentives, Agency theory, Collaboration, Adaptability

Title: Success only in the Numbers? A Critical Look at Outcome Metrics in Social Impact Bonds

Author: Tim Hölting

Resumo

A minha tese examina a definição e medição de métricas de resultados contratuais (COMs) e estruturas de pagamento em Títulos de Impacto Social (SIBs). Também investigo a relação entre esses fatores e o sucesso dos SIBs, além de considerar fatores não contratuais. Utilizando uma abordagem de métodos mistos, combino uma análise de uma base de dados de SIBs concluídos no Reino Unido com uma análise qualitativa dos seus relatórios de avaliação. A teoria da agência é utilizada para explorar como os SIBs visam alinhar os interesses das partes interessadas.

Constato que os COMs eram frequentemente irrealistas, levando a ajustes necessários e pressão financeira sobre os prestadores. Além disso, as partes interessadas frequentemente consideravam os resultados qualitativos como significativos, embora as métricas relevantes para o pagamento captassem principalmente apenas resultados quantitativos. A minha investigação também identifica desafios, incluindo incentivos perversos, desvio da missão e problemas de agência. Por fim, aspetos informais, como a adaptabilidade e a colaboração das partes interessadas, são fundamentais para o sucesso.

Defendo COMs mais realistas, que reflitam tanto os resultados quantitativos como o valor social mais amplo. As recomendações incluem a conceção de SIBs bem definidos, mas flexíveis, e a utilização de métodos de avaliação que permitam atribuir os resultados às intervenções.

Contribuo para a literatura ao aprofundar a compreensão dos SIBs a partir de uma perspetiva teórica da agência e ao destacar as diferenças ou semelhanças entre a literatura e a prática. Também ofereço orientações práticas para a conceção de SIBs mais eficazes e equitativos.

Palavras-chave: Títulos de Impacto Social, Métricas de resultados contratuais, Pagamento por resultado, Medição de impacto, Sucesso financeiro, Resultado qualitativo, Incentivos perversos, Teoria da agência, Colaboração, Adaptabilidade

Título: Sucesso apenas nos números? Uma análise crítica das métricas de resultados nos Títulos de Impacto Social

Autor: Tim Hölting

Table of Contents

Abstract i

Resumo ii

List of Figures vi

List of Tables vii

List of Abbreviations viii

1. Introduction 1

2. Impact Investing 3

2.1. The Impact Investing Ecosystem 4

2.2. Actors 5

2.3. Mechanisms 6

2.4. Financial Instruments 7

3. Social Impact Bonds 8

3.1. Development and Current Landscape of Impact Bonds 9

3.2. Key Actors and Their Roles 12

3.2.1. Commissioner 12

3.2.2. Investor 13

3.2.3. Intermediary 13

3.2.4. Social Service Provider 13

3.2.5. Evaluator 14

3.3. SIB Lifecycle 14

3.4. Agency Theory and Social Impact Bonds 15

3.5. Alignment of Interest 16

3.6. Varying Structures of Social Impact Bonds 17

3.6.1. Social Impact Bond Models 17

3.6.2. Payment Structures 18

3.6.3. Other Types of Impact Bonds 20

3.7. Critical Reflection of Social Impact Bonds 20
3.7.1. Advantages 20
3.7.2. Disadvantages 21
4. Impact Measurement 22
4.1. Defining Success in Social Impact Bonds 23
4.2. Designing Appropriate Outcome Metrics 24
4.3. Conceptual Approaches to Impact Measurement 26
4.4. Evaluation Design 28
4.5. Methods and Tools for Impact Assessment 28
4.6. Methodological Issues 30
5. Methodology 32
5.1. Research Design 32
5.2. SIB Database 33
5.2.1. Data Sample 33
5.2.2. Variables 34
5.2.3. Data Collection 35
5.3. Qualitative Content Analysis 37
5.3.1. Evaluation Report Corpus 38
5.3.2. Coding Agenda 39
5.4. Data Analysis 40
5.5. Methodological Limitations 40
6. Results 41
6.1. Findings from the SIB Database 41
6.1.1. Structural Data of Relevant SIBs 41
6.1.2. Insights on Contractual Outcome Metrics 42
6.1.3. SIB Evaluation Methods 46
6.1.4. SIB Success & Achievement of Contractual Outcome Metrics 48

6.2. Findings from the Evaluation Report Analysis 51
6.2.1. SIB Design, Construction & Delivery 51
6.2.2. Measurement & Evaluation 58
6.2.3. Perception of Success 61
6.2.4. Design Recommendations 70
7. Discussion 72
7.1. Definition and Measurement of Contractual Outcome Metrics 73
7.2. SIB Design, Impact Measurement, and Success 75
7.3. Unintended Consequences and Perverse Incentives 79
7.4. Integration with Agency Theory 80
8. Conclusion 82
8.1. Main Findings 82
8.2. Implications for Theory and Practice 84
8.3. Limitations and Future Research 86
9. References 89
10. Appendix 100
Appendix A: Variable overview from the SIB database 100
Appendix B: Full List of COM Grouping 103
Appendix C: Full List of Validation Method Grouping 106
Appendix D: List of evaluation reports used for the content analysis 107
Appendix E: Content Analysis Codebook 109
Appendix F: Structural Data of Relevant SIBs 114
Appendix G: COM Subgroups and COM Achievement 118

List of Figures

Figure 1: The Impact Investing Ecosystem Spectrum. 4
Figure 2: Different Types of Impact Investing..... 8
Figure 3: Number of Impact Bonds by Stage of Development..... 10
Figure 4: Number of Impact Bonds by Start Year. 10
Figure 5: Impact Bonds by Location..... 11
Figure 6: Number of Impact Bonds by Policy Sector. 11
Figure 7: The SIB model..... 12
Figure 8: Logic Model and Theory of Change..... 26
Figure 9: Frequency of Intermediaries. 42
Figure 10: Fund Overview. 42
Figure 11: Overview of the Number of COMs per SIB. 43
Figure 12: Frequency of COM Groups. 44
Figure 13: Frequency of Validation/Verification Methods..... 46
Figure 14: Overview of Qualitative Evaluation Methodologies. 46
Figure 15: Overview of Quantitative Evaluation Methodologies. 47
Figure 16: SIB Success Overview..... 48
Figure 17: Overview of COM Achievement..... 48
Figure 18: COM Groups and COM Achievement. 49
Figure 19: COM Payment and Achievement. 50
Figure 20: COM Achievement per Subgroup. 51

List of Tables

Table 1: Example of COM Grouping. 36
Table 2: Overview of the Codes for the Content Analysis. 39
Table 3: Exemplary Coding Agenda. 40
Table 4: Overview of COM Payments. 45

List of Abbreviations

ABLE	Adolescent Behavioral Learning Experience
BIT	Behavioural Insights Team
CBO	Commissioning Better Outcomes Fund
COM	Contractual Outcome Metric
CR	Counterfactual Risk
CSR	Corporate Social Responsibility
DIB	Development Impact Bond
DWP	Department of Work and Pensions
EIB	Environmental Impact Bond
ESG	Environmental, Social, and Governance
GBP	Great British Pound
GIIN	Global Impact Investing Network
IF	Innovation Fund
IPS	Individual Placement Support
MHEP	Mental Health and Employment Partnership
NEET	Not in Education, Employment or Training
NPO	Nonprofit Organization
NQF	National Qualifications Framework
NVQ	National Vocational Level
PbR	Payment by Results
PPP	Public-Private Partnership
PSM	Propensity Score Matching
QCF	Qualifications and Credit Framework
RCT	Randomized Controlled Trial
SIB	Social Impact Bond
SOF	Social Outcomes Fund
SPV	Special Purpose Vehicle
SRI	Socially Responsible Investment
SROI	Social Return on Investment
SSP	Social Service Provider
VCSE	Voluntary, Community, and Social Enterprise
YEF	Youth Engagement Fund

1. Introduction

Social Impact Bonds (SIBs) are a relatively new financial mechanism in impact investing for the delivery of social services (Fraser, Tan, Kruithof, et al., 2018). They were first introduced in the UK in 2010, but have gained significant attention among practitioners and researchers. The number of SIBs has been growing massively over the last 15 years. The INDIGO impact bond database (2025), a public database on impact bonds operated by the Oxford Government Outcomes Lab, currently lists 319 impact bonds, including SIBs and Development Impact Bonds (DIBs), with a total capital raised of more than $867 million. On the academic side, this gained momentum for SIBs, also led to an increasing amount of literature being published on the topic (cf. Broccardo et al., 2020; Carè et al., 2023; Chiappini et al., 2023; Dahbi et al., 2024; Fraser, Tan, Lagarde & Mays, 2018).

SIBs are often introduced as a 'win-win' instrument for the parties involved (Berndt & Wirth, 2018; Fraser, Tan, Lagarde & Mays, 2018). However, widespread use is yet outstanding, and the current peak of implementations of social projects happened in 2017 (Dahbi et al., 2024), which also highlights doubts and challenges that SIBs face (Fraser, Tan, Lagarde & Mays, 2018; Lake, 2015; Sinclair et al., 2014). The relatively small SIB market also demonstrates the missing diffusion compared to the overall impact investing market. Hand et al. (2024) estimate the global impact investing Assets Under Management to be around $1.57 trillion, meaning that the current SIB market size of a little bit more than $867 million is less than 1% of the impact investing market size.

Impact measurement and success verification are key challenges SIBs face, and the current literature does not adequately address this issue. However, for a growing acceptance and increased attractiveness of SIBs, it is imperative to have reliable impact measures to assess the success of the SIBs to gain the trust and confidence of the stakeholders involved. To develop improved measurement systems, it is necessary to understand the currently used contractual outcome metrics (COMs) and evaluation methods (Dahbi et al., 2024). At the same time, it would be helpful to look into reasons for successful or unsuccessful SIBs by comparing their contractual forms (Chiappini et al., 2023).

From a theoretical perspective, no studies have done this so far. Existing papers investigating specific SIBs have focused on individual SIBs (e.g., Millner & Meyer, 2022) or small sub-samples (e.g., Rizzello & Kabli, 2020). Therefore, the current literature is missing a broader overview of the used COMs, payment structures, measurements, and their connection to the

success of the SIBs. Success in this thesis is seen from two sides. First, there is the achievement of the COMs and the connected financial payments, i.e., quantitative outcomes. Second, there are qualitative outcomes, like improved self-confidence, that are often not reflected in the COMs.

From a practical perspective, it is important to understand how to optimally design SIBs, including robust metrics that achieve desired outcomes while limiting the risk of perverse incentives. Even though the number of SIBs is constantly growing (INDIGO, 2025), there still seems to be a lack of understanding about the important success factors, including the specific COMs to be used. Additionally, having more robust metrics and measurements could increase the attractiveness of SIBs among investors and thus could lead to stronger future SIB market growth. Furthermore, understanding why SIBs have been successful or failed in the past allows refinement of new and existing SIB contracts.

Therefore, this thesis aims to address the knowledge gap on COMs, payment structures, measurement, and success by looking at key contractual elements of SIBs and providing a holistic overview of the used elements based on available data from already completed SIBs from the employment and training policy area in the UK. For this, the available data from the INDIGO Impact Bond Dataset is used. Additionally, a qualitative content analysis with publicly available evaluation reports from the sampled SIBs was conducted. The reports are a mix of qualitative and quantitative reports and, except for one intermediary process evaluation, assess the overall performance of the SIBs after the completion or towards the end of the intervention. To prevent a too narrow view on potentially influential factors for the performance of SIBs, I also consider other factors like stakeholder collaboration and perverse incentives of SIB designs in my qualitative content analysis. The results of the thesis shall inform the design of future SIBs. This is because a better understanding of successful SIB designs might be able to reduce transaction costs before the implementation, implementation costs, e.g., due to extensive discussion between the different stakeholders, and measurement costs of the intervention. The thesis contributes to the literature on SIBs and impact measurement, an under-investigated area (Dahbi et al., 2024). It also adds to the growing number of empirical studies on SIBs where theoretical contributions dominate yet (Carè et al., 2023). The insights from the study can also inform the overarching impact investing area on best practices for successful contract specifications, especially related to the definition and measurement of COMs.

Specifically, the research questions for the thesis are the following:

RQ1. How are contractual outcome metrics defined in Social Impact Bonds, and how are they measured?

RQ2. How is success related to the contractual design of Social Impact Bonds and impact measurement?

The thesis also contributes to the literature by analyzing the unintended effects caused by the SIB designs and experienced during the real-life interventions. Exemplary risks mentioned in the literature are the potential for perverse incentives (Sinclair et al., 2014) or the subordination of financial interests over social outcomes (Lake, 2015).

Furthermore, the thesis provides empirical insights into the agency problems within SIBs and how they are aimed to be resolved. I demonstrate that SIBs can generally align the interests of various stakeholders with diverging motivations, but also show that challenges exist.

The remainder of the thesis is structured as follows: Chapter 2 introduces the broader area of impact investing, of which SIBs are a subsection. Chapter 3 defines and introduces agency theory and SIBs, including a development since their introduction, the key stakeholders, payment structures, and types of SIBs, as well as a critical reflection present in the debate in the current literature. Chapter 4 dives deeper into COMs, impact measurement, and evaluation, reflecting the current literature on these topics. Chapters 5, 6, 7, and 8 explain the methodology for the empirical part, present and discuss the results, and conclude by summarizing the key findings and highlighting the implications of the thesis.

2. Impact Investing

The Global Impact Investing Network (GIIN) defines impact investing as "investments made with the intention to generate positive, measurable social or environmental impact alongside a financial return" (GIIN, 2025a). This approach highlights the dual objective of positively impacting society or the environment while generating financial value (Chiappini et al., 2023; Evans, 2013; Spiess-Knafl & Scheck, 2023). However, there is debate within the industry about whether this objective is achievable. While some see impact investing as a way to overcome the seemingly contradictory goals, others believe that it is impossible and that there are different types of investors, with some focusing primarily on financial returns and others emphasizing social returns (Evans, 2013; Ormiston et al., 2015). This might lead to reduced returns on either end.

2.1. The Impact Investing Ecosystem

The impact investing ecosystem encompasses all forms of investing that aim to create social or environmental value by addressing pressing societal challenges. Figure 1 provides an overview of the ecosystem. It becomes evident that there are often overlaps between the different types of investing, making it difficult to distinguish them fully.

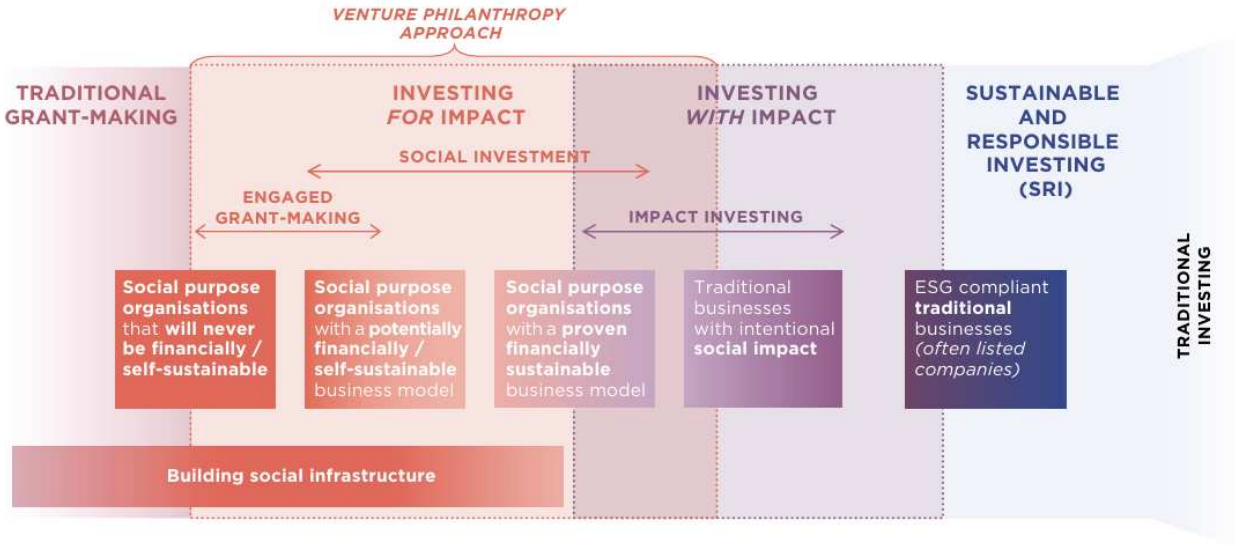


Figure 1: The Impact Investing Ecosystem Spectrum.
Source: Gianoncelli et al. (2019).

The overlaps are especially evident in *Investing for Impact* and *Investing with Impact*. *Investing for Impact* focuses on building social infrastructures and supports testing novel solutions that are not proven and potentially without a self-sustainable business model (Gianoncelli & Gaggiotti, 2021). Consequently, the investors in this part of the spectrum are willing to take higher risks, and social impact is more important than financial returns (Gianoncelli & Boiardi, 2018). *Investing with Impact*, on the other hand, is primarily concerned with financial outcomes, but also aims to achieve social impact simultaneously (Gianoncelli & Boiardi, 2018). The investors are more risk-averse and look for proven solutions they can scale (Gianoncelli & Gaggiotti, 2021).

Concerning other areas of investing, the distinctions are more apparent. While socially responsible investment (SRI) is similar to impact investing, as it also integrates "certain nonfinancial concerns, such as ethical, social or environmental, into the investment process" (Sandberg et al., 2009, p. 521), there are also several differences. Höchstädter and Scheck (2015, p. 456) summarize them as "a greater proactiveness of impact investing to solve social and/or environmental challenges (rather than improving corporate practices in terms of ESG

criteria), differences in the size and nature of investments (small versus large investees, investments in publicly listed companies versus direct investments in the form of private debt or equity), as well as differing returns expectations and risk-return profiles". Ormiston et al. (2015) also note that SRI does not aim for a defined and measurable social impact.

Philanthropy and government funding also aim for positive social impact, but unlike in impact investing, no financial returns are expected, i.e., the money is typically given as grants or donations (Ormiston et al, 2015). Therefore, this form of financing is typically used for organizations that cannot sustain themselves through market income (see Figure 1) and can be complementary to impact investing (Spiess-Knafl & Scheck, 2023).

Lastly, impact investing should also be distinguished from Corporate Social Responsibility (CSR). The latter is not focused on tackling social problems but on reducing negative externalities from companies and creating positive ones (Spiess-Knafl & Scheck, 2023).

Overall, the impact ecosystem consists of several types of investing, which are generally distinct but also overlap and complement each other. Compared to philanthropy, SRI, and CSR, impact investing is unique because it simultaneously focuses on achieving social impact and financial returns.

2.2. Actors

Various important actors are involved in the impact investing ecosystem, which can be broadly categorized into investors, intermediaries, and facilitators. The key actors for SIBs will be introduced later in this thesis; hence, this section will only provide a high-level overview of the broader variety of actors in the impact investing ecosystem.

Investors provide the capital needed for social enterprises to fulfil their mission and achieve the social and financial outcomes. Spiess-Knafl & Scheck (2023) distinguish three general types of investors: (1) Donors or charitable foundations, who do not expect financial returns; (2) investors with reduced financial return expectations, i.e., investors that value social impact more than financial return and would also be willing to sacrifice return for impact; and (3) investors with market-oriented financial return expectations, i.e., investors that do not necessarily care about social or ecological impact in their investment decisions. More specifically, there is a wide range of different investors, including foundations, social impact funds (e.g., social venture capital funds that apply the concept of venture capital in the social sector), banks (especially ethical banks, but also conventional ones as they are increasingly

considering ESG criteria in their investments), crowdfunding platforms (four different types depending on the type of funding: equity, debt, donation, reward-based), corporate social investors (e.g., corporate foundations or corporate social businesses), public sector, institutional investors (e.g., pension funds, insurance companies, or sovereign wealth funds), asset management companies, family offices (i.e., organizations managing private wealth e.g., from high-net-worth individuals), non-governmental organizations, and development finance institutions (i.e., institutions that invest in less developed countries to generate economic development, but also increasingly tackle societal challenges; Gianoncelli et al., 2019; Spiess-Knafl & Scheck, 2023).

Investors can use different financing instruments and be active in more than one area of the ecosystem. An example would be a foundation that gives grants while also using debt or equity instruments in *Investing for Impact* (Gianoncelli et al., 2019).

Network and social investment advisors are intermediaries and facilitators in the impact investing ecosystem. Both types fulfil an important task in the social capital market by bringing together the demand and supply sides. Social investment advisors do what is usually done by an investment bank in a traditional investment setting. The networks connect different actors, such as investors and social enterprises, and aim to enhance collaboration, thereby reducing transaction costs, which are typically high in the social environment (Spiess-Knafl & Scheck, 2023).

2.3. Mechanisms

Spiess-Knafl & Scheck (2023) mention four types of mechanisms in their book chapter on the impact investing market: SRI, outcome-based financing, guarantee schemes, and catalytic structures. Technically, SRIs are different from impact investing, as explained in chapter 2.1. The authors also note this necessary differentiation. Hence, for this thesis, SRIs are not seen as a mechanism within the broader spectrum of impact investing. In outcome-based financing models, payments are, as the name suggests, dependent on the outcomes achieved by the financed programs. Over the years, different models like Payment by Results (PbR), SIB, or DIB have emerged. Chapter 3 will go into more detail on SIBs and the differences between SIBs and the other models. Guarantee schemes are not a financing mechanism per se, but they enable the investee to raise more debt capital. By taking over the risk of capital repayment, the guarantor lowers the financing costs for the organization, thereby increasing the available funds. Catalytic structures encompass public subsidies and grants. Even though these financing

mechanisms are not directly part of impact investing, they complement it. They can act as catalysts for further funding with private capital and therefore leverage the social impact generated by the social enterprise (Spiess-Knafl & Scheck, 2023).

2.4. Financial Instruments

Impact investors utilize various financial instruments to fund social enterprises, depending on various factors, such as the specific situation of the enterprise or the sector they are investing in. Common instruments fall into one of these categories: equity, loan/debt, grant, or hybrid capital. The individual instruments will be briefly explained in the following:

Equity capital means investing in ownership stakes in social enterprises. This gives the investor (partially) control over the company and can lead to agency conflicts if the interests of the investee and the investor are not aligned. The investor is not guaranteed a fixed return on the investment, making it the riskiest form of investment (Spiess-Knafl & Scheck, 2023). Within the impact investing ecosystem, equity capital is seen as rather critical due to the perceived risk of mission drift, which means that social enterprises might shift their focus away from their social mission towards achieving financial returns (Achleitner et al., 2013).

Spiess-Knafl & Scheck (2023) further mention debt capital and grants are the other plain forms of investments. Debt capital is a loan that has to be repaid, typically with various interest payments throughout the loan period. This form of financing requires a more stable business model and predictable cash flows. Unlike debt, grants do not have to be repaid and therefore do not lead to any cost for the social enterprise receiving them. Because of that, they are not part of impact investing. However, they are very popular in the social sector, where many organizations do not have a self-sustaining business model.

Next to these plain investment forms, Spiess-Knafl & Scheck (2023) name many hybrid forms. One is mezzanine capital, which falls between debt and equity and allows the investor to convert debt to equity. Furthermore, there are recoverable grants (i.e., grants to be repaid if the investee achieves certain milestones) and convertible grants (i.e., grants converted into equity capital if the company is successful). Forgivable loans are debt capital that may be forgiven or at least reduced under specific, pre-defined conditions, which can lower the incentive for mission drift. Another instrument is revenue share agreements, which have a flexible repayment structure. The investor provides money upfront and is given a pre-defined share of the company's revenues over some time. This set-up is attractive to social enterprises because the repayment adjusts with the revenues, leading to flexible costs.

Overall, impact investing is embedded in a complex ecosystem of socially oriented investment practices that overlap to a certain extent but also serve different parts of the social sector. Impact investing differentiates itself by aligning measurable social impact with financial returns. There are many different actors in this sector, especially investors, and they all use various mechanisms and financial instruments. The following chapter will dive deeper into SIBs, a specific area within impact investing.

3. Social Impact Bonds

SIBs belong to the broader field of impact investing (see Figure 2) and can be seen as a form of a public-private partnership (PPP), albeit being very different from the well-known infrastructure PPPs in their contractual design (Wang et al., 2018; Warner, 2013). Their main goal is to provide preventive social services for underserved groups with the help of private capital in areas such as homelessness and recidivism, which are underinvested by regular government programs that typically focus on remediating existing problems (Carè et al., 2023; Del Giudice & Migliavacca, 2018; Fraser, Tan, Lagarde & Mays, 2018; Warner, 2013).

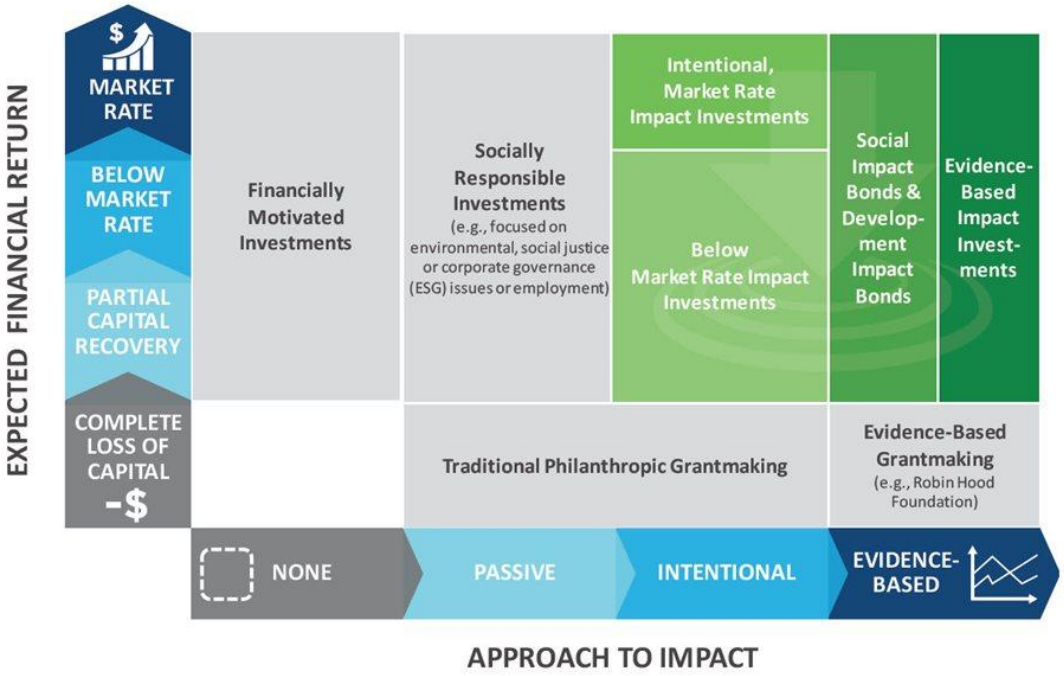


Figure 2: Different Types of Impact Investing.
Source: The Bridgespan Group (2018).

Unlike the name suggests, SIBs are not like typical bonds known from the financial industry. In a SIB, the return depends on the outcomes of the project. This means that the financier is not guaranteed to be paid back. Therefore, the mechanism is comparable to equity (Fraser, Tan,

Lagarde & Mays, 2018) or at least a hybrid financial instrument containing equity and debt logics (Arena et al., 2016). Overall, SIBs rather reflect a "collaborative partnership between public, private and/or third-sector stakeholders intended to achieve better social outcomes" (Carè et al., 2023, p. 2101) by leveraging private resources instead of public ones (Carè et al., 2020). The outcomes of the intervention need to be measurable and attributable to attract private investors. Therefore, SIBs often utilize proven interventions that increase the likelihood of success or reduce the risk for the investor, respectively (Warner, 2013). Santos et al. (2015) also state that outcome-based contracts are an ideal financing mechanism for scaling up of coupling hybrid organizations, i.e., organizations with business models in which the clients are not the beneficiaries and the value spillovers are contingent on additional interventions, when the outcomes are comparable.

SIBs use outcome-based contracts to achieve their goals and, as such, align with the public sector reform of the New Public Management movement (Fox & Morris, 2019; Warner, 2013). Outcome-based contracting generally has two approaches: Payment-by-Results and SIBs (Fox & Morris, 2019). The main difference between the two concepts is the existence of private investors in SIBs who cover the costs of the intervention upfront (Broccardo et al., 2020). In PbR contracts, the payments are only made after the services are provided, which might favor certain providers with sufficient capital available (Fox & Morris, 2019; Mulgan et al., 2011).

The rest of this chapter provides an overview of SIBs, how they work, how they are structured, and who the key actors are.

3.1. Development and Current Landscape of Impact Bonds

Since introducing the first impact bond in 2010, impact bonds, i.e., SIBs and DIBs, have gained increasing attention worldwide. As of May 14th, 2025, INDIGO (2025) lists 319 impact bonds that have been contracted, implemented, or completed. However, it should be remembered that this database is potentially not definitive, and the actual number might be even higher. The included impact bonds have raised a total capital of at least $867m, reaching more than 3.1m beneficiaries (INDIGO, 2025).

Of the 319 impact bonds listed in the INDIGO (2025) database, roughly one-third are completed, and two-thirds are currently in implementation (Figure 3). Only one impact bond is currently contracted, i.e., the service provision has not been started yet. This shows that the current pipeline is quite empty and could indicate a slowing momentum of this investment

mechanism. However, an explanation could also be that information about impact bonds in preparation is often not publicly available, so they cannot be included in the database.

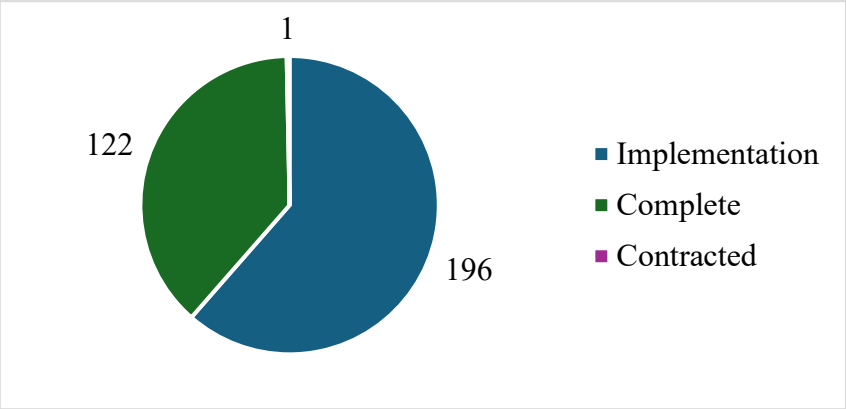


Figure 3: Number of Impact Bonds by Stage of Development.
Source: Author's elaboration from INDIGO (2025).

This impression is also reflected in the number of impact bonds per year (Figure 4). After a pilot phase in 2010 and 2011, impact bonds gained traction in 2012, and the number of started bonds per year increased significantly until 2018, when it reached its maximum with 47. However, the number of impact bonds declined massively and fell to only 4 in 2024 – a decline of more than 90% compared to 2018 and the lowest number since 2011. The considerable variations between individual years are partly attributable to the introduction of impact bonds facilitated by funds, a practice particularly prevalent in the UK. For instance, in 2012, the Innovation Fund (IF), administered by the Department for Work and Pensions (DWP) in the UK, commissioned 10 SIBs simultaneously, explaining the substantial increase compared to the preceding year (INDIGO, 2025). Nevertheless, a general trend is visible that impact bonds have been losing momentum over the last years.

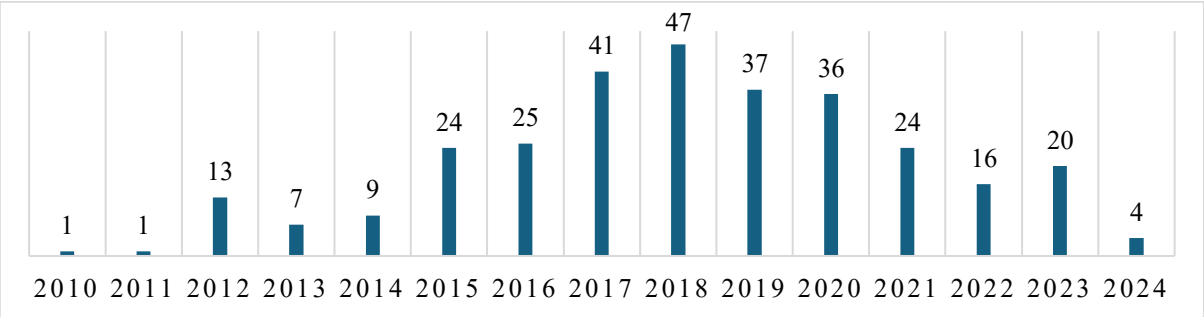


Figure 4: Number of Impact Bonds by Start Year.
Source: Author's elaboration from INDIGO (2025).

Figure 5 shows the geographic distribution of impact bonds. It can be seen that they have been implemented on all inhabited continents. However, more than 60% of them have been

commissioned in Europe, especially in the UK, where roughly one-third of all impact bonds are located. The top nine countries account for nearly 80% of all impact bonds, and all are developed countries. In many developing countries, the impact bonds take the form of DIBs, which function similarly to SIBs but also strongly focus on development within the countries (Spiess-Knafl & Scheck, 2023). From the geographic distribution, it could be assumed that impact bonds are only in the early stages of development across numerous continents and countries. However, considering the overall development over the past few years (Figure 4), there can be doubt whether that will happen.

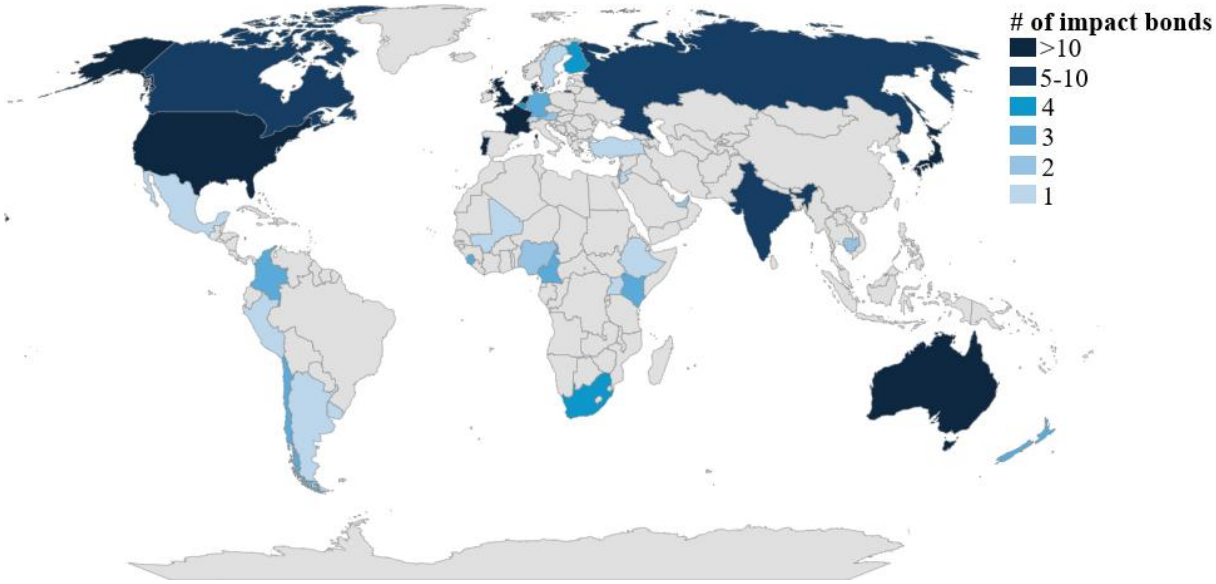


Figure 5: Impact Bonds by Location.
Source: Author's elaboration based on INDIGO (2025).

SIBs can be used to address a variety of social problems. Figure 6 shows the number of impact bonds by different policy sectors, ranging from poverty reduction to criminal justice and education. The most important one is, by far, employment and training, with 95 impact bonds.

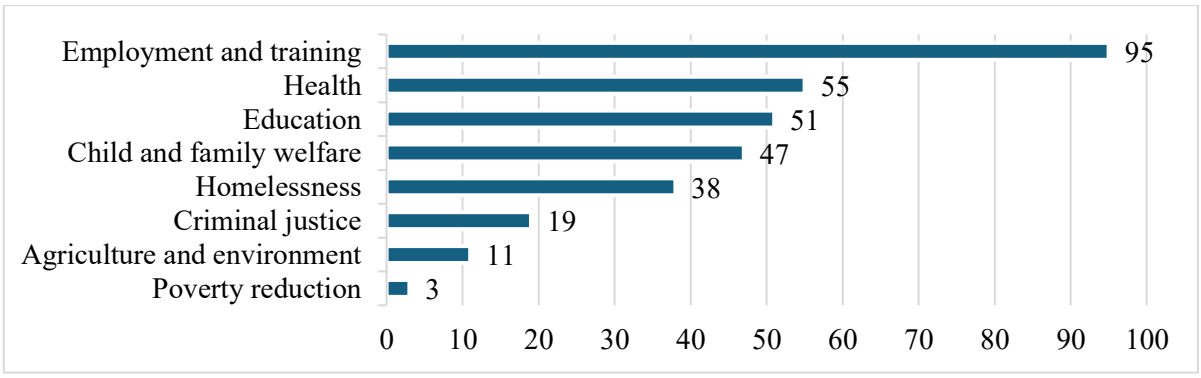


Figure 6: Number of Impact Bonds by Policy Sector.
Source: Author's elaboration from INDIGO (2025).

This general overview shows that impact bonds are a global phenomenon by now, reaching a substantial number of beneficiaries, but practitioners' interest seems to have decreased in recent years. The next chapter will examine the key actors within SIBs and their respective roles.

3.2. Key Actors and Their Roles

Figure 7 depicts a typical SIB structure with five key actors: A commissioner, an investor, an intermediary, a social service provider (SSP), and an evaluator. The following will examine the different actors and their relationships in more detail. The target population, which is also shown in Figure 7, is not seen as an actor since these individuals are merely recipients of social services and do not have an active role in it. This also highlights a special characteristic of social interventions, especially SIBs, that the users of the services are not the ones paying for them.

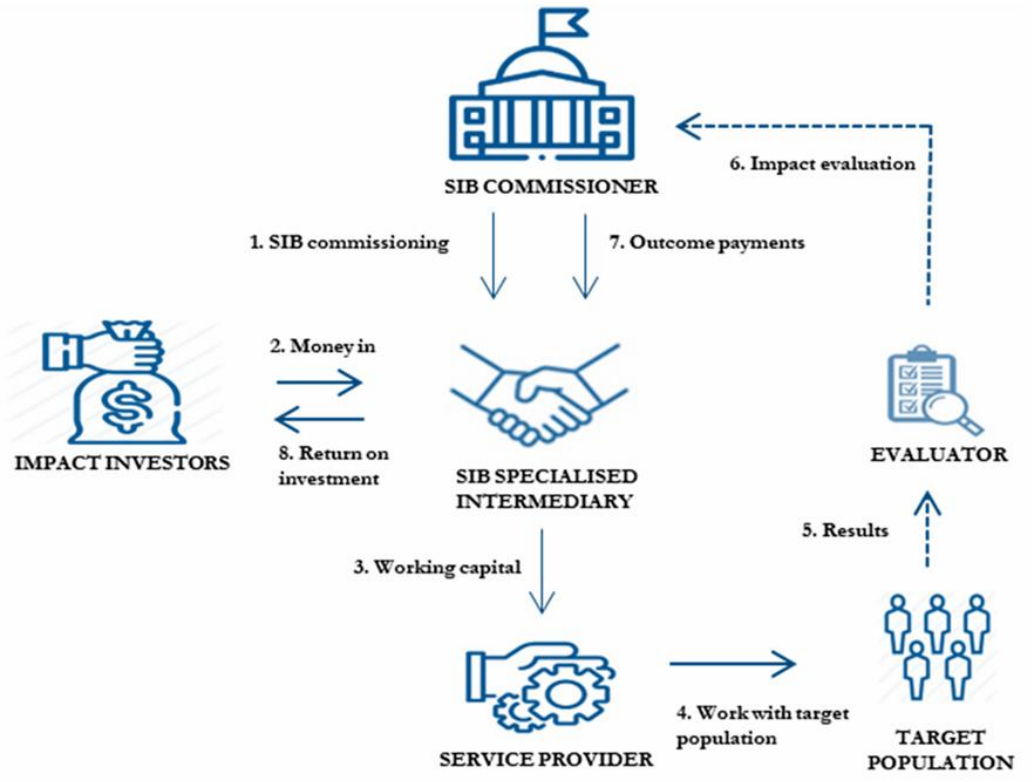


Figure 7: The SIB model.
Source: Rizzello & Kabli (2020).

3.2.1. Commissioner

The commissioner, typically a government agency at the national or local level, identifies the need for a SIB, determines the target population, and enters a contract with an intermediary (Carè et al., 2020). If the project is successful, i.e., if the pre-determined COMs have been achieved, the commissioner ultimately refunds the private investor's original investment plus a surplus (Carè et al., 2020; Rizello & Kabli, 2020). The commissioner is typically a public

administrator because the targeted policy sectors of the SIBs (see Figure 6) normally lie within the responsibility of the public administration. Due to the fundamental characteristics of government programs, the focus is predominantly placed on costly remediation measures instead of proactive prevention strategies (Warner, 2013), leading to potential cost reduction through SIBs.

3.2.2. Investor

The investor provides up-front the capital needed to perform the social services to target the identified social issues. He can be philanthropic, profit-oriented, or anything in between (Maier & Meyer, 2017). The investor is not guaranteed a fixed refund; it depends on the achievement of pre-determined outcomes of the intervention and is also contingent on the estimated cost savings the public administration has because of the preventive measures (Fraser, Tan, Lagarde & Mays, 2018; Nazari et al., 2018; Warner, 2013). The refund to the investor also includes a pre-specified return on investment (Del Giudice & Migliavacca, 2018; Rizello & Kabli, 2020; Warner, 2013). If the social project is unsuccessful, the investor loses all or parts of his money, depending on the set-up structure of the SIB (Berndt & Wirth, 2018), meaning he is generally assuming the risk of the program (Rizello & Kabli, 2020). However, there is variation in the risk the investor assumes, and he could choose a longer or shorter time to maturity, or demand a guarantee (Maier & Meyer, 2017).

3.2.3. Intermediary

The intermediary is responsible for arranging the SIB and aligning the different actors and their interests (Carè et al., 2020; Logue et al., 2024). Exemplary tasks of this actor are to engage with the commissioner to set up the SIB scheme with all its specifications, support raising capital by bringing in a potential investor, design the features of the financial instrument, help find a suitable SSP delivering the social services, facilitate negotiations with the stakeholders, and contract the evaluator (Del Giudice & Migliavacca, 2018; Logue et al., 2024; Rizello & Kabli, 2020). The intermediary also ensures the project's delivery (Rizello & Kabli, 2020) and manages the financial flows between the stakeholders (Del Giudice & Migliavacca, 2018). As shown in Figure 7, the intermediary is the central element in the SIB structure and therefore plays an important role in bringing the project to success.

3.2.4. Social Service Provider

The SSP, usually a social enterprise or a non-profit organization (Maier & Meyer, 2017), receives the money from the investor or the intermediary and is the actor that is delivering the

social service to the target population over a specified period (Del Giudice & Migliavacca, 2018; Millner & Meyer, 2022; Rizello & Kabli, 2020). The SSPs are usually not involved in contract negotiations, but they are of utmost importance to the performance of the SIB and its reputation due to their work.

3.2.5. Evaluator

SIBs can have different types of evaluators. The most prominent one is the external success evaluator or verifier. This evaluator has an important role in the SIB model by providing an independent assessment of the interventions' impact, verifying the achievement of the contractual targets, and determining whether the outcomes are attributable to the interventions (Millner & Meyer, 2022; Rizello & Kabli, 2020). His assessment determines whether the commissioner repays the investor and how much the investor receives (Carè et al., 2020; Warner, 2013). Warner (2013) stresses the importance of external evaluators for assessing the results of the SIBs because of the collusion risk among the different stakeholders, given a limited oversight by the government and an interest in positive evaluation results. However, Warner also highlights that it might not be enough to prevent perverse behavior.

Next to the success evaluator, there can be other evaluators, such as a process evaluator and an accompanying evaluator, as in the Austrian SIB PERSPEKTIVE:ARBEIT (Millner & Meyer, 2022). In this SIB, the stakeholders decided to have all three types of evaluators on board, but that is generally not the case for all SIBs. The process evaluator was responsible for analyzing the underlying processes of the SIB throughout its entire lifetime to identify improvement opportunities for future SIBs. The accompanying evaluator "investigated the experiences and conditions of the target group with the wider effects of the services provided" (Millner & Meyer, 2022, p. 726).

3.3. SIB Lifecycle

A typical SIB lifecycle involves the steps shown in Figure 7. It starts with identifying the social issue and the target population. Following, the commissioner engages in a contractual relationship with an intermediary, who, in turn, deals with finding an investor and SSP for the project. The SSP starts the work after the investor provides the necessary money. After the contractually agreed work has been done (Warner, 2013), an independent evaluator examines the project's success. In a successful assessment, the investor receives the pre-defined payment from the commissioner (Carè et al., 2020).

3.4. Agency Theory and Social Impact Bonds

Given the focus of the thesis on COMs and the payment structures of SIBs, which are integral parts of SIB contracts and inherently aim to align the interests of different involved stakeholders, agency theory seems appropriate to discuss the study's findings.

Agency theory, which originated in the 1970s and 1980s (Fama, 1980; Fama & Jensen, 1983; Harris & Raviv, 1978; Jensen & Meckling, 1976), has evolved into one of the most influential economic theories. Building on earlier research on private property, managerial intentions, and risk sharing among individuals or groups (cf. Bendickson et al., 2016; Eisenhardt, 1989), it conceptualizes the relationship between a principal, who delegates work and authority, and an agent, who is tasked with performing that work (Eisenhardt, 1989; Jensen & Meckling, 1976; Mitchell & Meacheam, 2011).

Although agency theory is frequently used to explain relationships between managers and shareholders, it can be applied to numerous other contractual scenarios (Eisenhardt, 1989; Hill & Jones, 1992; Jensen & Meckling, 1976). A central concern is the "agency problem," which occurs due to a conflict of interest between the principal and the agent, and monitoring the agent's behavior is either difficult or expensive, potentially leading to an opportunistic behavior of the agent to the detriment of the principal (Eisenhardt, 1989; Jensen & Meckling, 1976; Mitchell & Meacheam, 2011). Risk sharing is also an issue the theory addresses, as the principal and agent may hold different attitudes toward risk (Eisenhardt, 1989).

To address these problems, agency theory seeks to determine the most efficient contract to govern the principal–agent relationship. This approach is based on several assumptions, including self-interest, goal conflict, bounded rationality, information asymmetry, an emphasis on efficiency, risk aversion, and the treatment of information as a commodity (Eisenhardt, 1989). In practice, to mitigate the agency problem, the theory suggests incentivizing the agent's behavior, leading to bonding costs, and monitoring the agent, incurring monitoring costs. A residual loss inevitably remains since it is impossible to perfectly align the agent's actions with the principal's interests without cost. This residual loss decreases the principal's welfare. The monitoring costs, bonding costs, and residual loss collectively make up the overall agency costs (Jensen & Meckling, 1976).

Agency theory is commonly divided into two complementary streams: positivist and principal–agent (Jensen, 1983). The positivist stream highlights governance mechanisms, especially in large corporations, and identifies alternative contractual arrangements to mitigate agency risks

(Eisenhardt, 1989). In contrast, the principal–agent stream uses mathematical models to determine the optimal contract for a specific situation (Eisenhardt, 1989). Given the outcome-based nature of SIBs, which require careful impact measurement, applying agency theory to assess how well current COMs and payment structures function offers a promising way to evaluate their effectiveness.

3.5. Alignment of Interest

The alignment of interest is central for the success of SIBs, which involves a highly complex principal-multiagent relationship with the risk of agency problems due to diverging interests and information asymmetries (Becchetti et al., 2021; Del Giudice & Migliavacca, 2018). It was found that the two primary motivations for stakeholders engaged in SIBs are either socially oriented or financially oriented (Carè et al., 2020). A discrepancy in motivation can lead to conflicts, especially when it is not adequately addressed. Additionally, due to the financial risk transfer to the investor, it is important to ensure that the commissioner and the SSP are encouraged to direct their activities towards achieving the objectives (Albertson et al., 2020). It is believed that a well-defined SIB can fulfil this purpose of aligning the different interests by bridging the financial interests of the investor with the desired social outcomes and ensuring that the interventions are focused on achieving the COMs (Arena et al., 2016; Maier & Meyer, 2017; Rizello & Kabli, 2020). For this, it is also important to have sufficient monitoring and evaluation mechanisms to attribute the outcomes to the intervention and validate them (Rizello & Kabli, 2020).

A problem for aligning interests is that only selected stakeholders, such as the commissioner, the intermediary, and the investor, are usually involved in setting up the formal SIB contract (Maier & Meyer, 2017). Maier & Meyer (2017) derive two potential perils from this partial alignment of interests. First, they argue that the specific COMs chosen can influence the behavior of the SSP and can create perverse incentives. Taking binary metrics, which are easier to implement and measure, instead of frequency metrics could lead to situations in which the SSPs prioritize clients with a higher likelihood of success to ensure the COMs are achieved. This would simultaneously contradict the traditional beneficiary-centered focus of the SSPs. Second, the authors see a potential disadvantage for taxpayers when public agencies are getting a so-called “SIB fever” (Maier & Meyer, 2017, p. 6), in which they want to introduce SIBs whenever possible, even when they could lead to higher costs compared to traditional funding if the COMs are too easy to achieve. Giacomantonio (2017) also highlights a fundamental

conflict. He argues that the rational choices of commissioners and investors conflict, making it difficult to attract private capital. He also stresses that this issue persists even when transaction costs or investor risks are reduced (Giacomantonio, 2017). Morley (2019) highlights that informational asymmetry and power imbalances between the different actors create potential agency problems where the incentives of financially driven investors might not prioritize the best interests of the beneficiaries. This might lead to a focus on easily measurable COMs or practices of "cream-skimming" (Morley, 2019, p. 55) whereby clients are picked by how likely they are to succeed. The misalignments can lead to individual or societal harm (Morley, 2019).

Overall, a well-defined, transparency-enhancing SIB can reduce agency problems, but the actors should engage in extensive negotiations to reach the best alignment possible (Maier & Meyer, 2017). This would also attract more private capital as institutional investors favor SIBs with fewer agency problems (Del Giudice & Migliavacce, 2018). Carè et al. (2020) also emphasize the importance of a robust contract to reduce conflict potential. Research by Hevenstone et al. (2023) suggests that the positive effects of SIBs could be explained through agency theory by increasing the pressure on SSPs and governments to perform. However, the same research also finds evidence for a stewardship model.

3.6. Varying Structures of Social Impact Bonds

SIBs are the most common type of outcome-based payment program. SIB contracts are an important tool for aligning the interests of the various stakeholders. Given the multitude of different use cases for SIBs and the need to account for individual circumstances, they can vary substantially in their characteristics, e.g., in the specific stakeholders involved or the contractual characteristics, such as COMs or payment structures (Berndt & Wirth, 2018; Del Giudice & Migliavacca, 2018; Fraser, Tan, Lagarde & Mays, 2018; Nazari et al., 2018). Additionally, there are similar types of impact bonds, such as DIBs. This chapter will examine the different variations and types in more detail. The COMs will be covered separately as part of Chapter 4.

3.6.1. Social Impact Bond Models

Mulgan et al. (2011) distinguish three types of SIBs: Philanthropic SIBs, public sector SIBs, and commercial SIBs. There are also hybrid SIBs, which combine at least two of these types. In *philanthropic SIBs,* the funding happens through donations via a special purpose vehicle (SPV). This model allows for more innovation and risk-taking since philanthropists are more flexible than traditional funders, and it shifts financial risks away from the government, making it very appealing for them. *Public sector SIBs* involve local authorities funding programs

themselves and being paid back by the government if they succeed. It is a simpler model compared to the philanthropic SIB with low overhead, but would be better suited for large, experienced authorities. Liebman & Sellman (2013) and Nazari et al. (2018) also mention this SIB model. *Commercial SIBs* involve private investors and focus on proven programs with reliable partners for better risk assessment. Investors expect financial returns if the SIB succeeds (Mulgan et al., 2011).

The literature also mentions different SIB models regarding the structure of the relationships between the involved parties. Mulgan et al. (2011) distinguish between the "streamlined approach, [the] lead delivery agency approach, and [the] special purpose vehicle" (p.9). In the *streamlined approach*, no investor exists. The lead agency funds the intervention by itself. It may subcontract another SSP or deliver the services directly. Like the typical SIB structure, the government will repay the agency if pre-defined outcomes are reached. In the *lead delivery agency approach*, an investor directly funds the intervention by giving the money to the lead agency, and the agency, or a subcontractor, then implements the intervention. The commissioner repays the investor. Lastly, the *SPV approach* represents the typical SIB structure explained above. The SPV, a separate legal entity, acts as the intermediary and coordinates the SIB. However, a SIB does not always have to have a SPV. Often, the intermediary is a separate organization, like a nonprofit organization (NPO).

In addition to the *public sector SIB*, Nazari et al. (2018) mention two more SIB models, which show similarities to the ones mentioned by Mulgan et al. (2011). Their *no intermediary model* can be compared with the *lead delivery agency approach*, if the lead agency delivers the services itself. Their last model states a case in which no contract between the commissioner and the investor is in place, and the SSP is the focal point of all transactions (Nazari et al., 2018). This model is most comparable to the SPV. However, in the SPV, the SSP is only subcontracted and is not responsible for all transactions.

3.6.2. Payment Structures

Payment structures are an important element of SIB contracts and influence the financial success of a SIB, depending on whether a repayment took place or not. In this study, the payment structure describes the structure when outcome payments are triggered. Even though Del Giudice & Migliavacca (2018) find in their study on the critical success factors of SIB funding that contractual characteristics like the number of investors or having a local authority as part of the SIB are more important than financial characteristics in attracting institutional

investors, the investors still choose the least risky form of funding. Both types of characteristics influence the risk of a SIB. The authors do not include the payment structure as a variable in their quantitative model. However, this characteristic is situated at the intersection of financial and contractual characteristics, and it influences the risk perception of the investor.

From the scientific literature and the SIB evaluation reports, there are generally two common types of payment structures visible. The first widespread structure, especially in the UK, is a *staged payment*. The outcome payments are tied here to specific milestones, meaning that the investor will be reimbursed with parts of the total repayment every time a milestone is achieved. This structure reduces the risk for the investor, as it is unlikely that the investor will lose the total investment (Berndt & Wirth, 2018). This structure should therefore be better suited to attract institutional investors. Within this structure, there are two prominent sub-structures – *payment-per-outcome* and *percentage-based*.

In *payment-per-outcome* structures, the investors are reimbursed for every COM achieved, which is often done on the level of individual persons. For example, in the SIBs commissioned by the IF from the DWP, there were ten different COMs. For each individual achieving one of the COMs, the investors were paid a fixed price per person, depending on the specific COM (Salis et al., 2018). The price per achieved COM is fixed in a rate card prior to the start of the SIB.

In the *percentage-based system*, the payments are triggered when the COMs are improved by a specific pre-defined percentage value. This payment structure is, for example, frequently used for prisoners' recidivism. Good examples are the HMP Peterborough, the first SIB ever implemented, or the New York City Adolescent Behavioral Learning Experience (ABLE) (Disley et al., 2015; Parsons et al., 2016). In the ABLE program, the minimum decline in juvenile recidivism for repayment was 10%. Then there were different stages up until a decline of 20% that would increase the return to the investor (Parsons et al., 2016).

The other typical payment structure type is an *all-or-nothing* agreement. This describes a structure in which the investor is repaid the entire investment plus an additional return or nothing. This also means that all the COMs must be achieved for repayment. This approach is more common in central European countries like Austria or Germany (Millner & Meyer, 2022). In their study, Millner & Meyer (2022) also found that the investor of their studied SIB, which did not meet all COMs, did not favor this payment structure and would not engage again in a

SIB with this structure. The reaction is understandable as the investor must bear a substantially higher risk in this agreement than in the other payment structures.

Berndt & Wirth (2018) also mention the utilization of caps on returns. Considering the high risks involved in SIBs, this is not a good way to attract institutional investors, who typically demand higher returns as compensation for bearing more risks.

3.6.3. Other Types of Impact Bonds

SIBs have been implemented in other areas as well. One example of this is DIBs. DIBs generally have a similar goal to SIBs in tackling social issues. However, they focus on developing countries and improving development assistance in those countries. In contrast to SIBs, the outcome payer is typically not the host-country government, but government agencies from developing or donor countries or private institutions, such as philanthropic foundations (Dey & Gibbon, 2018; Spiess-Knafl & Scheck, 2023). Another variation is Environmental Impact Bonds (EIB), which focus on environmental projects, but operate similarly to SIBs. One potential advantage of EIBs over SIBs is that there already is a plethora of standardized metrics on environmental outcomes, which may be applicable for EIBs and therefore facilitate impact measurement (Dey & Gibbon, 2018). This study will focus solely on SIBs to allow for a better comparison of the COMs and the results.

3.7. Critical Reflection of Social Impact Bonds

Fraser, Tan, Lagarde and Mays (2018) identified three narratives about SIBs: Two of them – the *public sector reform narrative* and the *private financial sector reform narrative* – have a very positive view of the mechanism and are very prominent among grey literature, while the *cautionary narrative* is more critical about adopting "'private sector' value and mechanisms in the field of public services" (p. 12) and is more prevalent in the academic literature. This discussion of the advantages and disadvantages of SIBs is very present in the literature.

3.7.1. Advantages

On the one hand, SIBs offer several potential advantages by aligning public and private interests to drive social innovation. Shifting to outcome-based funding enhances goal clarity and financial stability for non-profits (Fraser, Tan, Lagarde & Mays, 2018; Warner, 2013). SIBs also transfer financial risk from the state to private investors. This enables funding for flexible and experimental interventions by the SSPs that are tailored to the local needs and might otherwise be too risky for public investment (Andreu, 2018; Fraser, Tan, Lagarde & Mays,

2018; Maier et al., 2018). This possibility for experimentation allows for testing existing interventions in different settings and thereby could also lead to a scaling of successful interventions (Maier et al., 2018). It is also argued that the opportunity for new approaches leads to better results for the beneficiaries (Andreu, 2018). They strengthen performance measurements, support evidence-based policymaking, and create new revenue opportunities for private financiers while improving their public image (Barajas et al., 2014; Broccardo et al., 2020; Fraser, Tan, Lagarde & Mays, 2018; Wilson, 2014). Additionally, SIBs promote preventative measures and cost-effective delivery models, reducing long-term public costs (Andreu, 2018; Millner & Meyer, 2022). Further, socially motivated investors in SIBs need to be more patient and willing to accept lower returns and higher risks (Carter, 2019). They are often more involved in the SIB to make it a success (Carter, 2019; Griffiths et al., 2016).

3.7.2. Disadvantages

On the other hand, SIBs come with significant drawbacks, and some of the arguments used as advantages of SIBs are simultaneously viewed critically by other scholars. The emphasis on outcome-based funding raises concerns about measurement accuracy, attribution of success, and potential gaming of performance metrics (Fox & Albertson, 2011; Warner, 2013). SIB contracts are often complex, which leads to high transaction costs, especially during the development period, and challenges impact measurement with a standardized risk assessment, thereby making long-term scalability difficult (Andersen et al., 2020; McKay, 2013; Millner & Meyer, 2022; Pauly & Swanson, 2013).

Additionally, while SIBs aim to shift financial risk to private investors, they make investments unattractive for many profit-oriented investors due to the high risks involved. In practice, government and philanthropies are sometimes required to underwrite investments, reducing the supposed transfer of risk (Fraser, Tan, Lagarde & Mays, 2018; Warner, 2015). Investors also want to be compensated for taking the risks, and given that public agencies have cheaper access to capital, the private investments could lead to higher costs in the end (Mulgan et al., 2011). There is also concern about the privatization of social services that lie in the responsibility of public agencies (Edmiston & Nicholls, 2018), and that financial incentives could lead to a subordination of public policy to financial interests (Lake, 2015) or distort the social missions of non-profits (McHugh et al., 2013), leading to detrimental activities like ‘creaming’ (Carter, 2019; Fraser, Tan, Lagarde & Mays, 2018; Warner, 2013).

Furthermore, focusing on measurable outcomes may stifle innovation by favoring interventions with proven metrics over more experimental approaches (Warner, 2013). This also indicates a conflict between the evidence-based interventions and the desired flexibility, which is seen as an advantage of SIBs (Maier et al., 2018). Some SIB contracts define the interventions upfront, entirely removing the flexibility (Maier et al., 2018). SIBs also lack direct consumer input, giving primary decision-making power to the other actors rather than the communities affected by these interventions (Warner, 2013).

Another issue is that SIBs often focus on tackling surface problems by focusing on individual outcomes and do not address the underlying structural issues that lead to the surface problems (Andreu, 2018). This means the problem is not solved, and the long-term cost reduction is not as high as possible.

These challenges highlight the need for careful design and oversight to prevent unintended consequences in social service delivery. Addressing the concerns of SIBs could also increase the capital stemming from institutional investors, but to date, the success in that regard is limited (Bafford, 2012; Del Giudice & Migliavacca, 2018). It could also be that SIBs primarily offer philanthropic and charity funders a way to efficiently use their granting capital, potentially leading to a different dynamic of interest alignment (Giacomantonio, 2017). The next chapter will dive deeper into impact measurement and outcome metrics.

4. Impact Measurement

After implementing SIBs, impact measurement and the evaluation of the results become important to determine success. Impact is defined as "significant or lasting changes in people's lives, brought about by a given action or series of actions" (Roche, 2000, pp. 545-546). Impact measurement is repeatedly highlighted as a complex issue in the literature. However, it is important to attract institutional investors that rely on quantitative results and therefore need defined metrics, evaluation methods, and goals (Berndt & Wirth, 2018). Liebman (2011) notes that a credible impact assessment is one of the necessities for a successful SIB, besides having high net benefits and short-term payouts, working performance measures, and a clearly defined treatment population. However, it can be challenging to be confident that the desired outcome has been achieved because of the actions of the SSP, i.e., there is a counterfactual risk (CR) (Fox & Albertson, 2011, 2012).

The basis for impact measurement are COMs. COMs are defined in the SIB contract and are pre-defined targets that must be achieved to initiate payments. In that way, COMs can be seen as indicators of impact. The stakeholders assume that following and achieving these targets will lead to the desired impact of the SIB. The achievement of these COMs is measured and evaluated in the end. There is widespread agreement that outcome measures must be carefully defined, measurable, and agreed upon by all relevant actors (especially commissioners, SSPs, and investors; Berndt & Wirth, 2018; Fox & Albertson, 2012; Warner, 2013). The development of robust outcome measures can be time-consuming and complex (Disley et al., 2011), given the divergent interests and risk attitudes of the different stakeholders (Warner, 2013), leading to potentially complex negotiations (Fraser, Tan, Lagarde & Mays, 2018). It is also debatable whether the COMs measure the full impact a SIB creates or merely focus on financial success. This differentiation will be further discussed in the following sub-chapter.

4.1. Defining Success in Social Impact Bonds

It becomes apparent from the existing literature that COMs and the impact measurement and evaluation are important for determining the success of SIBs. From a financial perspective, success is linked to achieving the metrics and the subsequent repayments to the investors. However, it remains arguable whether these measures fully capture success, and Broccardo et al. (2020) also call for attention in future research when using payment triggers as proxies for success. Millner and Meyer's (2022) evaluation of an Austrian SIB also highlights potential shortcomings of the success criteria employed in that SIB. They state that only economic impact targets were used, which raises the question of whether this equals success, especially when considering qualitative, social outcomes. This is also reflected in a quote from the participating SSP, which stated that the target beneficiaries benefited anyway, regardless of the achievement of the COMs. Insofar, the project was still considered a success even though the COMs were ultimately not met. Hence, investigating the success of SIBs not only from the perspective of achieving COMs would add value to the current literature, especially because achieving COMs does not have to be equal to addressing the root causes of the issues and generating social impact (Andreu, 2018).

Considering the above, the definition of success is crucial for this thesis. The easiest way to define success would be to determine whether the pre-defined COMs have been achieved and, subsequently, if the investors received money. However, this approach might be shortsighted for several reasons. First, from a social perspective, the SIBs might still be considered

successful due to the generated social benefit even if a predetermined threshold defined in COMs has not been achieved (Giacomantonio, 2017). Second, there is a lot to learn from 'failed' SIBs that can improve the design of future contracts (Giacomantonio, 2017). Third, there is the phenomenon of impact washing and the broader aspect of the ethics of impact investing (Chiappini et al., 2023; Dahbi et al., 2024). It could be that the COMs are ill-defined from the beginning and incentivize questionable behavior like gaming or creaming. A prominent example is the London Homelessness SIB, where the deportation of homeless people was one of the defined COMs and considered a success (Pequeneza, 2019).

To facilitate the analysis of the evaluation reports in this study and build a database for a quantitative analysis, success is seen as achieving the COMs if it cannot be determined otherwise, e.g., by using the results of an SROI analysis. However, to account for this approach's shortcomings, the used COMs are also critically assessed, and the reports qualitatively analyzed to evaluate whether qualitative outcomes are also discussed and if the perceptions of these outcomes differ from the achievement of the COMs.

4.2. Designing Appropriate Outcome Metrics

COMs are a crucial element of SIBs as they are key to aligning the interests of the different actors and defining target outcomes to be achieved. Therefore, COMs must incentivize actions in favor of the beneficiaries, e.g., by connecting the metrics to expected savings for the commissioner (Social Finance, 2013). Generally, COMs should be based on the objectives of the SIB, whether it is a scale-up of evidence-based interventions or testing innovative services (Social Finance, 2016). The more innovative-oriented a SIB is, the more qualitative and outcome-oriented the COMs should be in focusing on the quality of the service delivery and giving the SSP more flexibility (Andreu, 2018; Social Finance, 2016). A stronger evidence-based approach defines more strictly and narrowly how the interventions are to be carried out (Social Finance, 2016). However, in practice, there is not only an either-or perspective regarding quantitative and qualitative metrics, but they are often used simultaneously (Berndt & Wirth, 2018). An example is the Essex Edge of Care SIB, which used the aggregate number of care placement days saved as a primary outcome and was relevant for the repayment of the investors (Social Finance, 2021). Additionally, wider emotional well-being outcomes and the impacts on children and families were tracked (Social Finance, 2021).

The specific COMs should be measurable, objective, carefully defined, and agreed upon by all stakeholders involved in the SIB contract (Berndt & Wirth, 2018; Social Finance, 2013;

Warner, 2013). A potential bottleneck here is the stakeholders' divergent perspectives and risk attitudes. While a public commissioner is more risk-averse and requires high success rates of the intervention, a private investor is generally more risk-tolerant and used to lower success rates (Warner, 2013). This can lead to difficulties in the definition of the COMs. There are principally two types of outcome metrics, between which a choice must be made – binary and frequency metrics. Binary metrics, e.g., whether an unemployed person has found employment, only have two outcomes – either the individual achieves the metric or not (Maier & Meyer, 2017). Frequency metrics take a different approach and measure how often an event occurs or does not, compared to a situation without intervention (Maier & Meyer, 2017). An example is the number of care days saved in the Essex Edge of Care SIB (Social Finance, 2021). Frequency metrics are usually preferable to binary metrics but are also harder to establish (Maier & Meyer, 2017). Binary metrics are more likely to create perverse incentives that lead to actions like cream-skimming. One way to deal with that limitation is to introduce several counterbalancing binary metrics. While this seems initially advantageous, it creates another problem. SSPs could take this opportunity to focus on easier-to-achieve outcomes, thereby not achieving the initial intended results (Maier & Meyer, 2017).

Oftentimes, several possible proxies can be used to measure the same outcome (Fox & Albertson, 2012), complicating comparisons of the outcomes of different SIBs. Additionally, some scholars assume that different SIBs in the same policy area will use a similar set of COMs to simplify the process of target definition and measurement (Fox & Albertson, 2012). Others note that the COMs are very individual, encompassing qualitative and quantitative measures, and that SIBs often fail to address the structural, underlying problems (Andreu, 2018; Berndt & Wirth, 2018; Cooper et al., 2016; Morley, 2019), which is a problem especially for quantitative metrics that might simplify complex problems (Dayson, 2017). Outcome-based systems could also incentivize SSPs to focus only on achieving pre-defined outcomes, neglecting other important measures (Fox & Albertson, 2012).

Overall, choosing the correct form of the COMs is difficult and depends on the specific situation. However, it is always important to specify the metrics, including the target group, as much as possible to avoid perverse incentives and a manipulation of the metrics (Fraser, Tan, Lagarde & Mays, 2018; Liebman, 2011). Additionally, the number of COMs should be kept as low as possible so that the core objectives can be focused on (Social Finance, 2016). Looking at SIBs that address similar issues and their respective success can help define COMs. This is also what the second part of the thesis tackles. It provides an overview of COMs used in a

selected set of completed SIBs and qualitatively analyses the stakeholders' perceptions towards the COMs.

4.3. Conceptual Approaches to Impact Measurement

Impact measurement is a challenge for SIBs and impact investing in general. It describes the process of identifying the results achieved during or after a given activity. It can generally be done on a factual or counterfactual basis. Factual means that the baseline is the original state, and counterfactual takes as the baseline the situation that would have occurred without the intervention. Two important concepts for impact measurement are the *Theory of Change* and the *Logic Model*, which complement each other (see also Figure 8).

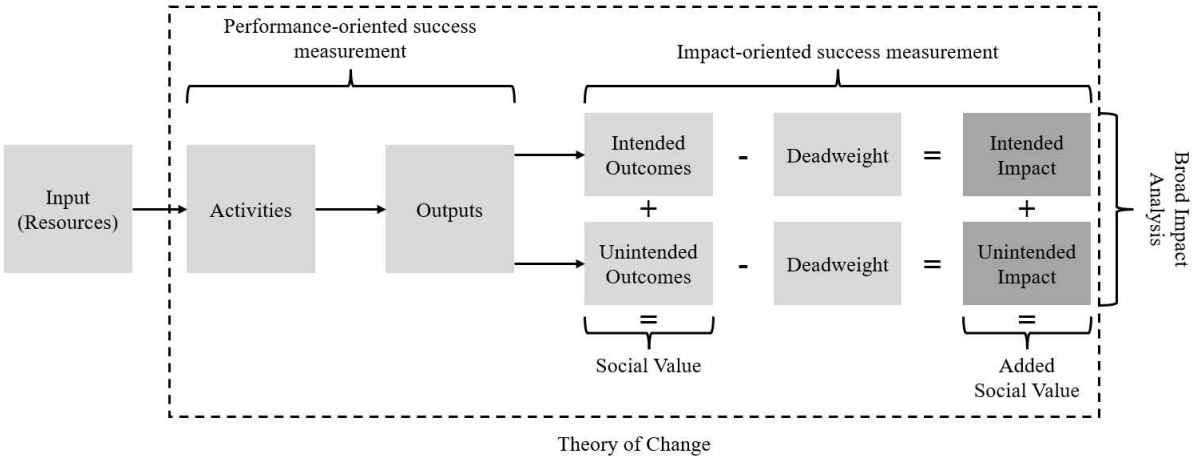


Figure 8: Logic Model and Theory of Change.
Source: Adapted from Grünhaus & Rauscher (2022).

As illustrated by Grünhaus & Rauscher (2022), the *Logic Model* describes the way from inputs to impact. Resources, such as funds or know-how, are used as inputs for activities to produce outputs, which in turn lead to outcomes and ultimately impact. To get from outcomes to impact, the so-called deadweight, i.e., the counterfactual, has to be deducted from the outcomes to receive the net effect. Every intervention has intended outcomes and impacts defined prior to the program, but there can also be unintended results. Both intended and unintended outcomes and impacts can be negative or positive, and they can also be contradictory to each other. Grünhaus and Rauscher (2022) provide an example of the inclusion of disabled people into the workforce. If disabled people find a job and are included, this is a positive intended outcome. However, if formerly employed people lose their jobs as a result, this is a negative unintended outcome. The sum of the two types of impacts is the added social value of the intervention. Considering only the intended outcomes when evaluating an intervention is too narrow a

perspective and could potentially be too positive. This could be a potential problem with evaluating SIBs if they only consider the intended outcomes.

Outputs are generally easier to measure than outcomes or impact, because they can usually be quantified. An exemplary output could be the service hours provided (Grünhaus & Rauscher, 2022). Outcomes and impact are harder to measure because they describe the changes for the beneficiaries due to the intervention. As described above, the impact measures the change by considering the counterfactual, i.e., what would have happened without the intervention (Grünhaus & Rauscher, 2022).

Grünhaus and Rauscher (2022) also differentiate several dimensions of impact. Time-wise, impacts can be short, medium, or long-term. They can also be on different levels, from the individual to the organization and society at large. Finally, the authors define different categories of impact – economic, social, political, ecological, cultural, and physical/psychological.

While the *Logic Model* explains what happens in the different stages, the *Theory of Change* provides the why perspective. According to Weiss (1995), it describes the causal mechanisms of how the impacts are assumed to be achieved through the inputs and activities. A well-defined theory of change helps to understand the impact generated by the intervention and the factors that led to failure or success. It is facilitated by a rigorous evaluation that uncovers the underlying assumptions and investigates whether they hold. Weiss (1995, pp. 69-72) further argues that there are four primary purposes of why it is helpful to base evaluations on the theory of change: (1) it helps to focus the evaluation on the key aspects of the program, (2) it facilitates the aggregation of evaluation results to build overarching theories of change, (3) it helps to align the interests and perspective of the different actors and ensure that they work towards the same goal, and (4) the evaluations might have a more substantial influence on policy and popular opinion. Grünhaus & Rauscher (2022) note that a theory of change often only focuses on the intended outcomes, impacts, and selected stakeholders. Therefore, they use the term *Impact Model*, which allows for a broader impact analysis. According to their understanding, an impact analysis consists of seven steps, starting with identifying the stakeholders and beneficiaries and defining the impact model. The remaining steps concern identifying, measuring, and evaluating the impacts. Prior to the start of the analysis, the purpose should be defined.

4.4. Evaluation Design

The evaluation design should ensure a high internal validity of the assessment, so that the measured effects can be attributed to the intervention (Fox & Albertson, 2011). High internal validity requires using experiments, primarily randomized controlled trials (RCTs), or quasi-experiments (Fox & Albertson, 2011). The main advantage of these designs is that they use control groups to assess the effect of the intervention, thereby allowing for attribution of the intervention effects (Fox & Albertson, 2011; Mulgan et al., 2011; William, 2021). The control group should be as similar as possible to the treatment group to isolate the treatment effect, i.e., the SIB intervention (Social Finance, 2013). However, they also come with disadvantages. Such evaluation designs are more complex, more costly, and can also be harder to establish due to the difficulty of finding a control group or a lack of quantitative data (Fraser, Tan, Kruithof, et al., 2018; Fraser, Tan, Lagarde & Mays, 2018; Mulgan et al., 2011; Williams, 2021). Williams (2021) also notes that RCTs require large sample sizes, leading to higher costs and longer timelines, making real-time adjustments to the SIB program less feasible. Investors also see a problem in the high statistical requirements, which makes it less likely that payments are triggered (Williams, 2021). Another potential design is to use historical data as benchmarks. This option is less costly and allows the same individuals to be in the treatment and comparison group. However, it also has a lower internal validity because external effects are not controlled, and the data quality can be an issue. Furthermore, it is possible to do pre- and post-intervention measurements for the target group. While this approach is easy to implement, there is also the problem of a lack of control over external effects, and the attribution of the results is difficult (Williams, 2021). Using control groups is a more scientific approach where the evaluation requirements are more demanding than more practical approaches like historical benchmarks (Grünhaus & Rauscher, 2022).

4.5. Methods and Tools for Impact Assessment

Social impact measurement is not standardized, and a large variety of methods and tools can be used (Arena et al., 2016; Grünhaus & Rauscher, 2022). This complicates the development of SIBs, and since not all methods measure impact similarly, their impact assessments are also less comparable (Arena et al., 2016; O'Flynn & Barnett, 2017).

Grünhaus and Rauscher (2022) identified different clusters of methods for impact assessments. First, they differentiate between simple and comprehensive impact analyses. While simple impact analyses focus on individual effects, comprehensive impact analyses, such as Social

Return on Investment (SROI) and cost-benefit analysis, consider impact mechanisms across several stakeholders and quantify as well as evaluate. Next to these two clusters, there are different types of tools. Reporting and rating tools, like the Social Reporting Standard or the IRIS+ from the GIIN (GIIN, 2025b), try to alleviate the lack of standardization and provide guidance for actors in the social impact area (Grünhaus & Rauscher, 2022). However, the authors criticize the focus on performance indicators and the missing inclusion of factors of social value. The remaining clusters relate to strategy and control tools, like the impact model or the logical framework, which are seen as conceptual approaches to impact measurement in this thesis.

Since tools for comprehensive impact analyses are highlighted as advantageous, examining them more closely is useful. The focus will be on SROI, the most comprehensive tool (Grünhaus & Rauscher, 2021). SROI "is a methodological framework for estimating the value created by an intervention across three realms: social, economic, and environmental – referred to as the triple bottom line" (Fischer & Richter, 2017, p. 105). SROI has a broader focus than SIBs, and while a SIB is a financing tool for social interventions, SROI is used to estimate the impact of interventions. This method aims to monetize the value of long-term outcomes of interventions and contrast it to the initial investment costs, known as the SROI ratio (Fischer & Richter, 2017; Grünhaus & Rauscher, 2021; Maier et al., 2015). Counterfactuals are very important in calculating the SROI because the net impact of the intervention is of interest to this metric (Grünhaus & Rauscher, 2021). SROI can be done retrospectively, focusing on existing outcomes, or predictively, trying to estimate future outcomes. Compared to a cost-benefit analysis, which also aims to monetize outcomes, SROI has a stronger focus on the social sector and involves a broader range of stakeholders (Fischer & Richter, 2017).

SROI has significant potential upside, but this method also comes with challenges. On the upside, this method can help make efficient and effective decisions on allocating resources or encourage cross-sector partnerships (Fischer & Richter, 2017; Maier et al., 2015). Maier et al. (2015) argue that the process of SROI analysis may lead to NPOs clarifying their goals and underlying assumptions and shifting the focus from outputs to impacts. However, they also note that the internal learning for the NPO is not as important as the signaling effect to external parties. Another advantage mentioned by Maier et al. is that an SROI analysis can help NPOs or funders to legitimize their actions towards other stakeholders, potentially leading to better funding conditions and a better perception in the court of public opinion. However, the evidence is lacking, especially regarding the funding (Maier et al., 2015; Millar & Hall, 2013).

On the other hand, Maier et al. (2015) identify three groups of limitations. The first one concerns fundamental controversies of SROI. The method is rooted in utilitarianism and can be ethically criticized. Additionally, calculating the SROI ratio implies that qualitative and priceless outcomes must be quantified. The resulting numbers should therefore be treated with caution. This is in line with Millar and Hall (2013), who note that the focus on monetization often contrasts with the value of social entrepreneurs and could lead to a lack of understanding. Additionally, there are also practical issues. Especially in social entrepreneurship, outcomes are often 'soft', making them hard to monetize (Millar & Hall, 2013). Secondly, Maier et al. (2015) identify a group of limitations inherent to the method, but do not prohibit using it. The SROI should not be used as a single indicator for the performance of the intervention and needs to be accompanied by more qualitative information. Furthermore, the SROI ratio can hardly be compared across different contexts, because the calculation is very individual. This also makes it hard to standardize the measurement, reducing the reliability and validity of the SROI. More thorough evaluations can reduce the need for discretionary assumptions and proxies. However, this increases the necessary time and resources, which is already a limitation of the SROI and something that not every organization possesses (Maier et al., 2015; Millar & Hall, 2013; New Philanthropy Capital, 2010). Third, Maier et al. (2015) mention the group of solvable technical difficulties. This group includes problems with identifying causality and the need to include deadweight, attribution, and displacement to fully understand causality. Temporal aspects like the use of linear return assumptions also pose challenges, as well as the lack of standardization. Too much standardization, however, inhibits the possibility of adjusting the SROI to the individual context and reduces the learning potential. The range of limitations also leads to concerns about the quality of the method. Another issue can be a lack of high-quality data and counterfactuals (New Philanthropy Capital, 2010), complicating the comparison of SROIs (Ryan & Lyne, 2008).

After examining how an evaluation design should be optimally structured and discussing available methods and tools, the next chapter examines methodological issues inherent in impact measurement.

4.6. Methodological Issues

COMs, impact measurement, and validation are subject to several methodological issues that should be considered. One key concept is CR, the challenge of attributing the outcomes to the services provided (McHugh et al., 2013; Social Finance, 2016). This is the risk of not knowing

what would have happened without the intervention (Social Finance, 2016). As a result, the actual effect of the SIB is hard to determine and could be over- or undervalued due to the possibility of external factors influencing the outcomes (Social Finance, 2016). Social Finance (2016, p. 13) identifies five factors affecting CR: availability of historical data on outcome, dependence of outcome on external events, strength of evidence base for the target group in the relevant context, scale of service provision or social issue, and duration of the impact bond. All the different factors can have a high or a low risk, influencing the overall CR. Social Finance also notes that properly evaluating the COMs with the proper methodology is key to reducing CR. Experimental evaluation designs like RCTs are beneficial methods. Counterfactuals without live control, like historical benchmarks, or even no counterfactuals, have no or only a minor positive influence on CR (Fox & Morris, 2019; Social Finance, 2016). Even though there is evidence that historical data can be a way to identify impacts, there is always the risk related to the selection of beneficiaries on the side of the SSPs, which can distort the outcomes and therefore also the evaluation (Hevenstone et al., 2023). Additionally, the benchmarks must be defined before implementing the SIB, to have a common baseline against which the intervention outcomes are measured (Becchetti et al., 2021). Empirical studies also show that counterfactual groups are hard to find and, therefore, not a common approach for impact measurement (Fraser, Tan, Lagarde & Mays, 2018).

Another methodological issue is the difficulty in defining and measuring the outcomes. The different actors must agree on what exactly to measure, by whom, and in what intervals (Fraser, Tan, Lagarde & Mays, 2018). Additionally, the general risk still holds that quantitative measures might overlook the underlying issues (Dayson, 2017). In their literature review, Andersen et al. (2020) identified several approaches with which SIBs try to tackle this methodological issue, which range from integrating more activity-based and outcome-based indicators to disconnecting social outcomes and the savings or focusing on the financials. There is generally a lack of research on how to collect and analyze data in social impact (O'Flynn & Barnett, 2017). Data quality is often subject to variability (Fox & Morris, 2019). Fox & Morris (2019) mention some possible ways to reduce the problem, i.e., by considering existing evidence, collecting primary data suitable for the purpose, reducing reliance on less fitting secondary data, and taking conceptual approaches like the theory of change into account. However, it is uncertain whether transferring knowledge between different SIBs is possible. In evaluating the 'Trailblazer' projects, the evaluators find that knowledge transfer is rarely possible (Tan et al., 2015).

Along with measuring outcomes comes the required time and associated costs. More reliable evaluation designs require more time and produce higher costs (Fox & Albertson, 2011). Overall, these costs should be proportional to the SIB's outcome payments, and more costly evaluations are, therefore, only useful for high-risk and unproven interventions where the attribution is less clear (Social Finance, 2016).

5. Methodology

The previous chapters have shown that SIBs are a complex instrument with advantages and disadvantages. They have significant potential upsides and could be a win-win situation for all actors if the challenges are adequately addressed. However, the current literature does not provide a holistic overview of payment structures, COMs, and impact measurement methods used in SIBs and how these factors relate to the success of the impact bonds. It is therefore helpful to investigate this further and consider additional potentially important factors like stakeholder collaboration. This chapter will explain the methodology for the empirical research of this study.

5.1. Research Design

This study adopts an explanatory mixed-methods design. On the one hand, a quantitative description of SIBs based on the INDIGO database was undertaken. On the other hand, a qualitative content analysis of selected final evaluation reports was conducted to enrich and contextualize the statistical patterns observed. The reports are a mix of qualitative and quantitative reports and, except for one intermediary process evaluation, assess the overall performance of the SIBs after the completion or towards the end of the intervention. The design directly addresses the two research questions introduced at the beginning of the thesis. The quantitative part identifies how COMs are defined and measured, and whether the SIBs achieved their targets. At the same time, the qualitative analysis allows me to consider underlying stakeholder perspectives, e.g., regarding the assessment of appropriateness of COMs or the perceived achievement of quantitative and/or qualitative outcomes. The data set consists of two complementary data sources – a SIB database for the quantitative part and an evaluation report corpus for the qualitative part. Each of these sources will be explained in detail in the following chapters.

5.2. SIB Database

5.2.1. Data Sample

The starting point for the SIB database is the INDIGO Impact Bond Dataset from the Oxford Government Outcomes Lab (INDIGO, 2025). At the time of the data retrieval (14th May 2025), the dataset encompassed 319 impact bonds. Several filters were applied to narrow down the number of relevant SIBs for this thesis.

First, it was decided to only look at completed bonds. Since this study also considers the achievement of targets, the advantage is that outcome data is more available and final, allowing a more reliable assessment of the success. Applying this filter reduced the number of impact bonds to 122.

Second, the database contains SIBs and DIBs. Given the thesis's focus on SIBs, DIBs were discarded from further analysis. Ten DIBs were identified among the completed impact bonds, reducing the remaining number of impact bonds to 112.

Third, it was decided to reduce the number of SIBs by focusing on SIBs commissioned in the UK. The UK is the birth country of SIBs and by far the country with the most commissioned and completed SIBs overall. Of the 112 completed SIBs in the INDIGO database, 73 are from the UK. The focus on the UK also allows for a better comparison of the results since potential cross-country differences are controlled for. Furthermore, the availability of relevant data is significantly better in the UK than in other countries since many reports are publicly available and the scope of the reports is also more extensive. This is especially relevant for gaining relevant insights from the content analysis.

As a fourth filter, it was decided to focus on only one policy sector to reduce the potential influence of different external factors and enable more robust inferences from the data. *Employment and Training* was chosen as the policy sector because it is by far the most frequently represented policy sector overall (see Figure 6) and in the UK alone. This filter reduced the total number of SIBs to 28.

The following two sub-chapters explain the variables of the SIB database's relevant output sheets and describe the data collection process in detail.[1]

[1] The entire SIB database is available as an Excel file, as it is too extensive for the appendix. The database includes all 319 impact bonds from the INDIGO database, but the manual data collection was only applied to the final sample. By applying the filters as described in the text, the final sample of 28 SIBs is visible.

5.2.2. Variables

The SIB database has five main sheets: *Completed SIBs, Deep-dive on COMs,* and *Report Overview*. These sheets contain the necessary information for subsequent data analysis. A complete list of all database variables, together with descriptions and the data source, can be found in Appendix A.

The sheet *Completed SIBs* contains information about the general characteristics of the relevant SIBs. Key variables from this sheet for further analysis are the *Number of COMs* per SIB, the *Payment Structure* of each SIB, the *Quantitative and Qualitative Evaluation Methodologies* used, and the *Success* of the individual SIBs. It is important to mention that success is seen as the result of quantitative evaluations like SROI or, if not available, as the achievement of COMs in this case. Some variables, like the *Country*, the *Bond Type,* or the *Stage of Development,* are needed to filter for the relevant SIBs. The table in the appendix shows that for most variables, the source is the INDIGO database. However, for some variables, the cells had to be filled manually by taking information from the evaluation reports or based on my reasoning. The affected variables from the *Completed SIBs* sheet are *Payment Structure*, *Intermediary*, *Qualitative Evaluation Methodologies*, *Quantitative Evaluation Methodologies*, *Success*, and *Success Info (Number of COMs achieved)*. It must be highlighted that, because of the reliance on evaluation reports, it was impossible to gather all the relevant information for all SIBs from the final sample due to missing reports (reports could only be found for 19 of the 28 SIBs) or missing information within existing reports. Chapter 5.2.3. will go into more detail on how the data was collected and which problems occurred.

The *Deep-dive on COMs* sheet encompasses information about the identified COMs within the individual SIBs. The most important variables in this sheet are the *COM Group* and *Validation Method Group*, i.e., self-defined overarching groups of COMs and validation methods, the *Payment for COM* and *Payment Structure*, as well as the *COM Achievement*. The validation methods state how the achievement of individual COMs was verified in the SIBs. Like in the previously described spreadsheet, there are several variables included that are needed for filtering or connecting the spreadsheets. The source of many variables is again the INDIGO database. However, a large share of variables had to be filled manually with the help of evaluation reports or my reasoning. The data collection process will be detailed in Chapter 5.2.3. The affected variables are *COM ID*, *COM Group*, *Validation Method Group*, *Payment for COM*, *Payment Structure*, *Outcome Payment Cap*, and *COM Achievement*.

The chosen variables across both spreadsheets provide a broad overview of important contractual structures of SIBs, thereby addressing the two research questions.

The sheet *Report Overview* provides information about the evaluation reports used for the database and the qualitative content analysis linked to the individual SIBs. The key variables are the *SIB_ID* and the *SIB Name* to connect the reports to the individual SIBs, as well as the *Link*, which helps to find the reports easily. Furthermore, the variables *Report Type* and *Report Language* provide more information about the reports and were used to filter out useless reports. The *Report Type* describes the type of a particular report, e.g., a final evaluation, an intermediary evaluation, or a case study. Lastly, there are variables for filtering the suitable SIBs, like *Country* and *Bond Type*. The information for the sheet was taken from the INDIGO database, the reports themselves, or the webpage where the report can be found.

5.2.3. Data Collection

As described above, the data sources were primarily the INDIGO database and the evaluation reports. The data from the INDIGO database could be retrieved as CSV files. These files were included in the SIB database and can be found in the database segment 'Input Sheets'. The evaluation reports were sometimes directly linked within the INDIGO database, and, if not, had to be searched manually. For UK SIBs, evaluation reports are often publicly available on governmental websites, e.g., from the funds that commissioned the SIBs.

The data collection process for the manually created variables is described below. If no information is available, it is marked as '*N/A'*; if information is not applicable for a certain SIB, it is marked as '*N/AP*.'

Payment Structure: The payment structure of SIBs is not explicitly stated in reports but must be inferred from how payments are made or how the COMs are designed. An example for the latter is the 'IPS employment support for people with drug and alcohol addictions' SIB. The final report does not contain any information about the payment structure. However, the COMs available on the INDIGO database are similarly defined to other COMs from the *Employment and Training* policy area, which all belong to a payment-per-outcome structure. Therefore, it was assumed that this SIB also has a payment-per-outcome structure. All the SIBs were categorized into one of the three payment structures introduced in Chapter 3.6.2.

Success: To fill this column, the evaluation reports were critically assessed concerning a stated financial success of the SIBs in the first place. However, the available information was

insufficient, so an alternative approach had to be chosen. One fund with ten commissioned SIBs had conducted an SROI analysis, thereby providing a quantitative measure to base the success decision on. An SROI > 1 was considered a success because the benefits outweigh the costs. For the remaining COMs without any quantified impact measurement, it was assumed that a COM achievement of more than 50% in payment-per-outcome SIBs indicates a successful SIB. Based on the available data, it is unlikely that all COMs are achieved in payment-per-outcome SIBs. Hence, this threshold was defined. Unfortunately, ten SIBs remained without any information on success.

Success Info (# of COMs achieved): This variable is calculated by adding all the COMs from the *COM Achievement* variable.

COM ID: This variable assigns a unique ID to each COM from each SIB. Even when two different SIBs have the same COMs, which is very common for SIBs that are part of a fund, each COM has a unique identifier.

COM Group: This variable was added to make the data and the different COMs used easier to analyze. The 28 SIBs from the final sample have 210 COMs, of which 74 are unique. This is because SIBs that are part of a fund typically all share the same COMs. Since many COMs refer to similar outcomes, grouping them to get a better overview and make the data more comparable was useful. Ultimately, the 74 unique COMs were grouped into 13 distinct COM groups. Table 1 gives an example of how the COMs were grouped. A complete list of all groups can be found in Appendix B.

Table 1: Example of COM Grouping.

Contractual Outcome Metrics	COM Group
13 weeks apprenticeship sustainment	Sustained Employment
13 weeks full-time Employment	
26 weeks full-time Employment	
13 weeks part-time Employment	
26 weeks part-time Employment	
Job sustainment (< 16 hours per week)	
Job sustainment (6 weeks)	
Job sustainment (min. 13 weeks at < 16 hours per week)	
Entry Into Work (min. 3 months at > 16 hours per week)	
Job sustainment (> 16 hours per week)	
Job sustainment (6 months)	
Job sustainment (min. 13 weeks at >16 hours per week)	

Source: Author's elaboration based on INDIGO (2025).

Validation Method Group: The grouping process was similar to that for the COM groups, and the reasoning behind adding this variable was the same. However, grouping the Validation methods was more difficult because many COMs mentioned multiple validation methods. The complete list of validation method groups can be found in Appendix C and contains 13 unique methods.

COM achieved: The information for this variable was mainly taken directly from the evaluation reports. For the SIBs 'Unlocking Potential (Career Connect)' and 'FutureImpact', the final reports did not have all the final numbers yet, but gave expectations. These expected numbers were used to decide on the achievement of the individual COMs. Furthermore, not all evaluation reports directly state whether the individual COMs were achieved. This is especially true for the IF pilot, which combines 10 different SIBs. The quantitative evaluation report only differentiates between the two rounds of commissioning but does not give information on the performance of the individual SIBs. Furthermore, it often only provides graphs that differentiate between three different cohorts of beneficiaries for each round and show each cohort's development over time. Every year, e.g., 2012/13 or 2013/14, represents one cohort. However, the graphs do not give any information about the statistical significance of the results and sometimes also indicate contradictory, i.e., positive and negative impacts of the SIBs, over time. For this reason, it is difficult to assess whether the COMs were achieved or not. Therefore, *N/A* was chosen for many of those COMs. These missing values are also why, for this variable, there are more cells filled with *N/A* than with *Yes* or *No*.

As indicated, not all variables could be filled with information for every SIB, particularly when the final report was unavailable. The variables suffering the most from missing data, besides *COM achieved*, are *Payment for COM*, *Outcome Payment Cap*, *Success*, and *Capital Raised*. Regarding *Capital Raised*, the problem is that the report for MHEP projects from the Social Outcomes Fund (SOF) and the Commissioning Better Outcomes Fund (CBO) only gives an overall contract value per fund. However, there is no breakdown on the level of the individual SIBs. This affects a total of 6 SIBs. A SIB that should be highlighted in terms of missing data is the *DFN-MoveForward*. Unfortunately, no final or interim report could be found for this SIB, so only the variables with data availability from the INDIGO database are filled.

5.3. Qualitative Content Analysis

The qualitative content analysis was conducted according to Mayring & Brunner's (2006) step model and based on final evaluation reports. Based on the theoretical part of the thesis and the

research questions, the content analysis focused on identifying themes related to contractual characteristics of the SIB, the measurement and evaluation, and the perceived success of the SIBs. The advantage of the content analysis over the SIB database is that it provides insights into the background of the SIB design and captures different perspectives from different stakeholders, e.g., regarding the perceived performance of the SIBs, the things that went well, or the problems that arose throughout the SIB lifecycle.

5.3.1. Evaluation Report Corpus

For the qualitative analysis, final reports for the final sample of 28 SIBs were searched for. The general advantage of final reports is that the results discussed in them are final. Out of those 28 SIBs, final reports could be identified for 19 SIBs. However, as it is common in the UK that SIBs are commissioned as part of larger funds and those funds ultimately publish the reports, the total number of unique reports is 6. One report is quantitative, two only used qualitative methods, and the remaining three used a mixed-methods approach. The reports' authors are university researchers in three instances and consultants in the other three. In one case, the university researchers wrote the report jointly with researchers from a not-for-profit social research organization. Most reports evaluated the impact or performance of the SIBs. One of the qualitative reports did this only qualitatively. The other qualitative report conducted a process evaluation, something that was also done by one report with a mixed-methods approach in addition to an impact evaluation. For two SIBs, the reports were already written before the final numbers were available, so the authors estimated numbers for the quantitative part. The length of the main parts of the reports ranges from 32 to 78 pages, excluding appendices, references, executive summaries, and forewords. Additionally, it was decided to include one interim report. The reasoning behind this is that this report contains a process evaluation of the early years of the MHEP SIBs, complementing the final report and adding valuable insights for the analysis. Furthermore, parts of this report can also be found in the final report. This report was written by consultants and has a length of 46 pages. The total number of pages that were qualitatively analyzed is therefore 378. Since the final sample only contains SIBs from the UK, all reports were in English, meaning no reports had to be discarded because of language barriers. Even though the total number of relevant reports seems low, they are considered highly relevant and rich in information, given their scope and length. A table of all included reports can be found in Appendix D.

5.3.2. Coding Agenda

The content analysis was done in MAXQDA, adopted a structuring analytical procedure, and therefore used the deductive category application (Mayring & Fenzl, 2019). The length of the reports drove my rationale for choosing this approach to ensure a focus on the key areas identified during the literature review. The coding unit was the complete sentence.

Table 2: Overview of the Codes for the Content Analysis.

First-level Codes	Second-Level Codes
SIB Design, Construction & Delivery	SIB Rationale & Framing
	Stakeholder Collaboration
	Payment Structure
	COM Definition & Appropriateness
	Delivery & Adaptations
Measurement & Evaluation	Evaluation Methodology
	Limitations & Constraints
Perception of Success	Financial/Outcome Achievement
	Qualitative Outcomes
	Contextual & Critical Reflections
	Unintended Effects & Risks
	Stakeholder Development
Design Recommendations	Partnership & Process Recommendations
	Metric & Contractual Recommendations
Basic Information	

Source: Author's elaboration.

I assembled a category system from the literature review. Subsequently, a coding agenda containing definitions, coding rules, and exemplary quotes was drafted. The coding agenda was used to code the entire corpus once. Afterwards, I refined the coding agenda since the originally defined codes and rules did not fit well enough. As part of the refinement, I added new codes, deleted others, and fine-tuned the definitions and coding rules because the content from the evaluation reports differed from the initial expectations. Table 2 gives an overview of all codes from the final coding agenda, and Table 3 shows an exemplary coding agenda for one second-level code. The entire coding agenda can be found in Appendix E. After refining the codes, the entire corpus of reports was re-coded. It was sometimes necessary to map one segment to more than one code because some sentences and paragraphs contained information relevant to different codes.

Table 3: Exemplary Coding Agenda.

Second-Level Code	Definition	Decision Rules	Example Quote
Payment Structure	Covers how payments are designed and triggered in SIBs, including outcome-based payments, block payments, rate cards, staged return mechanisms, and the distribution of financial risk among stakeholders.	**Included:** PbR models, rate cards, block vs. outcome payments, sliding scales, caps, staged returns, and rationale for design choices. **Excluded:** Operational or contextual challenges that are unrelated to the structure of payment mechanisms.	"The payment of these intermediate outcomes made the business proposition more attractive to social investors by enabling cash flow to be generated quickly, acknowledging the fact that employment was likely to be a long way off for many IF participants, so too distant and risky a prospect for many investors." (Griffiths et al., 2016, p. 26)

Source: Author's elaboration.

5.4. Data Analysis

The two different strands of research required different analysis approaches. I analyzed the SIB database using descriptive statistics, such as the ranges of payments for the same COMs or the shares of different validation methods used for validating COMs. An important part of the analysis was also to examine whether the achievement of COMs might be related to different design characteristics of SIBs.

For the qualitative content analysis, on the other hand, after coding the evaluation reports, I condensed the gathered insights for further analysis. For this purpose, I created a summary grid of the coded segments. This summary grid has one row for each code and one column for each evaluation report. For every intersection of codes and reports with associated coded segments, I wrote a summary of those segments to reduce the segments to a manageable size that can be further analyzed. I critically assessed the summaries for the results part of the thesis to identify important statements within individual reports and patterns across the reports. These insights were then enriched with illustrative quotes.

5.5. Methodological Limitations

The methodological approach is subject to several limitations that should be highlighted. First, there are data gaps in the SIB database, especially for SIBs where the final reports were unavailable. This means that the findings for the incomplete variables will suffer from the lack of data due to fewer data points. Second, as already discussed above, some variables were

manually created. Even though the specific process is explained in detail for transparency purposes, there is still a certain extent of subjectivity involved and room for errors. This is especially true for the variables related to the payment structure, the COM groups, and the success of the SIBs. Third, the coding in the content analysis was only conducted by me. Therefore, there is a risk of a single-coder bias. It is not possible to completely resolve this issue, but to reduce the risk, the entire corpus of reports was coded twice.

After explaining the methodology of the empirical part of the study, the next chapter will examine the analysis's findings before discussing them.

6. Results

Given this study's mixed-methods approach, this chapter will examine the findings from the SIB database first before proceeding to the content analysis.

6.1. Findings from the SIB Database

The SIB database offers valuable information about the sampled SIBs. First, structural data about the SIBs will be provided before exploring the COMs and SIB evaluation methods in more depth. Ultimately, the SIB success and COM achievement will be investigated, including relationships to structural SIB aspects.

6.1.1. Structural Data of Relevant SIBs

Overall, looking at the database, it becomes apparent that the SIBs are relatively homogeneous in many basic characteristics. They all have the same payment structure, i.e., payment-per-outcome, are affiliated with at least one fund, and all but one have an intermediary (see also Appendix F). The only SIB without an intermediary is the Prevista SIB from the Youth Engagement Fund (YEF). The evaluation report revealed that this SIB was generally special in its set-up compared to the other three SIBs from the YEF. Unlike the other SIBs, this project was only partly run as a SIB, with only 4 out of 12 SSPs using upfront capital while the remaining SSPs operated with a regular PbR contract (Ronicle & Smith, 2020).

Figure 9 shows the intermediaries that are part of the SIBs and how often they participated. With 15 participations, Social Finance is by far the intermediary with the most involvement, nearly three times as much as the next intermediary, MHEP. It should be noted that some SIBs have more than one intermediary, so that the total number does not add up to 28, which is the number of distinct SIBs from the final sample.

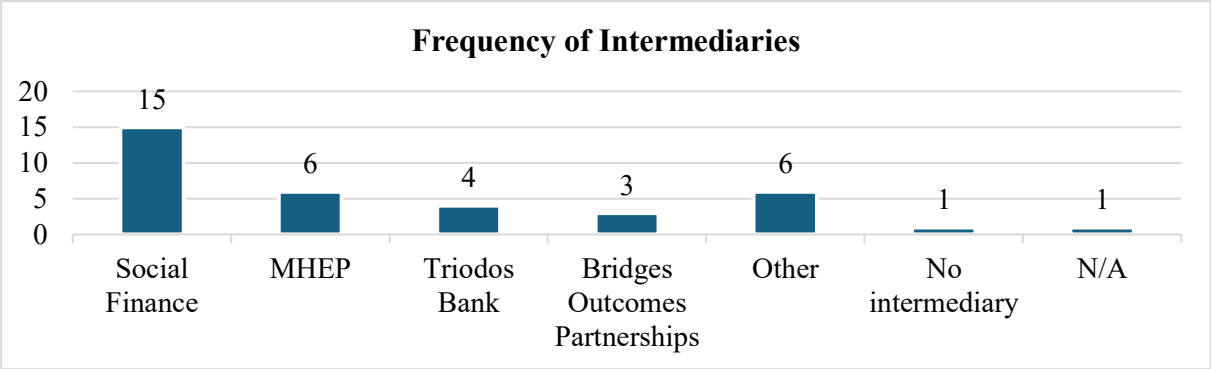


Figure 9: Frequency of Intermediaries.
Source: Author's elaboration based on INDIGO (2025).

Figure 10 gives an overview of the funds that commissioned the different SIBs. Five different funds commissioned all the SIBs from the policy area *Employment & Training* in the UK. Regarding total SIBs commissioned, the IF is the largest with ten SIBs, and the SOF and the YEF are the smallest with four commissioned SIBs each. As can be seen, the commissioning of SIBs through funds is very common in the UK. Often, several SIBs are commissioned simultaneously. The IF commissioned its ten SIBs over two rounds within the same year – the first round had six SIBs and the second round four. A special case are the MHEP SIBs because the first two rounds were funded by the SOF and the CBO together, while the Life Chances Fund funded the third round. However, the third round has not yet been completed, and an evaluation report is available only for the first round of three SIBs.

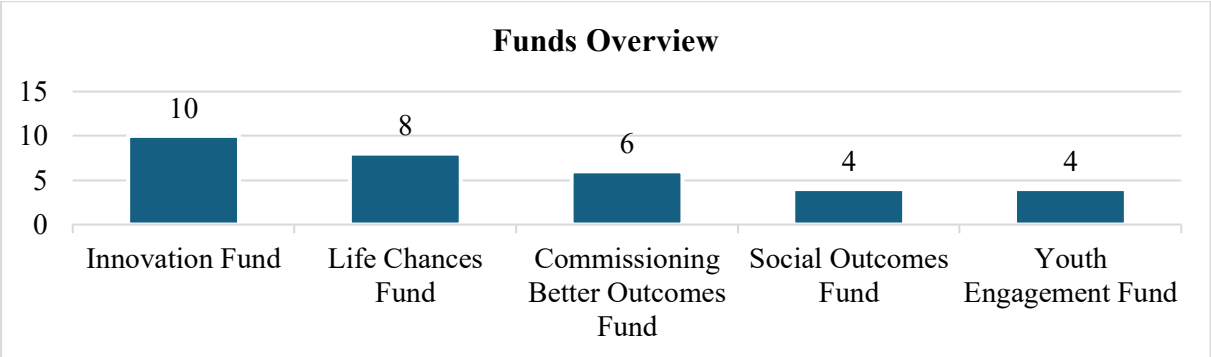


Figure 10: Fund Overview.
Source: Author's elaboration based on INDIGO (2025).

6.1.2. Insights on Contractual Outcome Metrics

COMs are important in SIBs because they define the payment trigger. The number of defined COMs varies considerably, as shown in Figure 11. The minimum number of COMs defined is four, while the maximum number is 19. However, while not being a statistical outlier, there is only one SIB with 19 COMs, and the second-highest number is 11. The figure also shows that the average number of COMs is 7.54, which is much closer to the minimum than to the

maximum and therefore stresses the exceptionally high number of COMs in the one project. 75% of all projects have between 4 and 10 COMs.

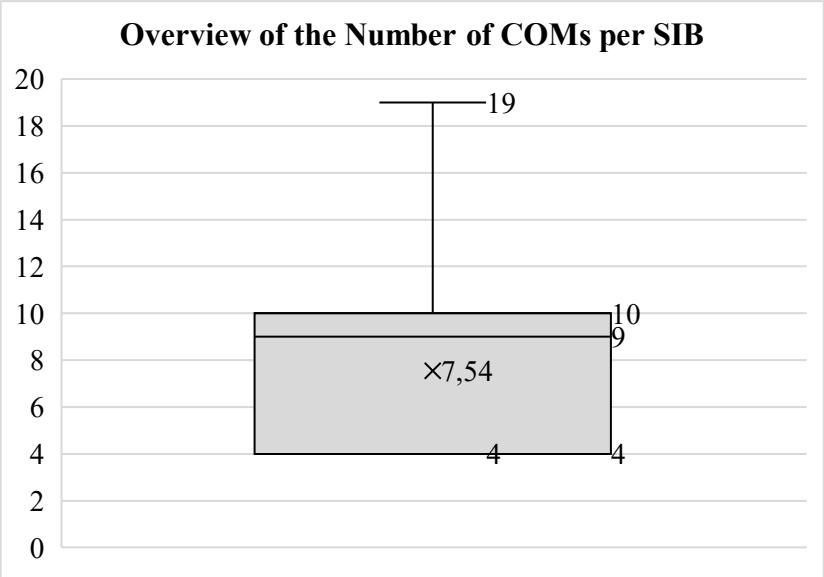


Figure 11: Overview of the Number of COMs per SIB.
Source: Author's elaboration based on INDIGO (2025).

Considering which COMs are frequently used in SIBs for *Employment & Training*, it is unsurprising that qualification and employment-related COMs dominate (see Figure 12). The two largest COM groups, *QCF Qualification* and *Sustained Employment,* comprise 57% of all COMs. Together with *Entry to Employment,* the qualification and employment-related COMs make up nearly two-thirds of all COMs used in the SIBs. Interestingly, the third largest COM group is not directly related to reaching qualification targets or employment, but is related to softer outcomes, i.e., behaviour, attendance, and attitude in school. This resembles the idea that the defined COMs are seen as gradual stages, and it is reasonable to assume that improved behaviour or attendance in school will lead to more positive qualification and employment outcomes later.

When looking at the specific COM definitions, it becomes clear that all COMs are binary, i.e., either they are achieved or not, and do not measure gradual improvements, except for one SIB that conducted wellbeing assessments. An example of the binary outcomes used is whether or not a particular qualification level is achieved. The definition of *School Attendance* is interesting because the beneficiaries have to pass a certain threshold based on the average attendance of the average student for the COM to be achieved. That means the threshold is irrespective of the starting point of the beneficiaries. Furthermore, *School Behavior* and *School*

Attitude are somewhat subjective COMs, because the improvements have to be noted and confirmed by the teachers or home tutors of the beneficiaries.

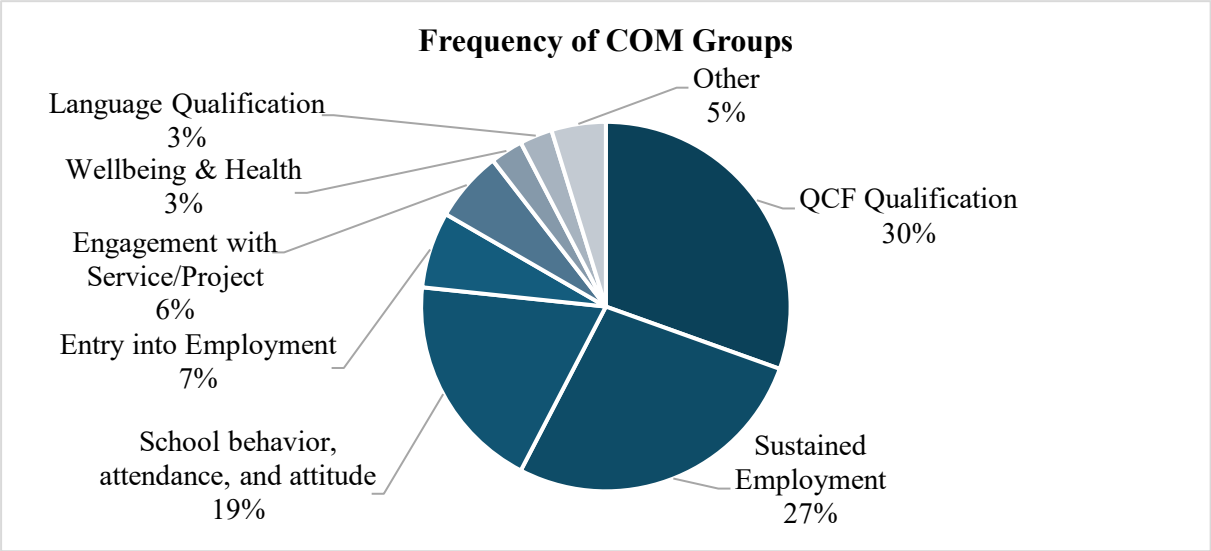


Figure 12: Frequency of COM Groups.
Source: Author's elaboration based on INDIGO (2025).

It is common in payment-per-outcome SIBs that different COMs are associated with different payments. Table 4 provides an overview of minimum, average, and maximum payments for all COM subgroups[2] for which the data were available. It becomes clear that the differences between the minimum and maximum payments can be very high. Three extreme examples for this are the subgroups *Engagement with Service/Project*, *Sustained Employment (< 16 hours/week),* and *Sustained Employment (>16 hours/week)*. The maximum payment in these groups is roughly five times higher than the minimum payment. Conversely, subgroups like *Improved School Attendance* and *QCF Entry-Level* have a tiny difference, indicating a consensus among projects and funds that this payment amount is sufficient. The subgroups *Basic Skills*, *Language Qualification*, and *QCF Level 4* only have one payment amount and were used in more than one project, but all of these SIBs were commissioned by one fund, which distorts the picture. It is also interesting to see the payment for *QCF Level 4* compared to the other QCF qualification groups. The average payment increases from entry-level to level 3, which intuitively makes sense because achieving a level 3 qualification is generally harder than an entry-level qualification. Of course, this also depends on the individual's starting point. However, for *QCF Level 4*, the payment is lower than for QCF *Level 2*. Similarly, the average

[2] The subgroups shown here are based on the COM groups shown in Figure 12, but are more detailed and differentiate, e.g., between the different QCF qualification levels.

payment for sustained employment ≥ 26 weeks is lower than for ≤ 13 weeks, even though sustained employment of at least 26 weeks is arguably harder to obtain.

Table 4: Overview of COM Payments.

COM Subgroup	Frequency[3]	Min. Payment (in £)	Avg. Payment (in £)	Max. Payment (in £)
Basic Skills	4 (4)	900	900	900
Engagement with Service/Project	13 (8)	414	810	2164
Entry into Employment	14 (5)	1145	2580	4569
Improved School Attendance	14 (14)	1300	1357	1400
Improved School Attitude	12 (12)	400	500	700
Improved School Behaviour	14 (14)	800	1029	1300
Language Qualification	6 (6)	1200	1200	1200
QCF Entry-Level	12 (12)	900	967	1000
QCF Level 1	15 (14)	700	1043	1500
QCF Level 2	15 (14)	2200	3000	3900
QCF Level 3	14 (14)	2700	3643	5100
QCF Level 4	6 (6)	2000	2000	2000
Sustained Employment (< 16 hours/week)	5 (5)	1993	4339	9655
Sustained Employment (> 16 hours/week)	5 (5)	2302	5315	10862
Sustained Employment (≤ 13 weeks payment trigger)	25 (17)	1025	2722	3500
Sustained Employment (≥ 26 weeks payment trigger)	22 (17)	1000	1540	2000

Source: Author's elaboration based on evaluation reports.

The methods to validate or verify the achievement of COMs are very diverse. As can be seen in Figure 13, there are many different ways to evidence a COM achievement. Often, the SSPs can choose between different options, e.g., a confirmation letter or a copy of a certificate. Nevertheless, looking at the graph, it becomes evident that confirmation letters and copies of certificates are by far the most mentioned validation methods, combining a share of more than 50%. The ex-post verification, which is also mentioned in 15% of the COMs, does not specify which evidence is needed and is therefore not a validation method like the others shown in the graph. It could be that using any other methods to verify the outcome achievement is possible. The ex-post verification entries were taken from the INDIGO database, and it was impossible to specify the methods due to a lack of data.

[3] The numbers in brackets indicate for how many of the COMs the payment data is available and therefore how many different data points were used to populate the payment data columns.

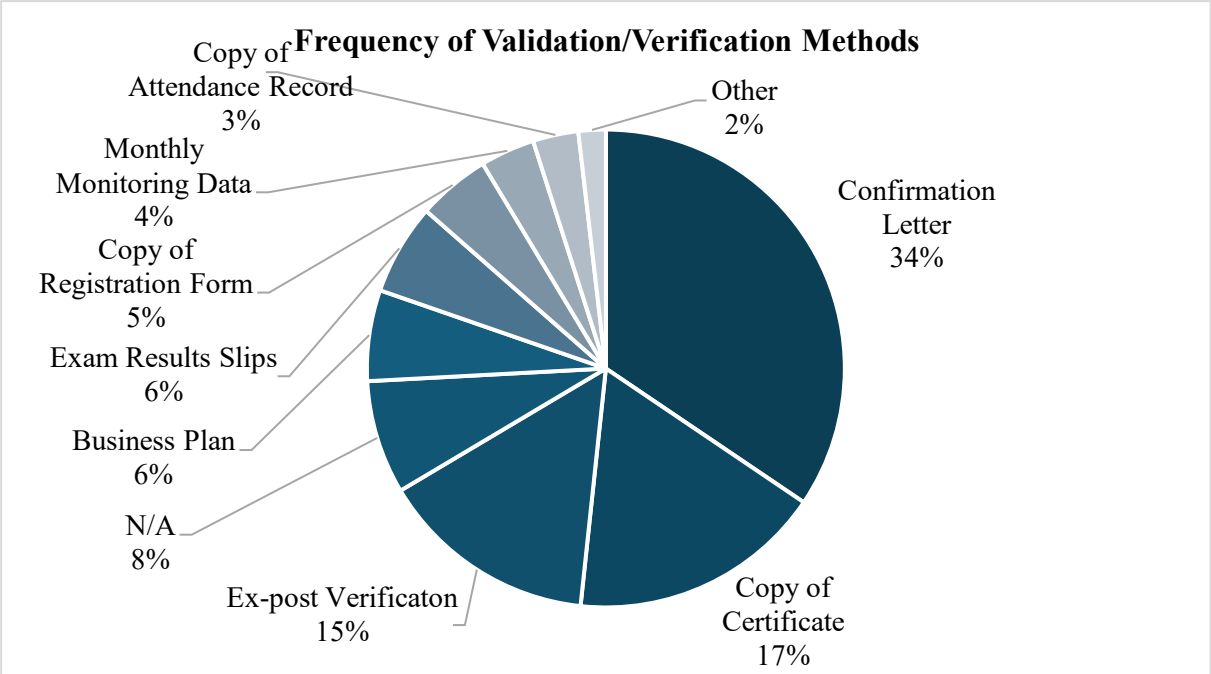


Figure 13: Frequency of Validation/Verification Methods.
Source: Author's elaboration.

Overall, it can be seen that UK SIBs from the policy area *Employment & Training* use mainly similar COMs and validation methods. The results are also driven by the fact that many SIBs are commissioned as part of larger funds, and the funds use the identical rate cards, i.e., COMs and associated payments, for each SIB within the fund. The number of COMs used can differ substantially, as can the payment for individual COMs. However, the influence of a higher payment on the achievement of a COM seems to be low, as can be seen in Chapter 6.1.4.

6.1.3. SIB Evaluation Methods

The investigated evaluation reports employ various qualitative and quantitative methods to evaluate SIBs as part of process or success evaluations. The latter are used to assess the overall impact of the SIBs. Figures 14 and 15 provide an overview of the different methods used.

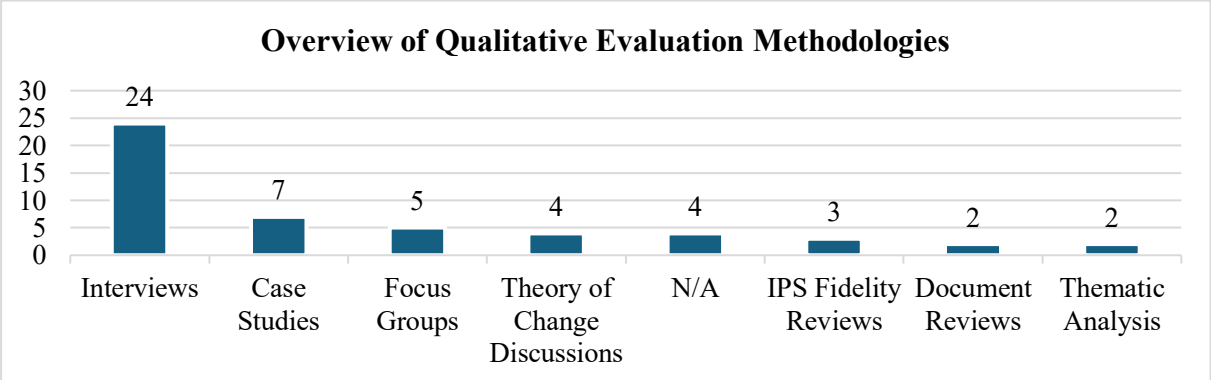


Figure 14: Overview of Qualitative Evaluation Methodologies.
Source: Author's elaboration based on evaluation reports.

Interviews are the most important method for qualitative evaluation. They were used in every qualitative SIB evaluation. Case studies and focus groups are the following most popular methodologies, but they were only used in 25% of the SIBs or fewer. For four of the 28 SIBs, it is unknown if a qualitative evaluation was conducted.

Performance data assessments were the most common method for quantitative evaluations, followed by SROI analyses and surveys. However, this result is again distorted by the fact that the IF, with its 10 SIBs, employed all three evaluation methods; four SIBs did not conduct any quantitative evaluation due to issues with data access, and for four other SIBs, it is not known if they did this kind of evaluation. This means that the most used evaluation methods for the remaining 10 SIBs of interest were still performance data assessment, totaling 7.

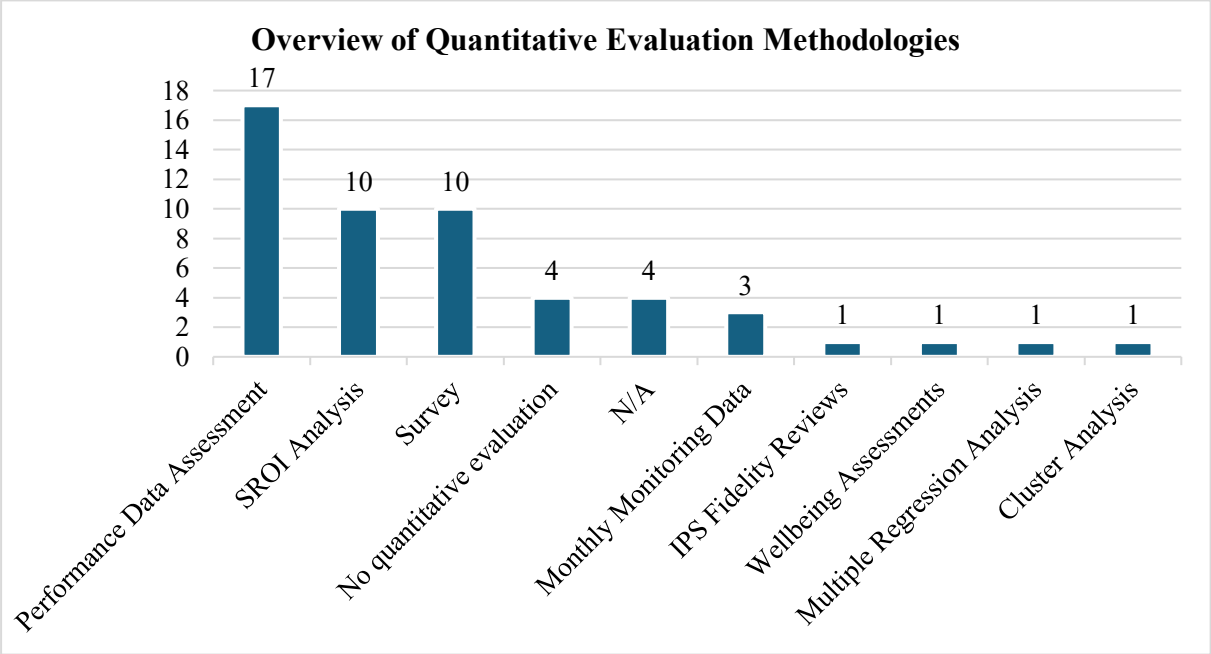


Figure 15: Overview of Quantitative Evaluation Methodologies.
Source: Author's elaboration based on evaluation reports.

Notably, no SIB used an RCT for evaluation. This evaluation type is more complex but leads to more robust results due to randomized control groups. The IF SIBs tried to identify matching comparison groups using Propensity Score Matching (PSM), but this is not as robust as randomized control groups.

A few methods dominate both qualitative and quantitative evaluations, but there are theoretically various options to choose from. As evaluation reports often employed several evaluation methodologies, it was impossible to detect whether specific methods are superior to others.

6.1.4. SIB Success & Achievement of Contractual Outcome Metrics

Ultimately, the success of SIBs and, consequently, the achievement of COMs in the first place is the most important factor for most stakeholders involved. Half of the sampled SIBs were successful (see also Figure 16). However, it must be kept in mind that this column of the dataset was filled manually, and the results should therefore be treated cautiously. Additionally, the SROI of the IF SIBs, which was used as a proxy for success, was very low for all SIBs, but larger than one (Salis et al., 2018). For over a third of the SIBs, success is unknown, highlighting the lack of data on this important matter.

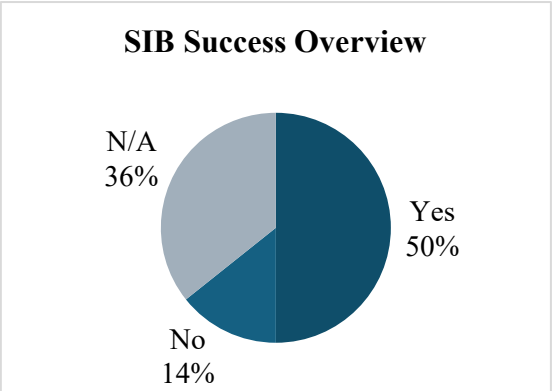


Figure 16: SIB Success Overview.
Source: Author's elaboration.

Moving on from the level of success on the SIB level to the COM level, Figure 17 shows that for more than 50% of the COMs, it is unknown whether they have been achieved. This, in turn, affects the analysis of the relationships between structural SIB characteristics, like validation methods used, and COM achievement. For the remaining COMs with available data, there were slightly more unachieved COMs (25% of the total) than achieved COMs (19% of the total).

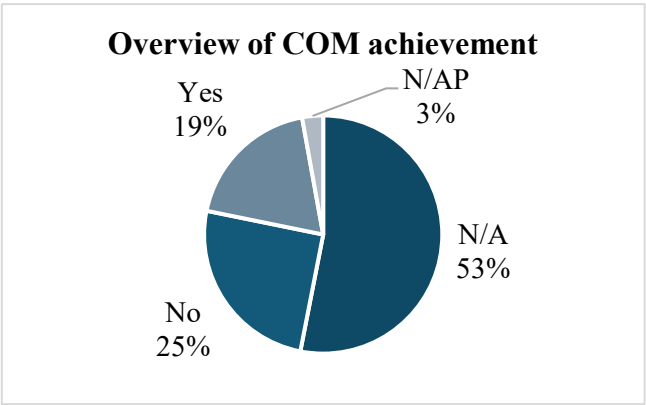


Figure 17: Overview of COM Achievement.
Source: Author's elaboration based on evaluation reports.

Figure 18 shows the achievement of COMs by the different COM groups. It is visible that for most groups, only up to five data points on the achievement of COMs are available, meaning that the robustness of the findings is low. The groups that stand out in this regard are *QCF Qualification*, *School Behaviour, Attendance, and Attitude*, and *Sustained Employment*, which also reflects that those COMs make up roughly 75% of all COMs used in the sample of this study (see also Figure 12). For five COM groups, the overall picture is positive, with more achieved than unachieved COMs. Three COM groups show the opposite results. This is especially true for *School behaviour, attendance, and attitude,* as well as *Sustained Employment,* where the difference between achieved and unachieved COMs is very high.

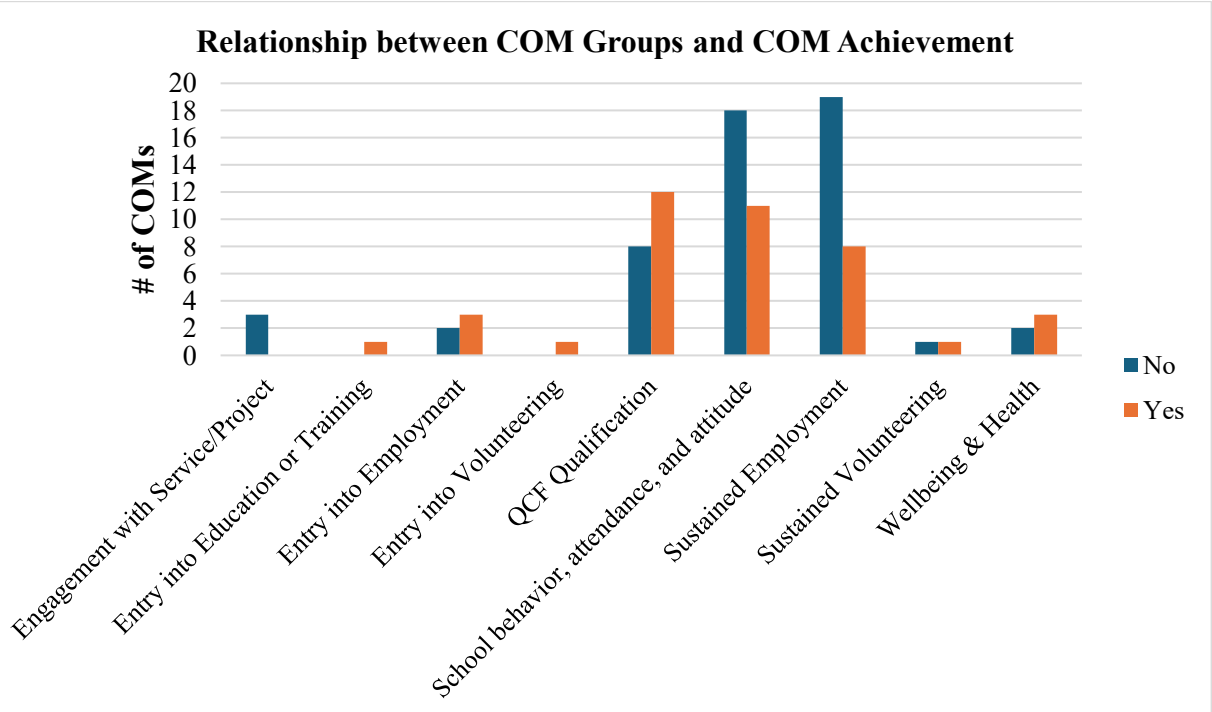


Figure 18: COM Groups and COM Achievement.
Source: Author's elaboration.

A view of the subgroup level reveals that, especially, the subgroups *Improved School Attendance* and *Sustained Employment (≤ 13 weeks payment trigger)* stand out negatively with a difference between achieved and unachieved COMs of 10 or more (see also Appendix G). Interestingly, the two remaining subgroups of *School Behaviour, Attendance, and Attitude* show positive or neutral results. In fact, *Improved School Attitude* is the most successful subgroup, along with *QCF Entry-Level*. For *QCF Qualification,* the overall picture is positive, with three of the five subgroups achieving a positive result.

Overall, the lack of data makes it difficult to make definite inferences about proper or less useful COMs regarding the ultimate achievement of the targets. The best overall impression makes

QCF Qualification. However, the results are probably also context dependent. This will be further investigated in the discussion.

Lastly, it is interesting to examine the relationship between COM payment and COM achievement, given that there can be considerable gaps between the minimum and maximum payments within the same COM subgroup.

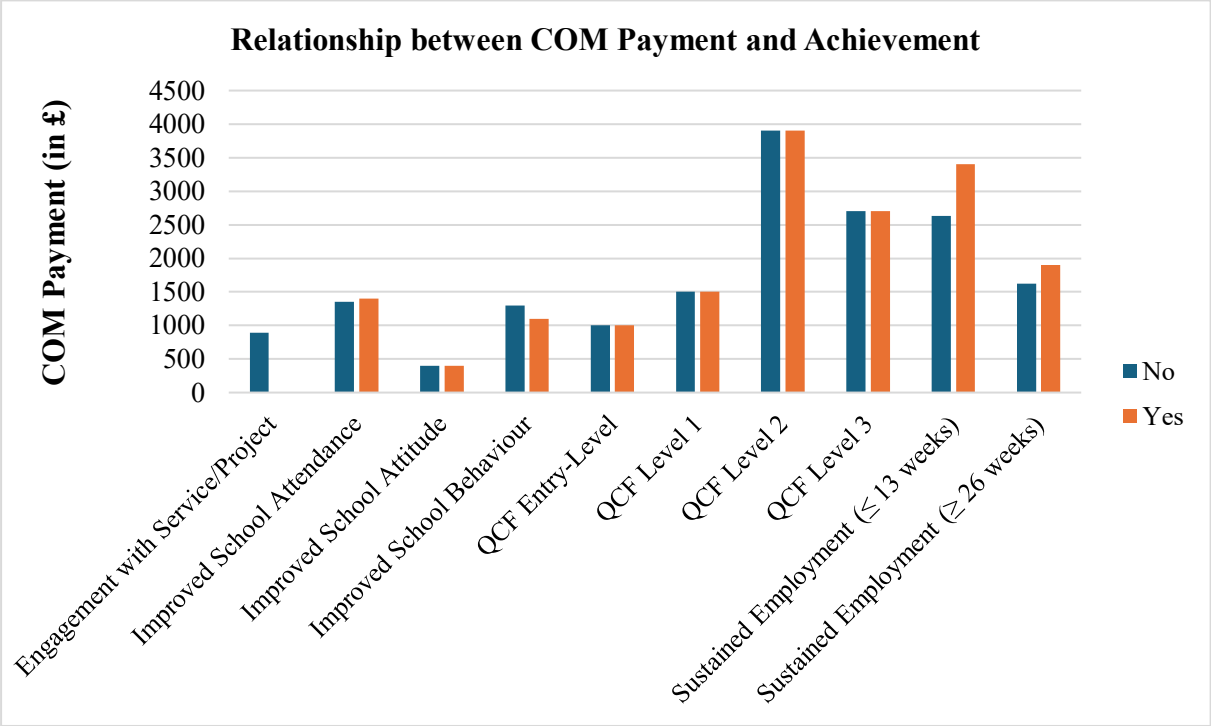


Figure 19: COM Payment and Achievement.
Source: Author's elaboration.

However, Figure 19 does not show significant potential impacts of the payment on the achievement of COMs. Only for *Sustained Employment*, the achieved COMs have a notably higher payment than the unachieved COMs. Given that *Sustained Employment (≤ 13 weeks)* was only achieved once and unachieved 15 times (see Figure 20), the perceived impact is distorted by a higher payment for the one achieved COM. The same distortions are also true for *Improved School Attendance* and, to a lesser extent, *Sustained Employment (≥ 26 weeks)*, which also exhibit slightly higher payments for the achieved COMs.

Overall, it seems there is limited evidence that the COM group or the payment structure are significantly related to the achievement of COMs. Specific COM groups might be better suited to be achieved than others, but the lack of data prevents definite inferences and makes it necessary to interpret the results with caution. It should also be noted that the results do not show any statistical significance.

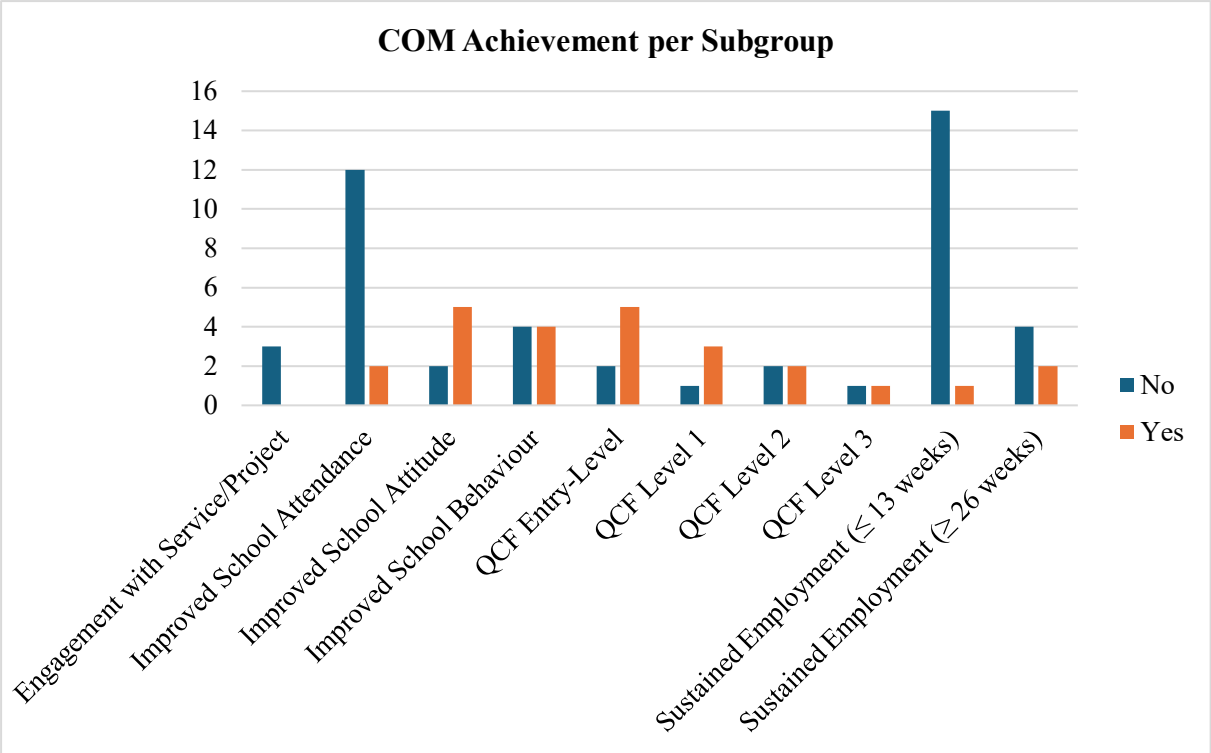


Figure 20: COM Achievement per Subgroup[4].
Source: Author's elaboration.

6.2. Findings from the Evaluation Report Analysis

This chapter summarizes the insights gained from analyzing the evaluation reports. Subchapters are provided for all the main codes defined in the coding agenda. The code *Basic Information* will not be covered here because these coded segments deal with general information about the projects and do not provide valuable insights for the study's goal.

6.2.1. SIB Design, Construction & Delivery

The SIB design is central to the success of an SIB, and setting up an SIB involves plenty of negotiations between various stakeholders, such as investors, commissioners, and government agencies. This subchapter summarizes findings from the reports about the underlying reasons for the SIBs, payment structures, and COM designs, as well as challenges and mitigations.

6.2.1.1.SIB Rationale & Framing

The rationale for using a SIB design is similar across the different funds and projects. All the funds (CBO, YEF, IF) essentially used the SIBs as test beds to determine if PbR contracts

[4] The graph only looks at the subgroups and COMs that were also included in figure 19 to keep the insights comparable. The data therefore differs from the one in Appendix G.

achieve better results than conventional fee-for-service contracts and whether they lead to fiscal savings.

> Test the extent to which a PbR approach involving social investors can drive improved outcomes for young people and generate benefit savings, as well as other wider fiscal and social benefits. (Ronicle & Smith, 2020, p. 11)

However, it must be mentioned that the savings were not always a central argument for the SIB structure.

> MHEP was expected to generate savings, both to DWP (…) and to local commissioners (. …) We do not, however, view it as a project which was primarily driven by an invest to save logic (. …) We think this because local commissioners do not appear to have been motivated by the promise of savings to engage with MHEP, and in two of the three sites the commissioners already had conventional IPS contracts in place. (Stanworth & ATQ Consultants, 2023, p. 37)

Another goal of using the SIB structure was to strengthen the social investment market and the capacity of SSPs. MHEP in particular wanted to combine national and local budgets to scale IPS services.

> Support the development of the social investment market, build the capacity of social sector organisations and contribute to the evidence base for SIBs. (Ronicle & Smith, 2020, p. 11)

> One of MHEPs core objectives was to combine sustainable sources of funding to grow the provision of IPS services (. …) Part of this process involved bringing together new sources of funding, combining national funding (…) and local health and Local Authority funding (…) to provide additional funds to IPS services. (Gadenne et al., 2020, p. 33)

6.2.1.2. Stakeholder Collaboration

Robust collaboration is repeatedly portrayed as the underlying force keeping SIBs on track. All reports emphasized the role of a clear governance structure, often housed in a SPV, and the critical coordinating, mediating, and performance monitoring function of intermediaries.

> Each SIB was managed through a Special Purpose Vehicle (SPV). (…) In the context of YEF, the SPV held the contracts with all the different parties, and provided intermediary support, including monitoring the performance of the service provider. (Ronicle & Smith, 2020, p. 16)

> For most stakeholders the primary function of intermediary action was to facilitate good communications amongst all bodies involved (…) a key role for intermediaries was to ensure and foster such commitment. (Griffiths et al., 2016, p. 38)

However, while the intermediary's role was appreciated mainly, some stakeholders questioned the high intermediary fees as an avoidable cost and suggested that a different setup could reduce the financial burden.

> In the view of several delivery managers, the high fees involved (…) had placed an additional financial burden on projects which they had not initially costed for. Some believed these costs could be reduced or avoided. (Griffiths et al., 2016, p. 37)

The depth of relationships between various parties was consistently identified as fundamental for effective collaboration and outcomes. Strong relationships and clear, shared (social) objectives were deemed critical to the success of the SIB model and a strength of SIBs, particularly during the early stages of the projects.

> The Central Government stakeholders reported that the SIB enabled good collaboration at the design stage, as it enabled departments to coalesce around a shared interest in the same outcomes. (Ronicle & Smith, 2020, p. 54)

> The genuine commitment to making the IF projects a success and to achieving social impact over and above making a commercial return on investment (…) had been something of a revelation to other stakeholders. It was seen as an important element in the development of strong and effective partnerships, and a reason in some projects for greatly improved relations between partner organisations as the programme progressed. (Griffiths et al., 2016, p. 29)

> It proves that rate cards are a good way of bringing departments together around a shared common interest…It feels to me that's a big take home – I think I'm increasingly of the view that where SIBs are quite powerful is the ability to coalesce people around a set of key interests in an area where that hasn't quite existed before. (Ronicle & Smith, 2020, p. 54)

However, the alignment of interests is not always as straightforward as this might suggest and in the YEF projects, problems occurred especially at later stages of the projects.

> This level of collaboration did not continue, however, during the delivery stage. (Ronicle & Smith, 2020, p. 54)

The sharing and utilization of performance data for monitoring, learning, and accountability were consistent across projects, serving as the standard language for partnerships and driving performance. However, the level of data analysis was unfamiliar for many SSPs, and data sharing and access to IT programs were challenges for some SIBs.

> The requirement for tight management of projects through the use of performance data was seen as axiomatic by investors and intermediaries but this 'granular' level of data collection and analysis was widely acknowledged as being new territory for most providers. (Griffiths et al., 2016, p. 29)

> In three of the SIBs the service providers would typically report to the board of investors on a monthly basis, including submitting latest performance results and forecasts. (Ronicle & Smith, 2020, p. 53)

> The ease to which this information could be ascertained from schools was to some extent depended on the presence of a local authority Service Level Agreement for annual return data. (ConnectMore Solutions & Richmond Baxter Ltd, 2023, p. 28)

Common friction points included misaligned priorities between stakeholders (e.g., providers prioritizing job outcomes, commissioners prioritizing mental health outcomes), rigid contract management by commissioners, and some stakeholders feeling marginalized.

> While the collaboration between the two parties was described as a positive, it transpired that their objectives were not perfectly aligned. Providers felt that the focus of the IPS service was on the delivery of a high-fidelity support service, which requires job outcomes (…), while the commissioners described their focus to be more on delivering positive mental health outcomes. (…) The relationships between the provider, MHEP and the commissioner were affected by this lack of alignment. (Gadenne et al, 2020, p. 41)

> It's felt like, as co-commissioner, we're not really round the table. (Ronicle & Smith, 2020, p. 46)

In conclusion, the evaluation reports consistently demonstrate that clear governance structures, shared interests, and consistent communication influence effective stakeholder collaboration in SIBs. While the overall picture of stakeholder collaboration across the SIBs seems positive, substantial variation exists. These differences were often caused by the mix of partners involved, the specific delivery set-up, and early performance trajectories of the SIBs.

6.2.1.3.Payment Structure

The exact payment structure differs between the SIBs at times. However, all SIBs, except for one, for which no information is available, use a payment-per-outcome structure with a rate card that defines different payments for various COMs. Given that some SIBs, like the MHEP SIBs, aimed to combine local and national funding, the payments were made by local commissioners and government funds (SOF, CBO).

In the MHEP SIBs, the financial risk distribution caused by the payment structure was extensively discussed. Initially, the MHEP model placed significant financial exposure on providers, resulting in cash-flow uncertainty and pressure to meet targets.

> The most important consequence of the design of MHEP 1 was that did not insulate the VCSE providers sufficiently from financial risk. At the low performance levels achieved this led to a downward spiral of additional pressure, increasing staff turnover and the creation of further undershooting of performance targets. (Stanworth & ATQ Consultants, 2023, p. 69)

Counterintuitively, the investor received the scheduled interest and a surplus despite lagging service performance.

> This imbalance of risk is further demonstrated by the fact that the social investor – BII – continued to receive interest payments on its loan capital and has made a small return on MHEP 1. (Stanworth & ATQ Consultants, 2023, p. 60)

Consequently, the block payments for the providers were increased, which is also reflected in subsequent MHEP contracts.

> However, we found over time that paying on engagements didn't reduce the risk for providers sufficiently and created a misalignment between the goal of the programmes (jobs) and what we were paying on. Therefore, over time, we have shifted provider payments towards block

> payments. In most of our more recent contracts, the provider receives 75-90% of their payments on block. (Stanworth & ATQ Consultants, 2023, p. 51)

The MHEP SIBs were not the only ones undertaking adjustments to their payment structure. The FutureImpact SIB introduced fixed payments after the outbreak of Covid, while the IF, which used a 100% PbR structure, changed the payments defined in the rate card for the SIBs of the second round of commissioning. This change seemed successful, since the Round Two projects did not need as much restructuring as the projects from the first round. These adjustments reflect learnings and reactions to changing circumstances.

> There was strong evidence of learning, with the Steering Group flexing programme design and the contract to better serve young people. Notable examples include (…) the introduction of fixed payments during Covid. (ConnectMore Solutions & Richmond Baxter Ltd, 2023, p. 27)

> The scale of payments for all outcomes was also increased for Round Two projects, arguably making it easier for these later projects to sustain cash flow. (Griffiths et al. 2016, p. 40)

The 100% PbR payment structure with staged payments employed in the IF SIBs was chosen because it is more appealing to investors due to more constant cash flows. Other stakeholders generally also positively evaluated it because it incentivizes better performance.

> The payment of these intermediate outcomes made the business proposition more attractive to social investors by enabling cash flow to be generated quickly. (Griffiths et al., 2016, p. 26)

> The funding model appeared to have created a high intensity of focus on performance across nearly all projects and PbR was widely seen as having incentivised better performance. (Griffiths et al., 2016, p. 30)

In the YEF, in addition to the payment cap, there was also a volume cap for every COM. The specialty of this cap was that the SSPs had to state the expected volume for every COM already during the bidding stage. This structure had detrimental effects, especially related to service provision.

> Service providers felt that the initial announcements from Central Government around the total number of young people the programme was going to support, coupled with the amount service providers could claim per outcome, inhibited their ability to engage very vulnerable young people in the programme. (Ronicle & Smith, 2020, p. 45)

In conclusion, the reports highlight the intricate nature of designing effective SIB payment structures. While initial models often aimed to transfer risk to providers to encourage PbR capability, this frequently led to significant cash flow challenges and misaligned incentives, ultimately resulting in adaptation to existing and subsequent SIB contracts. Additionally, early, lower-value outcomes proved crucial for generating initial cash flow in pure PbR environments.

6.2.1.4.COM Definition & Appropriateness

The design of the rate cards and definition of the COMs were influenced by the payment-per-outcome model structure, which needed to ensure sufficient cash flows and appropriate risks for the stakeholders involved. The YEF report states that this structure is necessary because reaching ultimate goals takes a long time.

> In recognition of the fact that it could take some years to track the final employment outcomes for young people, the YEF paid for a series of 'proxy' intermediate outcome measures. (Ronicle & Smith, 2020, p. 16)

However, as with the payment structure, the defined COMs have also caused problems during the intervention. The payment structure and the COMs are two sides of the same coin, and together, they form the rate card. Therefore, they are closely linked. The core challenge identified across the SIBs is the difficulty in designing and implementing COMs that are financially viable and genuinely reflect social impact.

A major problem of many COMs across the SIBs was that they put financial strain on providers due to unrealistic targets and timelines, resulting in low achievement rates and cash flow problems.

> The performance targets set for the provider turned out to be far above what was achievable. These initial performance expectations (…) were set by the commissioner based on previous performance under the existing, conventional contract, as reported by Making Space. (Stanworth & ATQ Consultants, 2023, p. 40)

> The sensitivity of the model to both under- and overestimating the number of outcomes that could be achieved, and the time required to achieve them, caused early and in some cases, ongoing, difficulties for several providers. (…) It further became apparent that within the model, the timing of predictions and achievements was critical to cash flow and the smooth running of projects. (Griffiths et al., 2016, p. 38)

These shortcomings made it necessary to reprofile and restructure the COMs, often by adding new intermediate metrics that better reflect the client's journey and the intended end goal. For the MHEP SIBs, these changes were undertaken during the projects' implementation, while for the IF SIBs, the adaptations happened in Round Two of the commissioned projects.

> MHEP starting to pay Making Space per user entering employment. (…) It meant that providers were paid for job starts (irrespective of duration) and thus further eased provider cashflow. (Stanworth & ATQ Consultants, 2023, p. 39)

> The outcomes matrix (…) was thus seen by many as needing revision if it was to accommodate the realities of working with this particular group of NEET young people. These included recognition of other key stages (…) such as entry-level and Level 1 qualifications, work experience and voluntary work placements. (Griffiths et al., 2016, p. 36)

On the other hand, some SSPs from YEF SIBs commented that the COM design lacked flexibility, referring to the inability to change pre-determined numbers of outcome achievements (see also Chapter 6.2.1.3 above).

Many providers also felt that the defined COMs did not capture all the support provided and the impact achieved through the interventions, especially lacking softer outcomes like confidence or motivation of the beneficiaries.

> The use of job outcomes as a proxy for assessing the performance of the service as a whole also had an impact on implementation. (…) Providers felt that the focus on job outcomes to calculate outcome-based payments was not representative of the entirety of the support they provided. (Gadenne et al.,2020, p. 42)

> The recognition that the strictly defined outcomes that are eligible for funding under the IF programme did not encapsulate the full depth and diversity both of individual and of broader social impact being achieved. (Griffiths et al., 2016, p. 30)

Lastly, SSPs also raised the issue that they felt some of the defined COMs were outside their control, perceiving it as unfair to attach payments to them.

> The intervention can only have a small impact on Level 2 qualifications; the largest influencer on Level 2 results are schools (. …) Therefore, the projects argued that their success against this metric was reliant on the effectiveness of the schools, and was to a degree beyond their control. (Ronicle & Smith, 2020, p. 39)

Despite a largely negative emphasis on the defined COMs being retrieved from the reports, the overall picture and perception of the stakeholders might not be as bad. The YEF report also notes that

> Overall, the investors and service providers were pleased with how the SIBs had been designed, mainly in terms of the rate card. (Ronicle & Smith, 2020, p. 44)

6.2.1.5.Delivery & Adaptations of Intervention

Being flexible and adapting SIB structures to changing circumstances or identified flaws is important, as can already be seen in the previous chapters about the payment structure and the COM design. The reports also reflect the advantages of adaptation, which is seen as a key to success for SIBs.

> The analysis of these two case studies shows how the flexibility of the SIB model, achieved through a proactive approach to collaboration, positively affected the delivery of IPS services. (Gadenne et al., 2020, p. 46)

> This said, the flexibility of providers to reprofile and reorientate projects was a key part of the success of the pilot programme. (Griffiths et al., 2016, p. 93)

The YEF report also portrays the SIB structure as an enabler of changes.

> Although many stakeholders reported that there was limited flexibility on the macro-level programme design and outcome projections, there was scope to be flexible in more micro-level elements of the programme, and the SIB facilitated this. This was as a result of the regular meetings with the investors, coupled with the greater focus on outcomes. (Ronicle & Smith, 2020, p. 63)

It is interesting to see the connection between the IF SIBs and the YEF SIBs regarding their stand on flexibility. On the one hand, the IF SIBs, which were commissioned first, had a very open approach towards adaptations and had very positive experiences with them.

> In addition to this and in recognition that the IF was operating in territory new to most of those involved, the programme was run very much as an 'active experiment', with considerable flexibility given by DWP commissioners to projects to enable them to re-design interventions, restructure delivery plans and implement operational modifications in response to changing circumstances and live experiences. (Griffiths et al., 2016, p. 39)

> After restructuring and reprofiling (…) all projects expected to exceed, reach or come close to meeting their targeted outcomes. (Griffiths et al., 2016, p. 92)

On the other hand, the YEF-commissioned SIBs, which were commissioned later and could already learn from the IF pilots, followed a different, more rigid approach regarding flexibility. The designers had their reasons for it, e.g., to avoid a potential de-risking of projects to the government's detriment, but it was not welcomed by many stakeholders, and evidence indicates that more flexibility might have led to better outcomes.

> Projects strongly disliked DWP's stance on contract variations. DWP reported that there could only be limited variations within the projects for a number of reasons. Firstly, they were constrained by public sector commercial procurement rules and regulations around variations within contracts (. …) Secondly, DWP's stance stems from the idea that testing whether an approach worked or not is easier when not changing things whilst in delivery. (Ronicle & Smith, 2020, p. 47)

> A key element of the SIB is risk transfer from government to the investors (. …) Central Government stakeholders felt there was too much flexibility in IF, which enabled the projects to de-risk, and essentially shift the risk back from the investor to government. (Ronicle & Smith, 2020, p. 47)

> However, there is evidence in the evaluation that delivery would have been greater had DWP accommodated more flexibility. (Ronicle & Smith, 2020, p. 48)

6.2.2. Measurement & Evaluation

The measurement and evaluation of SIBs are important to verify their success regarding achieved COMs and the overall impact generated. This subchapter looks into the qualitative insights into the evaluation methodologies used, in addition to insights from the quantitative research strand, and into the limitations and constraints of the approaches.

6.2.2.1.Evaluation Methodology

The analysis of the SIB database already showed that most SIBs employed a mixed-methods approach for evaluation, combining quantitative data, like administrative records, with qualitative data, especially from interviews.

While some evaluations, like the ones for the IF SIBs, attempted more rigorous quasi-experimental designs using Propensity Score Matching (PSM) to create comparison groups and account for selection bias, this was an exception and not comparable to more robust methods like an RCT.

> The comparison group was obtained by implementing the Propensity Score Matching (PSM) methodology, which entailed pairing (or 'matching') each IF participant with the most similar non-participant young person. (Salis et al., 2018, p. 15)
>
> Being able to compare IF participants to a refined, similar group of non-participants was key to building up a picture of how the IF had an impact on programme participants. (Salis et al., 2018, p. 18)
>
> In terms of its methodology, this evaluation is not able to draw upon a randomised controlled trial (RCT) programme design nor is it able to draw upon data of a comparable group of clients not served by an IPS service from which to build a statistically matched counterfactual. (Whitworth & Cullen, 2023, p. 18)

In the case of the MHEP SIBs, the CBO fund also criticized a missing rigorous evaluation design that would have allowed conclusions to be drawn on the SIBs' actual effect and facilitated learning.

> The CBO team also expressed some disappointment that there hadn't been more learning from the project, and in particular a robust local impact evaluation had not been commissioned. In their view while the BIT evaluation contains some useful insights, it does not provide strong conclusions about the effectiveness of MHEP either in its own right or compared to alternatives. (Stanworth & ATQ Consultants, 2023, p. 56)

When looking at the verification of COMs, the IF report notes differences in the effort needed for the SSPs to evidence claimed outcomes.

> Compared to job entry and sustained employment, which formed the key outcomes for young people over the age of 16, outcomes based on behaviour and qualifications were relatively easy to claim on a 'one-off' basis and did not require further evidencing at future points in time. This contrasted with job outcomes which had to be evidenced twice, once when the job started and again when it was sustained at 26 weeks. (Griffiths et al., 2016, p. 33)

Overall, the reports do not provide extensive information on the reasoning or thought processes behind the decision to use specific evaluation or verification methods. They demonstrate a consistent effort by the stakeholders to understand SIBs through various data. However, the methods used are subject to severe limitations, as can be seen in the following subchapter.

6.2.2.2.Limitations & Constraints

The SIBs faced three main issues regarding measurement and evaluation – difficulty proving causal impact, data limitations and quality issues, and a high administrative burden.

Most evaluations struggled to prove causation between the services offered and the outcomes, primarily due to the absence of robust experimental designs like RCTs or because the evaluations were conducted too early to see lasting impacts. Without these designs, testing the perceived casualties mentioned by some of the stakeholders in the interviews is impossible.

> It has not, however, clearly demonstrated that by doing so it has achieved better or more outcomes than could have been achieved by IPS delivered through a different mechanism, in part because the project was not designed with a counterfactual measure that could prove additionality and attribution. (Stanworth & ATQ Consultants, 2023, p. 67)

> The above calculations assume that all the positive outcomes recorded may be attributed to the programme. In other words, they imply that such outcomes would not have occurred in the absence of the intervention. (Salis et al., 2018, p. 65)

> Elsewhere a direct impact on levels of NEET-ness was more difficult to claim, mainly because such an impact will only become apparent in the longer term. (Griffiths et al., 2016, p. 74)

> The perceived improvements in attitude and behaviour were not always easy to measure with precision, but subject teachers would report on positive changes and many senior school staff said they had seen evident impacts and effects. (Griffiths et al., 2016, p. 84)

Some evaluators faced incomplete, missing, or poor-quality data. This included issues like unrecorded early exits, truncated participant records, and data linkage problems that shrunk sample sizes. In one case, the data complications meant a quantitative impact evaluation was impossible.

> The reason for young people exiting the programme early was not captured, a missed opportunity to learn. (ConnectMore Solutions & Richmond Baxter Ltd, 2023, p. 27)

> The commissioner also reported that the data set available to commissioners was narrowly focused on the contractual outcomes. This meant that Haringey's commissioners found it harder to establish whether the project had met their wider social objectives, such as widening equality of access. (Stanworth & ATQ Consultants, 2023, 50)

> Due to unforeseen complications with the data collected during the programme and changes in data sharing regulations an impact evaluation was not possible. (Ronicle & Smith, 2020, p. 18)

> Given some missing data (…) many of the individual categories (…) have relatively small sub-group sample sizes. As a result, it will inevitably be challenging (…) for the estimated odds ratios of categories (…) to be statistically significant. (Whitworth & Cullen, 2023, p. 16)

Lastly, measuring and evaluating the SIBs imposed a significant administrative and reporting burden on projects. This involved complex financial and outcomes reporting, duplicate data requests, and substantial staff time and project spending diverted from direct service delivery.

In some cases, this administrative burden also negatively affected the relationships with educational institutions.

> MHEP stated in their application to SOF/CBO that the establishment of a comparison group (…) would be explored 'subject to funding and feasibility' but this appears to have been ruled out on cost grounds. MHEP also explored the establishment of a comparison group as part of its local evaluation (…) but again appears to have ruled it out on cost grounds. (Stanworth & ATQ Consultants, 2023, p. 36)

> A significant cost is accounted for by the need to meet SOF/CBO reporting requirements rather than working with commissioner and providers and reporting to the MHEP board. (Stanworth & ATQ Consultants, 2023, p. 58)

> Whilst the additional data collection had positive side-effects, it also had negative side effects. The service providers felt the volume and detail of evidence required (…) was disproportionate. Service providers reported that this resulted in them having to ask for large amounts of data from education institutions (…). This created an additional administrative burden for educational intuitions, and some service providers felt this damaged their long-term relationship with these schools. (Ronicle & Smith, 2020, p. 58)

> The tasks involved had often been extensive and time-consuming, in particular ensuring up-to-date and appropriate data were available, developing tools for the analysis and understanding of the data gathered, and reporting ongoing situations and results back to investors. (Griffiths et al., 2016, p. 37)

To sum up, evaluating the SIBs is subject to several limitations, and most reports highlight that the findings stated should be interpreted with caution, particularly due to attribution problems or data issues. Furthermore, it becomes apparent that the requirements significantly burden SSPs, impacting their ability to focus on the interventions.

6.2.3. Perception of Success

The success evaluation in the SIB database was primarily focused on quantitative measures and the achievement of COMs. Since the respective cells had to be filled manually based on my own reasoning, there is room for errors. At times, it was difficult to state whether a project was successful, given a lack of data and due to the assumption made that a COM achievement of at least 50% equals success. Therefore, given that a main focus of the thesis is the success of the SIBs, it is also important to analyze the reports in that regard. During the analysis, I found that the perceived success can substantially deviate from the stated success based on hard outcomes and numbers.

6.2.3.1. Financial/Outcome Achievement

Financial success and, closely related to it, the achievement of COMs, are of utmost importance for SIB, because they are linked to payments and therefore needed to secure the financial viability of the projects.

The evaluation reports indicate mixed success of the SIBs. The FutureImpact SIB achieved full investor repayment, including a surplus, and overdelivered on many hard outcomes – even continuing service delivery after reaching contractual caps. Education and training, as well as work-related outcomes, were especially successful. However, it also has to be noted that not all hard outcomes could be achieved.

> The strong performance by Futures enabled them to repay the principal to SASC with interest. Futures delivered the programme within budget (. …) The contract generated value added for the commissioners as Futures' continued delivery after individual outcome caps had been reached. (ConnectMore Solutions & Richmond Baxter Ltd, 2023, p. 34)

> Claimed income by outcome was made available as at end March 2023. At that point nearly 40% came from education & training outcomes (. …) Similarly, work-related income (29%) reflected a high rates and good performance (. …) Volunteering claims were below 4%, underlining the difficulties in this area. (ConnectMore Solutions & Richmond Baxter Ltd, 2023, p. 35)

Similarly, the IF SIBs were widely viewed as a financial and outcome achievement success, even though this is primarily due to mid-program restructurings. Also, the SROI ratios were larger than 1 for every project. However, compared to ratios from other SROI evaluations, the results were very low, and considering additionality would probably lead to ratios below 1.

> All IF projects successfully operated for the full three years and expected, as a minimum, to repay investments to their social investors. Several generated a clear financial return for investors. After restructuring and reprofiling (…) all projects expected to exceed, reach or come close to meeting their targeted outcomes (or had indeed already done so). (Griffiths et al., 2016, p. 92)

> Given the results of the impact study imply few positive outcomes, and relatively small additionality from even those positive outcomes, estimates of SROI and benefit-cost ratios in general will be below 1. (Salis et al., 2018, p. 65)

On the negative side, many SIBs, particularly from MHEP and the YEF, consistently fell short of engagement and job-outcome targets, impacting payments and savings. Despite the overall positive judgement of the IF SIBs, the quantitative evaluation found that outcomes for SIB participants were sometimes worse than for comparison groups, e.g., negative impacts on school behaviour through increased absence and exclusions in school.

> As this shows performance in terms of both total cohort referred to the intervention and total users engaged was well below plan (. …) This mattered especially in relation to the user engagement target because user engagement was a key driver of payments to service providers. (Stanworth & ATQ Consultants, 2023, p. 43)

> Since user engagement was below plan it is not surprising that these subsequent outcome targets were also not achieved. (Stanworth & ATQ Consultants, 2023, p. 45)

> Basically, this finding suggests a negative impact of the IF pilot in that it indicates that, following participation in the programme, the proportion of young people attending school/college decreased. (Salis et al., 2018, p. 29)

A key reason for under-performance was unrealistic or over-ambitious targets. Additionally, lower-level qualifications were often achieved, but providers struggled with higher-level qualifications or employment outcomes.

> A major reason for the shortfalls across all three contracts appears to have been that the original forecasts were set high and have been described by stakeholders, especially commissioners, as 'unrealistic'. (Stanworth & ATQ Consultants, 2023, p. 44)

> The findings of this evaluation indicated that the IF pilot was beneficial to participants, in that it was found to have supported them to achieve NQF level 1 qualifications. However, there was also evidence suggesting that the proportion of participants achieving higher level qualifications was reduced as a result of the programme. (Salis et al., 2018, p. 69)

However, the problem of achieving more lower-level outcomes is not true for every project, as the FutureImpact SIB demonstrates. Here, the picture is the other way around, due to the participants' different starting points.

> The programme helped fewer than expected intro entry level qualifications, and more than expected at a higher level, demonstrating the barriers faced by young people were often social and emotional, not academic. (ConnectMore Solutions & Richmond Baxter Ltd, 2023, p. 38)

The WDP SIB evaluation showed that several factors, such as recent work history, service intensity, or ethnicity, significantly influenced job-start success, highlighting potential equity issues.

> As expected, a client's likelihood of job entry consistently decreases the longer the time since their last period of paid work, controlling for other factors (. …) A consistent finding in terms of client ethnicity is that all ethnic groups are around twice as likely on average to achieve a job start compared to clients of Black ethnicity, controlling for other factors. (Whitworth & Cullen, 2023, p. 17)

Some stakeholders also questioned whether SIBs truly created new outcomes or just improved the evidence of existing ones, meaning that the outcomes already existed but were not evidenced.

> Because of the improved monitoring the SIB did lead to more outcomes being evidenced, but their gut feeling was these outcomes were already being achieved, but just not being evidenced. (Ronicle & Smith, 2020, p. 69)

The central takeaway is that SIBs have a highly mixed and often challenging record of achieving their contracted outcomes and financial viability. While some excel, many struggle with ambitious targets and the complexity of serving the most vulnerable, leading to underdelivery and financial shortfalls.

6.2.3.2.Qualitative Outcomes

Apart from the COMs and the connected financial outcomes, the SIBs also achieved many qualitative outcomes, which are not visible directly from the SIB database. Many stakeholders, especially SSPs, emphasized these qualitative outcomes.

Across all SIBs, beneficiaries experienced substantial increases in confidence, self-esteem, and self-belief as well as improved emotional regulation, reduced anxiety, and better overall mental health.

> Confidence was the most commonly reported benefit. (ConnectMore Solutions & Richmond Baxter Ltd, 2023, p. 14)
>
> Service users also described feeling increased confidence, a newly found sense of independence, or reduced anxiety as a result of receiving this support. (Stanworth & ATQ Consultants, 2023, p. 53)
>
> The young people involved demonstrated improvements in their mental health and emotional wellbeing, primarily through having someone non-judgemental to talk to, but also as a result of being able to better manage their emotions. (Ronicle & Smith, 2020, p. 42)
>
> This attention to wellbeing had a positive impact on clients with reports of increased confidence, self-esteem and self-respect. (Whitworth & Cullen, 2023, p. 34)

Furthermore, especially in youth-focused SIBs, there were tangible improvements in behavior and a more positive attitude towards learning and life.

> There's definitely less young people leaving school at 16 and becoming NEET in the schools that we support as compared to neighbouring schools. (Griffiths et al., 2016, p. 76)

The participants also often reported improved relationships with their families, reduced conflict, and expanded social networks, which contributed to a greater sense of social inclusion.

> Improved social interaction and relationships were also a common theme in the research. Interviewees frequently talked about how they were much better at communication, particularly when in situations with strangers; (…) Others found that their improved management of their behaviour and emotions had impacted positively on their relationships with their parents. Finally, a number of participants noted that being involved with the project had extended their friendship groups. (Ronicle & Smith, 2020, p. 43)
>
> Although not outcomes that attracted payment in the IF programme, improved social functioning and family relations – often based on increased confidence and self-esteem – were seen by delivery staff as fundamental to long-term sustained positive change. (Griffiths et al., 2016, p. 80)

Regarding their professional future, SIBs broadened participants' horizons by showing them new possibilities, helping them articulate career paths, gain the required skills needed for their envisioned future, and generally be better prepared for the future.

> This was a common theme; participants in one focus group talked about how they were now considering new options for their future, such as entering a career in construction (or indeed, moving on to further education). (Ronicle & Smith, 2020, p. 38)

> That they had a stronger idea of what they wanted to do and a clear focus on what qualifications they would require to get there was also a common experience of participation. (Griffiths et al., 2016, p. 77)

Unfortunately, the qualitative outcomes were mostly unrelated to payments, even though they are highly valued by most stakeholders. The missing payments also meant that the support leading to these outcomes often had to be subsidized by other funding.

> I have seen case studies of young people who could not leave their bedroom and now come out and have communication. They have had progress on confidence and mental health but not so much paid outcomes. It is still impactful but for these cohorts the coach's time is being subsidised by other parts of the project. (Ronicle & Smith, 2020, p. 26)

Sometimes, the findings from qualitative and quantitative evaluations were contradictory. For example, in the case of the IF, the quantitative evaluation indicated negative impacts like increased school absences, while the qualitative report highlighted positive behavioral changes. It is, therefore, important to consider both financial success or COM achievement and qualitative outcomes for the success evaluation of SIBs.

The core takeaway is that SIBs consistently delivered significant, positive qualitative outcomes for beneficiaries, which are widely recognized as meaningful and important for their future success. However, these outcomes are often not reflected in payment structures and, therefore, do not automatically result in financial success.

6.2.3.3. Contextual & Critical Reflections

The previous two subchapters showed that while SIBs frequently foster positive qualitative outcomes for beneficiaries, quantitative outcomes often fall short. One reason is that success is heavily mediated by external factors such as economic conditions, policy shifts, and partner performance. These mediators also make it more complicated to attribute the outcomes to the interventions, particularly without a counterfactual.

> The largest influencer on Level 2 results are schools, as they spend substantially more time working with young people on educational progress than the projects. (Ronicle & Smith, 2020, p. 39)

> Over the period of the pilots, the policy and labour market context also changed due to the raising of the statutory school leaving age and the fact that there are virtually no job opportunities any longer for 16 and 17-year-olds. (Griffiths et al., 2016, p. 35)

While the SIB mechanism definitely has advantages compared to more traditional contracts, it also comes with considerable administrative burden and management overhead, leading to higher costs. Therefore, the question arises whether the overall value for money still justifies using the SIB model, especially given that some commissioners were more attracted by the external funding opportunities coming with the SIBs rather than the SIB mechanism itself.

> As it stands, the impact evaluation implies that most of the scheme's positive outcomes would have been achieved even if the scheme had not existed, and in some cases that the results from the statistical analysis suggests that participants in the programme achieved poorer outcomes than the comparison group. Moreover, the kinds of outcomes where the IF may have made a positive difference (e.g., lower level qualifications) do not provide particularly high rates of return either to society or to individuals. (Salis et al., 2018, p. 65)

> Haringey commissioner stakeholders saw a trade-off between the costs of negotiating and setting up a SIB and its operational overhead, and the benefits of pro-active contract performance monitoring and support that the SIB funded. In their view the trade-off might be marginal given the extra work that commissioners have to do to agree contracts, but it is a "question of what are you prepared to do" to secure additional funding. (Stanworth & ATQ Consultants, 2023, p. 50)

> The SIB led to a broad range of effects, some positive and some negative. Overall, stakeholders thought that the positives outweighed the negatives, and that the SIB, overall, improved service design, management and delivery. However, the ultimate indicator of whether the SIB was successful must surely be whether it achieved more outcomes than an alternative funding mechanism. Here the evidence is unclear. (Ronicle & Smith, 2020, p. 75)

Another critique brought forward by several projects is that the measured impact underestimated the actual impact of the projects.

> Indeed, several projects made the point that they felt they were achieving greater social impact than was being recorded by paid outcomes, both in terms of 'softer' measures (…) and in terms of outcomes achieved that could not be claimed for due to contractual restrictions. (Griffiths et al., 2016, p. 30)

Overall, many stakeholders generally expressed positive views on the SIBs, and many even intended to pursue further SIBs in the future. However, a recognized need was to redesign SIBs and implement lessons learnt.

6.2.3.4.Unintended Effects & Risks

While all stakeholders aimed for SIB designs that avoid unintended effects, this was not entirely achieved, as the evaluation reports show. However, it should be noted that only the reports from the IF and the YEF mention unintended effects.

One significant issue in some of the YEF SIBs was the cherry-picking of beneficiaries, resulting from the models' financial pressures and the need to achieve outcomes. Picking beneficiaries with a higher chance of success made it more likely and easier to achieve claimable outcomes

in the available timeframes. However, it led to under-serving the hardest-to-reach, which was the SIBs' actual aim.

> All four projects argued that it was unrealistic to expect disadvantaged young people to achieve some of the outcome triggers (. …) This disincentivised the projects from supporting young people who were unlikely to achieve these outcomes. (Ronicle & Smith, 2020, p. 26)

> [Cherry picking] was evident in three of the SIBs (…) some service providers had deliberately avoided accepting harder to reach young people, because they found they disengaged with the service and so the service provider could not claim outcomes for them. (Ronicle & Smith, 2020, p. 61)

In contrast to the YEF SIBs, cherry-picking did not seem to be a problem in the IF SIBs. However, the report highlighted the important role of schools in recruiting beneficiaries. While there were some pre-defined eligibility criteria, some schools refused to refer the most disadvantaged people. At the same time, others referred too complicated people for whom the SSPs did not have enough resources to support them adequately.

> There was no suggestion from any quarter that any projects were 'cherry picking' by choosing to recruit young people who presented them with the quickest prospect of an early outcome; programme eligibility criteria (…) and the social commitment of all partners (…) helped to ensure this. It was rather the school environment that played most part in determining the types and characteristics of young people recruited to projects. (Griffiths et al., 2016, p. 90)

> In virtually all projects, therefore, the very hardest-to-help young people who were not attending school and unlikely to achieve an outcome within the timeframe of the programme, did not generally find their way onto the programme. (Griffiths et al., 2016, p. 66)

Besides cherry-picking, the YEF report notes several other adverse, unintended effects. The evaluators identified behaviours of parking, i.e., reducing the support for beneficiaries who had already triggered payments, a narrow focus on payment-linked goals, and a provision of services to generate payments, without having the best interests of the beneficiaries in mind.

> [Parking] was evident in one SIB, in which they had stipulated in their model that they would support young people for a set period of time. For some young people they had already achieved the outcomes they were working towards, but were unable to exit the young person from support (. …) Consequently they were 'parked'; they remained on the programme but were only provided with light-touch, or no, support. (Ronicle & Smith, 2020, p. 61)

> This was evident in three SIBs. The outcomes-focus culture only extended to the outcomes that payments were attached to, and in some instances these were prioritised at the expense of other, secondary outcomes. (Ronicle & Smith, 2020, p. 62)

> In one SIB, the service provider was struggling to achieve its number of entry-level qualifications, as the schools were reluctant to release children from classes. In response, the provider designed its own course that they could deliver outside of school time (. …) However, (…) this qualification was not being provided to young people because there was a genuine belief that young people would benefit from this qualification; instead, it was primarily provided to hit the targets. (Ronicle & Smith, 2020, p. 63)

The evaluation of the IF SIBs found that the SSPs felt that the project structure diluted the quality of the services being offered, and some also expressed a mission drift as a result of the structure.

> The main concern of frontline workers was what high caseloads could reduce the quality of support they were able to deliver. (Griffiths et al., 2016, p. 42)

> The voluntary and community sector organisations most strongly identifying their 'mission' as helping the most disadvantaged and hardest-to-help young people believed that, in the need to generate cash flow quickly, extending support to larger numbers of school-aged participants had diverted them away from this mission and core target group where much of the social impact was to be gained. (Griffiths et al., 2016, p. 88)

Lastly, the reports suggest that the SIBs have systemic design flaws that drive risks and cause these behaviors. The SIB design should provide incentives to focus on the most disadvantaged people. For that, it is necessary that the payments are high enough to cover the costs, that targets are realistic, and that long-term outcomes are incentivized appropriately and cannot simply be replaced by achieving more easy-to-reach outcomes.

> If projects are able to make themselves financially viable simply by increasing the number of early outcomes achieved, there will be no incentive for them to strive for later or more risky outcomes more strongly predictive of the desired policy goal. (Griffiths et al., 2016, p. 93)

> Whilst YEF has demonstrated that provision can be delivered with these costs, the costs cannot cover the delivery of support for very disadvantaged young people. (Ronicle & Smith, 2020, p. 25)

> As mentioned in the social motivations of stakeholders, when the original outcome projections are inaccurate this fuels the perverse incentives, as projects seek to secure the outcome payments. (Ronicle & Smith, 2020, p. 69)

In essence, while SIBs aimed for impact, their financial structures frequently pushed providers to prioritize their survival and easy achievements over genuinely serving the most vulnerable, leading to unintended and counterproductive outcomes. More realistic SIB designs with a higher chance of success are important to reduce the need for perverse behavior.

6.2.3.5.Stakeholder Development

Apart from financial success or qualitative outcomes related to investors and beneficiaries, the SIB model positively impacted the development of various stakeholders, increasing their capabilities and putting them in a better position for future SIBs. This development was particularly evident in SSPs:

> Although something of a 'baptism of fire', the delivery organisation had come through the experience with its capacity significantly enhanced – and even ready to take on another SIB project. (Griffiths et al., 2016, p. 31)

> The more robust data collection and management systems, combined with the requirement to report to the investors on a monthly basis, coupled with a stronger focus on outcomes, all resulted in the service providers having a better understanding of the project performance compared to running services through standard contracts. (Ronicle & Smith, 2020, p. 56)

> Both Working Well and Twining stated explicitly that they had benefited from the additional discipline and scrutiny of delivering an outcomes-based contract. This had improved their culture and ability to deliver future contract on an outcomes basis. (Stanworth & ATQ Consultants, 2023, p. 58)

Across different SIBs, the trajectory for providers was similar: steep early learning leading to measurably enhanced capacity and readiness. However, a key difference was the divergent outcome for one MHEP provider (Making Space), where the contract became loss-making and ended early, showing a negative developmental impact, unlike the largely positive experiences of other providers, especially in YEF and IF SIBs.

> Instead, the change to the MHEP model with both challenging performance targets and increased scrutiny of validated employment outcomes, appears to have exposed delivery and uncovered some issues in Making Space's reporting, suggesting that previous performance was not at the level supposed. (Stanworth & ATQ Consultants, 2023, p. 40)

In the YEF, despite the positive experiences with the SIB, the SSPs lacked a suitable platform for learning and decided to create their own learning network to enhance the exchange between the different providers.

> Multiple service providers commented that they would have benefited from more opportunities for the providers from the four projects to convene and share learning. Eventually the projects proactively created their own peer learning network. (Ronicle & Smith, 2020, p. 46)

In addition to the SSPs, intermediaries like Social Finance underwent institutional learning and acknowledged the organizational developments that have taken place.

> In the process, it was acknowledged that much learning, capacity building and organisational development had taken place. (Griffiths et al., 2016, p. 29)

> SF also argue that it has refined and standardised the MHEP model so that "new contracts are much cleaner, simpler, and more consistent". Stakeholders argue that the strengths of its approach have been 'continuous improvement based on experience and testing, leading to a well-oiled model that is cheaper to set up and run'. (Stanworth & ATQ Consultants, 2023, p. 52)

In summary, SIBs acted as a powerful, albeit sometimes challenging, catalyst for the professional development and organizational capacity building of different stakeholders, particularly service providers.

6.2.4. Design Recommendations

Some of the reports already give recommendations on how to improve future designs. This study differentiates between *Partnership and Process Recommendations*, and *Metric and Contractual Recommendations*.

6.2.4.1. Partnership & Process Recommendations

One recurring theme across the reports is the stressed importance of relationships with local stakeholders and the embeddedness of the SIB in the local context.

> Using local contributors (…) in SIBs commissioned by central government increases the engagement of local public bodies and ensures to a certain degree that the SIB is embedded within the local ecosystem. (Ronicle & Smith, 2020, p. 23)

> With the hard to reach (…) you have to have community-based organisations who know these people, who can reach out to them, who have got the skills and the experience to be able to do that. (Griffiths et al., 2016, p. 35)

These relationships are important to success, making it easier to generate local support and reach the right target beneficiaries.

Furthermore, as already highlighted before, maintaining a flexible design and adapting to local contexts, changing circumstances, and participant needs is important, but the degree of flexibility should be agreed upon when setting up the SIB initially.

> The important learning point to note is that decisions around the degree to which the projects can be flexible and adapt needs to be agreed at the outset, communicated with all parties involved and built into legal documents to enable the flexibility to occur without challenge. (Ronicle & Smith, 2020, p. 48)

> When outcome numbers are set and there is no room for flexibility, service providers need to be able to ensure the profile of beneficiaries matches the outcomes they are trying to achieve (whilst avoiding cherry picking). (Ronicle & Smith, 2020, p. 39)

> This said, the flexibility of providers to reprofile and reorientate projects was a key part of the success of the pilot programme, and retaining flexibility in future SIBs will be important for promoting innovation and continued learning. (Griffiths et al., 2016, p. 93)

Connected to this, one report also points out that the delivery teams should be involved in discussions around the SIB set-up as early as possible and that mid-way changes should be kept to a minimum.

> While remaining flexible, future iterations of the model should involve delivery teams in discussions regarding model structures as early as possible after commissioning, and should try to minimise changes to the structure during implementation to avoid confusion. (Gadenne et al., 2020, p. 48)

Lastly, one finding from the IF SIBs is that clear role separations between the stakeholders are needed, particularly to enable the SSPs to focus on the actual interventions, as one interviewed intermediary put it.

> For somebody it has to be about the numbers (. …) That means that that group has to worry about numbers; people on the ground worry about the interventions (. …) So often, people that are trying to offer services are encumbered by the numbers that go with them – if you get your processes sorted out there's no need for that. (Griffiths et al., 2016, p. 38)

Summarizing, the reports mention several potential improvements for future SIBs related to partnerships and processes. Even though these suggestions are not always immediate contractual SIB characteristics, they are important for the success of future SIBs.

6.2.4.2.Metric & Contractual Recommendations

The different reports oftentimes make concrete suggestions relating to the specific circumstances of the SIBs. Generally, there is an advocacy for a more realistic, conservative, and stress-tested target setting informed by evidence of past SIBs.

> Use the insight from this programme to set realistic volunteering targets, including sustainment at 13 and 26 weeks. (ConnectMore Solutions & Richmond Baxter Ltd., 2023, p. 42)

> Therefore there is a high risk of targets being missed unless they are substantially de-risked and based on a pessimistic view of likely performance. (Stanworth & ATQ Consultants, 2023, p. 68)

Some SSPs also campaign for a cap on outcome-contingent income, arguing that they cannot absorb as much risk and that the risk should rather be shifted towards the social investors:

> Twining did not, however, enjoy the financial impact when the service was performing below expectations and, like Working Well, has arrived independently at a similar conclusion that the maximum proportion of payment at risk to the providers under contracts such as this should be 25%. (Stanworth & ATQ Consultants, 2023, p. 49)

Another suggestion is that payment triggers should move beyond binary triggers to sliding scales linked to baseline status, allow several possible employment destinations, and give longer windows to achieve high-level outcomes.

> One suggestion put forward was that there should be a sliding scale of payments that would reward the distance travelled in getting a young person to an outcome, as this could vary enormously in terms of time and effort depending on where they started out. (Griffiths et al., 2016, p. 36)

> Provide more time for high-level (i.e. level 3 qualifications or employment with training) post-16 destinations to be achieved, recognising that it may take time for disadvantaged young people to achieve these outcomes. (Ronicle & Smith, 2020, p. 41)

This ensures a focus on the primary targets of the SIBs and reduces the risk of mission drift or the nudging towards easier-to-achieve outcomes. The longer time frames for higher-level outcomes do not reduce the necessity of lower-paid, earlier payment triggers, which are important to sustain cash flow for the SSPs. However, it needs to be ensured that those intermediate outcomes lead towards the high-level outcomes:

> Though intermediate outcomes are necessary for healthy cash flow, there is a need to ensure that short-term outcomes are truly accurate proxies for the ultimate social benefits being sought. (Griffiths et al., 2016, p. 94)

However, simply setting more realistic targets and adjusting the triggers is insufficient. The YEF evaluation report further states that outcome prices must cover delivery costs. Furthermore, there has to be a link between the interventions provided by the SSPs and the outcome metrics, and the metrics and corresponding evidence demands should also be aligned to keep the administrative burden as low as possible:

> When attaching costs to outcomes, commissioners should consider the cost of delivery for the target cohort (…), as well as any possible cashable savings. (Ronicle & Smith, 2020, p. 25)

> There must be a strong link between the intervention provided by the service provider and the outcomes that payments are attached to. The intervention must be the main influencer on those outcomes. (Ronicle & Smith, 2020, p. 39)

> However, it needs to be borne in mind that the programme's evidence requirements placed an administrative burden on the educational institutions (. …) Future SIB designs need to seriously consider how the metrics will impact on the resources of other organisations. (Ronicle & Smith, 2020, p. 73)

Across reports, it is clear that metrics must be realistic, risk-sharing proportional, and rate cards progressive, ensuring sufficient cash flows while incentivizing the focus on the ultimate social benefits.

7. Discussion

With this thesis, I aim to understand how COMs, an important aspect of SIBs, are defined and measured in SIB practice. In a further step, I want to shed light on the relationship between SIB design, particularly COMs and payment structures, measurement, and the subsequent success of SIBs. After giving an overview of the current state of research and providing quantitative and qualitative findings from completed SIBs of the *Employment and Training* area in the UK, I will now integrate and discuss these insights.

7.1. Definition and Measurement of Contractual Outcome Metrics

The SIB database showed that the considered UK SIBs oftentimes use similar binary, quantifiable COMs like sustained employment, job entry, or achieved qualifications. This pattern closely reflects the finding from the literature that SIBs from the same policy area tend to use similar measurable and auditable outcomes (Berndt & Wirth, 2018; Fox & Albertson, 2012; Warner, 2013).

Most SIBs use between 4 and 10 COMs in a payment-per-outcome structure, highlighting a focus on staged COMs and subsequent payments. However, this approach somewhat contradicts the literature that advocates for as few COMs as possible to keep the focus (Social Finance, 2016). The evaluation reports highlight the use of intermediate outcomes to reflect the process beneficiaries undergo on the way to reach the ultimate outcomes, and this approach also ensures cash flow for providers. While sufficient cash flow is important and cash flow issues have frequently been raised in the reports, these intermediate outcomes also contain the risk of losing focus on the actual end goals of the SIBs and incentivizing providers to focus on easier-to-reach intermediate outcomes. Therefore, a thorough SIB design is crucial to minimize this risk. Generally, COMs should be linked to the intervention, but intermediate outcomes should also be logically linked to the end goals. The evaluation reports found, for example, that specific COMs like *User Engagement* are important predictors for achieving subsequent COMs. Thereby, reaching this COM makes it more likely that later COMs will also be achieved or at least highlights problems early on.

The lack of data makes it difficult to make general statements about more or less achievable COMs. What can be said is that *QCF Qualifications* have the best achievement rate overall. However, many external factors can influence the achievement of COMs, such as changing economic circumstances that make it challenging to place beneficiaries in jobs. This means that it is difficult to compare achievements across SIBs because the same COM might be achieved or not in different SIBs due to different circumstances. An example is job outcomes that were very successful in the FutureImpact SIB, while the achievement rate in the YEF reports is mixed and very negative in the MHEP SIBs. Furthermore, it also means that, depending on the context, it might not be feasible to use the same COMs for similar interventions. Also, the definition of the COMs is important and influences the possibility of achieving them. For example, to claim the *School Attendance* payment for beneficiaries, a threshold based on the average attendance of the student had to be passed, irrespective of the starting point of the individual. This makes

it much harder to achieve the COM with a more disadvantaged group of beneficiaries and leads to perverse incentives. Other metric designs that allow for different starting points of beneficiaries, like sliding scales or distance-traveled measures, were proposed during stakeholder interviews to account for this problem.

While the literature emphasizes that COMs should be co-designed by all relevant stakeholders (Berndt & Wirth, 2018; Fox & Albertson, 2012; Warner, 2013), qualitative findings reveal that, in practice, COMs are often shaped by funder requirements and the need for measurable outcomes and do not reflect the reality of service delivery. Since these initially defined COMs were often flawed, some SIBs underwent mid-way restructuring by adding intermediate COM stages. The adjustment processes were more collaborative as they took into account the feedback and struggles of the providers and were aimed at relieving their problems.

Another insight from the quantitative and qualitative analysis is the dominant focus on quantitative COMs. Even though qualitative outcomes such as confidence and well-being are highly valued by all stakeholders and social impact is the ultimate goal of the SIBs, this is rarely reflected in the COMs, which is also generally criticized by Grünhaus & Rauscher (2022). The only qualitative outcomes that are frequently used are *School Behaviour, Attendance, and Attitude*. However, in their current use, these COMs also have their flaws. While *School Behaviour and Attitude* are somewhat subjective metrics due to their evidence requirements, *School Attendance* uses the same threshold for all beneficiaries, making it harder to achieve for individuals with a lower starting point. Additionally, all of them are defined as binary metrics, ignoring the distance the individual beneficiaries traveled.

Measuring qualitative outcomes is much more complex and subjective than measuring quantitative outcomes, making it more challenging to attach payments to them. Nevertheless, neglecting these outcomes in the COMs might simplify complex problems and risks to divert the focus away from the underlying issues the SIBs try to tackle and the broader social value they aim to create (Andreu, 2018; Dayson, 2017). Considering the results from the content analysis, it becomes apparent that qualitative outcomes were achieved, even if they were not defined as targets in the COMs. I will discuss this in more detail in the following subchapter. However, the COMs focus on individual beneficiaries, meaning potential structural problems are not addressed, and this might limit the long-term effects once the interventions are completed.

Some service providers felt they had limited control over some COMs, raising concerns about fairness and achievability. As a result, this might reduce the motivation to perform from the SSPs. It also aligns with agency theory's warnings about misaligned incentives and the risk of perverse behaviors when metrics do not reflect actual provider control (Maier & Meyer, 2017; Morley, 2019). Chapters 7.1.3. and 7.2. will go into more detail on unintended consequences and agency theoretical perspectives.

The most common evidence for COM achievement is confirmation letters. However, SIB designers can choose from a variety of validation methods. Depending on the COMs used and the specific validation method, the reporting burdens for providers can be high, something noted in the evaluation reports and the academic literature (Fox & Albertson, 2011). For example, while some outcomes, such as qualifications, only had to be evidenced once, others, such as *Sustained Employment*, required it twice. These administrative burdens can detract from service provision, negatively impact relationships with partners, e.g., schools, as stated in the YEF report (Ronicle & Smith, 2020), and lower the overall impact generated. To avoid the negative impacts from reporting requirements as much as possible, the requirements should be streamlined, e.g., by keeping reporting demands proportionate and feasible, and avoiding duplication.

Summarizing, the definition and measurement of COMs reflect a balancing act between the needs for measurable, risk-managed outcomes and the realities of social service delivery, which values incremental, context-dependent, and often qualitative progress. While the real-world insights align partly with the literature, e.g., regarding the reporting burdens, there are also deviations. This is especially true for the proposed co-design of COMs, which considers all relevant stakeholders' viewpoints, but is not that evident in reality. While the achievement of COMs is highly context-dependent, it is important to set realistic and controllable targets for SSPs to enhance the motivation and the likelihood of COM achievement.

7.2. SIB Design, Impact Measurement, and Success

The UK Employment & Training SIBs are characterized by high structural homogeneity. All use a payment-per-outcome structure and are commissioned as part of larger funds. This standardization is reflected in the relatively narrow range of COMs and the dominance of a few intermediaries and evaluation approaches. This homogeneity also makes it challenging to analyze the impact of specific characteristics on the success of the SIBs.

The payment-per-outcome structure and the associated staged payments are designed to manage financial risk and ensure cash flow for providers. The individual payment amount for similar outcomes can vary considerably between SIBs or funds. Some differences result from lessons learned, as the IF shows, which adapted payments from round one to round two of the commissioned SIBs (Griffiths et al., 2016). Capping possible payments for each individual was a common structural design across the SIBs and funds. However, the YEF SIBs additionally had a volume cap on how often every COM could be claimed. The combination of both caps led to detrimental effects on service provision because it restricted the SSPs' ability to support the most vulnerable people (Ronicle & Smith, 2020). SIB designers, therefore, have to be careful which kind of caps they use.

The evaluation reports show that many SIBs had flaws in their initial designs, and the providers faced cash flow issues resulting from underperformance. This was driven by unrealistic targets and a share of outcome-dependent payments that were considered too high by some providers, especially in the MHEP SIBs (Stanworth & ATQ Consultants, 2023). This increased the financial exposure to be borne by the providers and transferred the risk more towards the providers, which is opposite to the central idea of SIBs, which is that investors carry most of the risk. As a result, the providers experienced enhanced performance pressure, which is already higher in SIB models compared to more conventional contract models. In response, several SIBs introduced mid-program adaptations, such as adding additional intermediate COMs or changing payment structures to alleviate the financial pressure for the SSPs. These adaptations improved the performance of the SIBs (e.g., Griffiths et al., 2016), highlighting the need and importance of flexibility in SIBs to react to changed circumstances or identified design issues, which is also emphasized in the literature (Maier et al., 2018). While important for success, a good SIB design should also set boundaries for adaptations to align the expectations of the different stakeholders and create more certainty. As can be seen, setting up an appropriate SIB design from the start is difficult. This also calls for a stronger learning from completed SIBs with similar backgrounds when designing a SIB to minimize avoidable mistakes. The designs of SIBs that were too rigid were also criticized by stakeholders, especially providers (Ronicle & Smith, 2020).

Most SIBs employed mixed-methods evaluations for their impact measurement, combining insights from different data sources like administrative data and interviews. The literature also highlights the importance of SIB evaluations considering counterfactual risk to attribute the SIBs' results to the interventions (e.g., Fox & Albertson, 2011; Mulgan et al., 2011). However,

the practical insights show that this rarely happens. This is a significant drawback of the methodologies used, as it leads to an inability for causal attribution, which significantly limits the meaningfulness of the evaluations and does not allow drawing definite conclusions about the actual impact of the SIBs (McHugh et al., 2013). Even though the IF used comparison groups for its evaluation by employing PSM (Salis et al., 2018), this is not comparable to an RCT, which delivers more robust results. The use of SROI for the IF evaluation has to be acknowledged since it is a comprehensive impact analysis that is more sophisticated than the evaluation methods used by other funds. However, the suitability of this method for precisely measuring the social impact of the SIBs can be questioned. In the IF, the results were very low, and it was also mentioned that they would probably be even lower when additionality is considered. This also means that how the SROI was employed in that particular evaluation did not capture the entire impact, and it is reasonable to assume that a more sophisticated SROI evaluation would lead to significantly more effort being needed, given that it is theoretically possible in the first place. The latter is also a challenge since qualitative outcomes have to be quantified. Additionally, due to its individualistic calculation, the SROI is hardly comparable across different contexts, reducing its informative value (Maier et al., 2015). The reasons given in the evaluation reports for not using more rigorous quantitative evaluation methods are the high costs, complexity, and data access challenges. The literature also mentions costs as a major barrier (Social Finance, 2016; Williams, 2021).

When looking at the success of SIBs, the reports demonstrate that it is important to consider both financial/COM achievements and qualitative outcomes, as was already found in the literature. It is also interesting that the success perspectives for both categories can deviate within individual SIBs. The findings for financial success were very mixed. In the end, many SIBs were considered financially successful, but achieving this was often only possible due to restructuring and redesigning. The SIB design, therefore, definitely influences SIB success. Poorly defined SIBs, e.g., through inadequate payment structures or COM targets that are too optimistic, result in poor performance. This highlights the need for targets based on the on-the-ground realities, but does not necessarily call for more COMs and tighter control, because keeping flexibility in delivering the intervention is also important.

At the same time, the qualitative analysis reveals that SIBs often generated substantial social impact for beneficiaries, such as increased confidence or better social relationships with their families, even when financial success was not achieved. As previously discussed, these outcomes are rarely tied to payments, with the exception of School Behavior, Attendance, and

Attitude, even though they are highly valued across stakeholders. As a result, some stakeholders questioned whether the impact measurement captures the full impact achieved by the projects. I also argue that COMs and impact measurement in their current form are insufficient for capturing change, primarily due to a lack of causal attribution and incomplete metrics. This tension between financial success and achieved social impact also reflects broader critiques in the literature about the adequacy of using binary metrics and payment triggers as proxies for social impact (Broccardo et al., 2020; Millner & Meyer, 2022).

In addition to the qualitative outcomes for the beneficiaries, the stakeholders also frequently reported that they developed themselves within these SIB pilots, meaning their capabilities increased, which equips them for future SIBs and makes success more likely. There is only one instance in which there was a negative development and an early termination of the SIB contract. Especially, SSPs also stressed the importance of learning opportunities, such as a structured exchange with other SSPs.

Since all but one SIB used an intermediary, it is impossible to conclude the importance of an intermediary for a SIB's success. However, the stakeholders sometimes critically questioned the intermediary's usefulness. Overall, there was a positive perception of using an intermediary as a connecting stakeholder within the SIB structure that communicates with all parties and fosters commitment. However, the value for money was questioned. The intermediaries produce significant costs, and some SSPs perceived them as too high for the output generated. Therefore, to justify the existence of an intermediary, it is important to stress its role and contribution to all stakeholders involved.

Overall, there is a clear connection between SIB design, impact measurement, and financial success. A well-designed SIB increases the likelihood of a SIB succeeding and achieving the COMs. However, the qualitative analysis reveals that improvements in SIB design and impact measurement are needed to enable better-performing SIBs and capture a project's full impact. Keeping flexibility to adjust to experienced design problems or changing circumstances is essential for a SIB’s performance, but SIB designers should also aim to improve the designs from the start. The lack of inclusion of qualitative outcomes in the design and measurement is a drawback of current SIBs, as social impact is regarded as very important but is not equally reflected in the current structures. Impact measurement also remains a significant challenge due to the lack of robust counterfactuals, data quality issues, and administrative burdens, which significantly limit the accuracy and meaningfulness of the current evaluations. However, the

overall perception of the SIB designs across stakeholders was positive despite the challenges identified.

7.3. Unintended Consequences and Perverse Incentives

While SIBs are designed to align the interests of commissioners, investors, and service providers through outcome-based payments, the literature and empirical evidence reveal that these structures often generate unintended consequences and perverse incentives. These effects can undermine SIBs' social mission, distort service delivery, and disadvantage the populations they aim to help.

The empirical evidence, especially the YEF report, discussed perverse behaviors in their SIBs (Ronicle & Smith, 2020). The reports document explicit cases of cherry-picking where providers prioritized beneficiaries who were more likely to achieve an outcome to increase the likelihood of achieving outcome payments. This, in turn, led to the neglect of the most disadvantaged people, whom the SIBs were initially intended to help. Moreover, the reports discovered behaviors of parking, mission drift, and gaming, all to hit target outcomes, and to the detriment of beneficiaries. It was also found that SSPs tended to focus primarily on payment-linked outcomes to neglect broader, non-incentivized goals. However, it must also be noted that some SSPs continued to provide services even after reaching contractually defined caps. While the use of contractual caps makes sense to limit the total possible amount of payments, the content analysis also revealed that especially a combination of payment and volume caps leads to negative outcomes for the beneficiaries.

The reasons for these perverse behaviors are primarily to be found in the SIB design, specifically the COMs and the associated payment structures. The literature also reflects this perspective. Fox & Albertson (2012) mention exactly the risk that outcome-based systems could lead to a neglect of non-payment-linked outcomes. Several research papers also highlight the risk that binary or hard-to-achieve COMs, or being primarily financially driven, incentivize cherry-picking, cream-skimming, and parking (Maier & Meyer, 2017; Morley, 2019). This can result in the most vulnerable group of people being sidelined, contradicting the very goal of the SIB and confirmed by the empirical findings. The risk of mission drift is prominent in the literature and frequently mentioned as a potential downside of equity capital (Achleitner et al., 2013). Therefore, the use of frequency COMs should be studied and tested, which are assumed to reduce the risk of perverse incentives (Maier & Meyer, 2017).

The empirical evidence shows that unintended consequences and perverse incentives are not hypothetical risks but empirically observed realities. The literature and evaluation reports demonstrate that the structure of outcome metrics and payment mechanisms can drive behaviors that undermine social interventions' equity, effectiveness, and integrity. Addressing these risks requires more nuanced, context-sensitive contract and metric designs, realistic target-setting, and mechanisms that reward genuine progress for the most disadvantaged, not the easiest wins.

7.4. Integration with Agency Theory

SIBs bring together various actors and need to align the interests of these stakeholders. In general, the outcome-based payment structure intends to align interests and incentivize performance. However, the empirical evidence shows friction, so it is interesting to investigate SIBs from an agency theoretical perspective.

Throughout the evaluation reports, stakeholders made many statements highlighting the positive, motivating aspects of SIBs that foster a stronger focus on performance. This highly resonates with agency theory, which has at its core that the agents perform well and work in the principal's best interest.

However, in SIBs, there is not just one principal-agent relationship. The presence of multiple principals (commissioners, investors) and agents (providers, intermediaries) in SIBs creates a web of agency relationships. The reports generally describe a shared interest of different stakeholders in achieving social impact. They also point out the willingness of investors to accept smaller financial returns and to actively engage to achieve the desired impact and make the projects a success, thereby aligning with the literature (Carter, 2019). This demonstrates the ability of SIBs to surround different stakeholders around the central goal of social impact, which is also seen as a strength of SIBs, particularly during the design phase. However, the alignment of interests is not always straightforward, as the problems in some SIBs reveal. In the MHEP SIBs, the SSPs described feeling caught between conflicting primary demands, such as prioritizing job outcomes for payment versus broader mental health outcomes for beneficiaries. In another case, there was a SIB with two funding streams, and the instructions of the two funders were not aligned, leaving the SSP in a difficult situation, unable to implement the SIB as intended (Gadenne et al., 2020). These examples show that it is important for the stakeholders to be aligned because otherwise, the service provision will suffer, making it unlikely that the SIB will achieve its ultimate goals.

The evaluation reports suggest that outcome metrics were imposed top-down on providers, with little previous negotiation involving them. This is evident in statements about COMs being outside provider control and that frequent restructuring was necessary to reflect on-the-ground realities. It is also contrary to the literature that suggests that all relevant stakeholders should be involved in defining the COMs. Furthermore, one evaluation report noted that SSPs should be involved in the SIB set-up as early as possible. It is not known whether early involvement of providers would have prevented these problems, but from an agency theoretical perspective, the lack of involvement led to residual losses.

The use of COMs and extensive validation also represents an attempt to reduce information asymmetry between principals and agents and monitor agent performance. However, this comes with the cost of an increased administrative burden, fewer resources for service provision, and, at times, strained relationships with partners. Additionally, using too much control by defining more COMs or precisely specifying how to deliver the interventions would take away flexibility, which is considered a strength of SIBs and is important for the SSPs to adjust their services to the experienced realities.

Another highly relevant issue is the existence of perverse incentives. From an agency theoretical perspective, contracts should usually prevent exactly this from happening. The fact that these incentives exist nevertheless indicates that the interests and motivations of the different parties are not perfectly aligned and that there are gaps in the contract.

Lastly, stakeholders emphasized the importance of trust, collaboration, and ongoing learning. These elements are less profound in agency theory, but more highlighted in stewardship theory, which was also highlighted by previous research (Hevenstone et al., 2023).

In summary, agency theory is well-suited to explaining many aspects of what works well in SIBs and where and why challenges exist. However, the empirical findings also highlight the limits of contractual solutions and the importance of ongoing adaptation, trust, and collaboration. To fully understand and improve SIB performance, agency theory should be complemented by other frameworks like stewardship theory that account for the dynamic, relational, and multi-stakeholder nature of social innovation.

8. Conclusion

This thesis analyzes how contractual outcome metrics in Social Impact Bonds are defined and measured, as well as how the contractual design of SIBs is linked to their success. Success can be further divided into financial success and qualitative outcomes. Financial success describes the achievement of COMs linked to outcome payments, while qualitative outcomes are typically not captured by the COMs and therefore not connected to payments. To critically assess the usage of COMs, I also looked at other factors that might influence the financial success and qualitative outcomes of SIBs, such as the flexibility of the SIB design and the stakeholder collaboration throughout the SIB lifecycle.

The central questions for this research were (1) how contractual outcome metrics in SIBs are defined and measured, and (2) how success is related to the contractual design of SIBs and impact measurement.

I focused on UK SIBs from the Employment & Training sector and employed a mixed-methods approach by combining a quantitative analysis of a self-constructed SIB database primarily based on the INDIGO Impact Bond Database with a qualitative content analysis of SIB evaluation reports. I chose this approach to get a holistic view of SIBs, particularly COMs. Especially, the content analysis allowed for generating new insights by uncovering perceptions of different stakeholders that cannot be retrieved by simply looking at numbers or facts from the database.

8.1. Main Findings

Regarding the first research question, my findings reveal that SIBs in Employment & Training often use similar, binary COMs that can be clustered into a limited number of COM groups. The most frequent groups are *QCF Qualifications*, *Sustained Employment*, and *School Behavior, Attendance and Attitude*. The binary COMs promote measurability, but provide a higher risk of perverse incentives. I also found that the investigated SIBs largely neglect qualitative outcomes in their COMs, even though these are highly valued. To verify the COMs, various validation methods, like confirmation letters and copies of certificates, are employed throughout the SIBs, and oftentimes, it is possible to use one of several methods for validation. The different COMs and validation methods have varying impacts on the administrative burden for the providers, with more effort being needed to verify some COMs compared to others.

As for the second research question, the quantitative data does, unfortunately, not show clear patterns of relationships between COM achievement and different COM groups or COM payments. The limited data points were a significant hurdle, resulting in weak quantitative results. However, even though the available data provides limited insights, my qualitative analysis shows that the wider SIB design is closely tied to financial success. Maintaining a flexible SIB design was found to be very important. For many SIBs, achieving outcome targets depended on mid-program adaptations, such as adding COMs or adjusting the payment structure, to react to flaws in the original designs. A key reason for the adaptations was that the original COMs were often unrealistic or overambitious and therefore resulted in lower performance. Despite consistent appraisal of qualitative outcomes from all stakeholders, the SIBs rarely recognized or rewarded qualitative outcomes like improved self-confidence of the beneficiaries.

I also found that the impact evaluation of SIBs at the end of the interventions lacks quasi-experimental or randomized approaches that can isolate the effects of the interventions, making it difficult to attribute the results to the interventions and ultimately to demonstrate the effect of the SIBs. Most investigated SIBs only relied on administrative data and interviews to assess the overall performance. These evaluation methods are easier to conduct but produce less reliable results.

From an agency theoretical perspective, I found an overall positive resume across the different stakeholders that SIBs align the different actors around the focus on performance and generating social impact. The COMs generally intend and manage to foster these aligned interests. However, I also found that metric development lacks involvement of the SSPs, leading to poorly defined metrics that do not reflect reality and result in the necessary adaptations throughout the implementation phases of the SIBs. Additionally, the SSPs report that the validation requirements, which aim to reduce information asymmetry, lead to significant administrative burden, which reduces the perceived benefit of the SIBs, as it takes away valuable resources to deliver interventions. The fact that SIB contracts do not manage to align the diverging interests perfectly becomes apparent through the existence of perverse behaviors, such as cherry-picking of beneficiaries. However, not all ongoing mechanisms within SIBs can be explained with agency theory. I also found elements of stewardship theory, which looks more at the relational aspects of SIBs, such as collaboration and trust among stakeholders.

In sum, I demonstrate in this thesis that SIBs can focus stakeholders' attention on performance and foster partnership across diverse stakeholders. COMs play an important part in directing the focus of the interventions, but oftentimes are not realistic and therefore fail to incentivize performance optimally. The content analysis also shows that COMs are less important for the ultimate success of the SIBs than anticipated. Other factors, like the flexibility of the SIB design, also seem very important and offer room for future research. Lastly, many SIB stakeholders recognize the qualitative outcomes achieved by the interventions, but acknowledge at the same time that the current COMs do not capture them.

8.2. Implications for Theory and Practice

I make several important contributions to the broader academic literature on SIBs with this thesis. First, I put an agency theoretical lens on SIBs and highlight that it is highly relevant to explain how SIB designs try to reduce agency problems through the alignment of interests by having a common focus on performance and social impact, clear outcomes and monitoring mechanisms in place. At the same time, the evidence of partial alignment, perverse incentives, and costly monitoring is consistent with classical principal-agent dilemmas and validates existing literature on SIBs (e.g., Carter, 2019; Fox & Albertson, 2011; Fraser, Tan, Lagarde & Mays, 2018). I also find that other theoretical frameworks besides agency theory are needed to better understand how SIBs work, especially on the relational side between the stakeholders.

Moreover, I stress the importance of including qualitative COMs in the SIB design. The investigated SIBs generally use binary, quantitative COMs that are easier to measure and validate, but, at the same time, constrain stakeholders in the variety of impacts that are measured, as they do not capture qualitative outcomes, which are also part of the overall success. As a result, I note that the use of COMs in these SIBs cannot capture the entire impact created, and a more holistic view on success, including quantitative and qualitative outcomes, is needed, which is in line with Millner & Meyer (2022). I suggest that the impact measurement literature should focus on identifying additional COMs that can also cover these qualitative aspects.

Additionally, I provide an overview of COMs used in SIBs and their validation methods. I show a significant overlap of COMs and validation methods across the investigated SIBs, which are all from the same policy area. Future research could build on this and investigate other policy areas of SIBs and whether certain COMs could be used across policy areas, potentially making the initial negotiation processes easier and the SIBs more comparable.

Another theoretical contribution of my study is the findings on the importance of mid-program adaptations and stakeholder engagement of the projects for achieving the targets of the SIBs, which reflects the existing literature (Maier et al., 2018).

Practically, the study demonstrates the need to design more effective SIB contracts. This does not necessarily mean increasing the control mechanisms within the contracts, such as the number of COMs, but designing contracts, particularly COMs and payment structures, that reflect the reality of the interventions and transfer the risk to the investors. Poorly defined contracts have substantial detrimental effects on the SIB's success and the beneficiaries due to a higher risk of perverse incentives, lower performance, and increased cash-flow issues for SSPs. SIB designers should draw on robust data from previous SIBs, similar interventions, and the input from providers to set reasonable targets. Additionally, they need to make sure that providers are properly incentivized to perform while simultaneously ensuring a stable cash flow through a mix of block payments and staged payments for intermediate outcomes in payment-per-outcome structures. Of course, if a different payment structure is used, the applicability of the findings is limited, but ensuring sufficient funds for the SSPs while incentivizing performance is important anyway.

A recurring theme across SIBs was the emphasis on a flexible and adaptive contract design as a key success factor. This insight is important for practitioners, as they should not create SIB mechanisms that are too rigid, but also ensure that certain boundaries exist. Therefore, future SIBs should formalize flexibility in SIB contracts, allowing for structured negotiation points.

Another major takeaway for practitioners is the need for improved outcome metrics in SIBs, including the incorporation of qualitative outcomes, and better impact measurement. This would help to capture the true impact of SIBs. Additionally, SIBs should move away from only using binary metrics and experiment with different metric designs, like sliding scales or distance-traveled measures, that allow for different starting points of beneficiaries and reward individual progress more. Ensuring COMs are closely linked to the intervention and within the SSPs' perceived control is also important to ensure sustained motivation. Intermediate outcomes should also be linked to the desired end goals to ensure the delivery of the intervention is directed toward them. Finally, SIB evaluators should use more robust evaluation designs like RCTs to improve causal attribution and reduce counterfactual risk. Linked to that, evidence requirements should be streamlined, and the administrative burden should be reduced as much

as possible. This leaves the SSPs with more time to deliver the interventions and generate impact.

However, improving metrics and measurements alone is not sufficient. It is equally important to ensure good stakeholder collaboration and governance structures. This includes building trustful relationships between the stakeholders, including all relevant stakeholders in negotiations, ensuring alignment on the desired outcomes, embedding the SIB in the local environment, and optimizing processes, like SIB design and performance reporting. Furthermore, offering opportunities for exchange across SIBs encourages cross-learning and further facilitates improvements.

Looking beyond SIBs, some implications are also relevant for the overarching field of impact investing. Defining appropriate, well-defined outcome targets that incorporate qualitative aspects and prevent perverse incentives is imperative for all forms of impact investing to ensure that the full impact is captured and the interventions are directed towards achieving social impact. Improvements in impact measurement are also an overarching topic in impact investing to be able to measure the generated impact. Developing standardized metrics and impact measurement techniques would allow for comparing the performances of different projects. However, standardization is limited at the same time by the need to adapt the projects to local conditions to ensure high relevance and optimal delivery of the interventions. Lastly, practitioners in impact investing need to consider that achieving success takes more than just well-defined contracts. Informal aspects like having adaptive capabilities and being flexible to react to changing circumstances, building trustful relationships and good collaboration among stakeholders, and including all relevant interest groups are equally important to foster commitment and align the interests and actions of the different parties.

8.3. Limitations and Future Research

The findings must be interpreted cautiously, as this study has several important limitations. First, data availability is a major limitation. Even though the INDIGO dataset already provides a large amount of information, the quantitative analysis was limited by incomplete or missing data for several SIBs – even after extensive manual research. Many SIBs lacked comprehensive final evaluation reports, meaning a deeper content analysis was only possible for a smaller sub-sample. Furthermore, the depth of the reports differed significantly, impacting the quality and comparability of the results and making statistical analyses challenging. Two SIB evaluations

also took place before the intervention was finished and worked with estimations, which entails the risk of over- or underestimating the final effects.

Additionally, several SIB database variables, such as *COM Group* and *Success*, were manually filled based on evaluation reports and my judgment. This introduces a risk of subjective interpretation and leaves room for errors. Furthermore, *Success* was measured as the achievement of COMs if no other information was available. This approach risks overlooking the broader social impact generated. Qualitative evidence showed that SIBs often delivered meaningful social impact, which was not reflected in payment structures. Future research should investigate how to incorporate qualitative outcomes better into SIB design to ensure this impact is also measured.

Given that I conducted the qualitative content analysis alone, there is a risk of single-coder bias. Even if I tried to minimize this bias by double-coding all reports, this cannot be completely ruled out.

Another major limitation is the focus on UK SIBs in the Employment & Training area. While this approach allows for a better comparison of the results and rules out different country-specific factors as potential, influential variables, it simultaneously limits the generalizability of the findings to other policy areas or countries. The UK context is unique due to its relatively mature SIB market and high data transparency. Pronounced differences from other countries are the payment structures. Whereas all SIBs considered in this study had a payment-per-outcome structure, SIBs in countries like Germany or Austria tend to have an all-or-nothing payment structure. Future research could start here and investigate potential differences between different payment structures and the success of SIBs. Additionally, research is needed on SIBs in different policy areas and countries to understand how contractual designs and COMs vary and what works best in different contexts.

SIB outcomes were often also shaped by factors outside providers' control or the SIB mechanism itself, such as labor market conditions, policy shifts, or the performance of partner organizations. These contextual influences complicate attribution and make it difficult to assess the real impact of the intervention. As this study was explorative, I did not control for such factors. In the future, quantitative studies controlling for external factors are needed to test if there are statistically significant influences of SIB design and impact measurement on SIB success.

In addition to the already mentioned future avenues for research, longitudinal studies that track beneficiaries over extended periods, also after the end of the SIBs, are needed to assess the durability of outcomes achieved through them. This includes examining whether qualitative outcomes translate into sustained social and economic benefits and how SIBs contribute to systemic change.

Furthermore, different theoretical frameworks should be applied in future research. As I have shown, agency theory can explain many SIBs' incentive and monitoring challenges, but other relational dynamics are also at play that can be better explained with frameworks such as stewardship theory.

The evaluation reports highlighted that substantial organizational capacity development has taken place throughout the lifetime of a SIB, enabling different stakeholders to manage future SIBs better. Future research could go into more depth here and investigate how participation in SIBs influences organizational learning and capacity building among different SIB stakeholders, especially providers and intermediaries.

Lastly, the presence of unintended consequences and perverse incentives calls for more insights into the ethical implications of SIBs to inform more just and inclusive contract designs for the future.

9. References

Achleitner, A.-K., Lutz, E., Mayer, J., & Spiess-Knafl, W. (2013). Disentangling Gut Feeling: Assessing the Integrity of Social Entrepreneurs. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, *24*(1), 93–124. https://doi.org/10.1007/s11266-012-9264-2

Albertson, K., Fox, C., O'Leary, C., & Painter, G. (2020). Towards a Theoretical Framework for Social Impact Bonds. *Nonprofit Policy Forum*, *0*(0). https://doi.org/10.1515/npf-2019-0056

Andersen, M. M., Dilling-Hansen, R., & Hansen, A. V. (2020). Expanding the Concept of Social Impact Bonds. *Journal of Social Entrepreneurship*, *13*(3), 390–407. https://doi.org/10.1080/19420676.2020.1806100

Andreu, M. (2018). A Responsibility to Profit? Social Impact Bonds as a Form of "Humanitarian Finance." *New Political Science*, *40*(4), 708–726. https://doi.org/10.1080/07393148.2018.1528533

Arena, M., Bengo, I., Calderini, M., & Chiodo, V. (2016). Social Impact Bonds: Blockbuster or Flash in a Pan? *International Journal of Public Administration*, *39*(12), 927–939. https://doi.org/10.1080/01900692.2015.1057852

Bafford, B. (2012). The Feasibility and Future of Social Impact Bonds in the United States. *Sanford Journal of Public Policy, 3*(1), 3–19.

Barajas, A., Barajas, L., Burt, K., Harper Jr, T., Johnson, P., Larsen, E., Licona, J., Nije, C. O., Parker, W., Reid, K., Sturtevant, L., Tokunaga, M., Vien, Q. & Yeh, C. (2014). Social Impact Bonds: A new tool for social financing. New Jersey. NJ: *PPIA Program. Princeton University*.

Becchetti, L., Pisani, F., Salustri, F., & Semplici, L. (2021). The frontier of social impact finance in the public sector: Theory and two case studies. *Annals of Public and Cooperative Economics*, *93*(4), 887–912. https://doi.org/10.1111/apce.12357

Bendickson, J., Muldoon, J., Liguori, E., & Davis, P. E. (2016). Agency theory: the times, They Are a-changin'. *Management Decision*, *54*(1), 174–193. https://doi.org/10.1108/MD-02-2015-0058

Berndt, C., & Wirth, M. (2018). Market, metrics, morals: The Social Impact Bond as an emerging social policy instrument. *Geoforum*, *90*, 27–35. https://doi.org/10.1016/j.geoforum.2018.01.019

Broccardo, E., Mazzuca, M., & Frigotto, M. L. (2020). Social impact bonds: The evolution of research and a review of the academic literature. *Corporate Social Responsibility and Environmental Management*, *27*. https://doi.org/10.1002/csr.1886

Carè, R., Carè, S., Lévy, N., & Fatima, R. (2023). Missing finance in social impact bond research? A bibliometric overview between past and future research. *Corporate Social Responsibility and Environmental Management*, *30*(5), 2101–2120. https://doi.org/10.1002/csr.2496

Carè, R., Rania, F., & De Lisa, R. (2020). Critical Success Factors, Motivations, and Risks in Social Impact Bonds. *Sustainability*, *12*(18), 7291. https://doi.org/10.3390/su12187291

Carter, E. (2019). More than marketised? Exploring the governance and accountability mechanisms at play in Social Impact Bonds. *Journal of Economic Policy Reform*, *24*(1), 1–17. https://doi.org/10.1080/17487870.2019.1575736

Chiappini, H., Marinelli, N., Raja, & Giuliana Birindelli. (2023). Past, present and future of impact investing and closely related financial vehicles: a literature review. *Sustainability Accounting, Management and Policy Journal*, *14*(7), 232–257. https://doi.org/10.1108/sampj-09-2022-0471

ConnectMore Solutions, & Richmond Baxter Ltd. (2023). *Future Impact Programme Evaluation*. https://golab.bsg.ox.ac.uk/documents/FutureImpact_LCF_-_Final_report.pdf

Cooper, C., Graham, C., & Himick, D. (2016). Social impact bonds: The securitization of the homeless. *Accounting, Organizations and Society*, *55*, 63–82. https://doi.org/10.1016/j.aos.2016.10.003

Dahbi, F., Carrasco, I., & Petracci, B. (2024). A systematic literature review on social impact bonds. *Finance Research Letters*, *62*, 1-8. https://doi.org/10.1016/j.frl.2024.105063

Dayson, C. (2017). Evaluating social innovations and their contribution to social value: the benefits of a "blended value" approach. *Policy & Politics*, *45*(3), 395–411. https://doi.org/10.1332/030557316x14564838832035

Del Giudice, A., & Migliavacca, M. (2018). Social Impact Bonds and Institutional Investors: An Empirical Analysis of a Complicated Relationship. *Nonprofit and Voluntary Sector Quarterly*, *48*(1), 50–70. https://doi.org/10.1177/0899764018797480

Dey, C., & Gibbon, J. (2018). New development: Private finance over public good? Questioning the value of impact bonds. *Public Money & Management*, *38*(5), 375–378. https://doi.org/10.1080/09540962.2018.1477676

Disley, E., Giacomantonio, C., Kruithof, K., Sim, M., & Europe, R. (2015). *The payment by results Social Impact Bond pilot at HMP Peterborough: final process evaluation report*. https://assets.publishing.service.gov.uk/media/5a7f190fed915d74e33f4513/social-impact-bond-pilot-peterborough-report.pdf

Disley, E., Rubin, J., Scraggs, E., Burrowes, N., & Culley, D. M. (2011). *Lessons learned from the planning and early implementation of the Social Impact Bond at HMP Peterborough*. RAND Europe. https://www.rand.org/content/dam/rand/pubs/technical_reports/2011/RAND_TR1166.pdf

Edmiston, D., & Nicholls, A. (2018). Social Impact Bonds: The Role of Private Capital in Outcome-Based Commissioning. *Journal of Social Policy*, *47*(1), 57–76. https://doi.org/10.1017/s0047279417000125

Eisenhardt, K. M. (1989). Agency Theory: An Assessment and Review. *Academy of Management Review*, *14*(1), 57–74. https://doi.org/10.5465/amr.1989.4279003

Evans, M. (2013). Meeting the challenge of impact investing: how can contracting practices secure social impact without sacrificing performance? *Journal of Sustainable Finance & Investment*, *3*(2), 138–154. https://doi.org/10.1080/20430795.2013.776260

Fama, E. F. (1980). Agency Problems and the Theory of the Firm. *Journal of Political Economy*, *88*(2), 288–307. https://doi.org/10.1086/260866

Fama, E. F., & Jensen, M. C. (1983). Separation of Ownership and Control. *The Journal of Law and Economics*, *26*(2), 301–325. https://doi.org/10.1086/467037

Fischer, R. L., & Richter, F. G.-C. (2017). SROI in the pay for success context: Are they at odds? *Evaluation and Program Planning*, *64*, 105–109. https://doi.org/10.1016/j.evalprogplan.2016.11.012

Fox, C., & Albertson, K. (2011). Payment by results and social impact bonds in the criminal justice sector: New challenges for the concept of evidence-based policy? *Criminology & Criminal Justice*, *11*(5), 395-413. https://doi.org/10.1177/1748895811415580

Fox, C., & Albertson, K. (2012). Is payment by results the most efficient way to address the challenges faced by the criminal justice sector? *Probation Journal*, *59*(4), 355–373. https://doi.org/10.1177/0264550512458473

Fox, C., & Morris, S. (2019). Evaluating outcome-based payment programmes: challenges for evidence-based policy. *Journal of Economic Policy Reform*, *24*(1), 61-77. https://doi.org/10.1080/17487870.2019.1575217

Fraser, A., Tan, S., Kruithof, K., Sim, M., Disley, E., Giacomantonio, C., Lagarde, M., Rubin, J., & Mays, N. (2018). *Evaluation of the Social Impact Bond Trailblazers in Health and Social Care: Final Report*. London: PIRU. London School of Hygiene & Tropical Medicine and RAND Europe.

Fraser, A., Tan, S., Lagarde, M., & Mays, N. (2018). Narratives of Promise, Narratives of Caution: A Review of the Literature on Social Impact Bonds. *Social Policy & Administration*, *52*(1), 4–28. https://doi.org/10.1111/spol.12260

Gadenne, V., Maglicic, M., Nolan, D., Wright, H., & Frerichs, J. (2020). *The Behavioural Insights Team / Individual Placement Support: A Social Impact Bond Model*. https://golab.bsg.ox.ac.uk/documents/Social_Impact_Bond_IPS_Research_Report_Final.pdf

Giacomantonio, C. (2017). Grant-Maximizing but not Money-Making: A Simple Decision-Tree Analysis for Social Impact Bonds. *Journal of Social Entrepreneurship*, *8*(1), 47–66. https://doi.org/10.1080/19420676.2016.1271348

Gianoncelli, A., & Boiardi, P. (2018). *IMPACT STRATEGIES How Investors Drive Social Impact EVPA REPORT*. European Venture Philanthropy Association. https://www.impacteurope.net/sites/www.evpa.ngo/files/publications/EVPA_Impact_Strategies_2018.pdf

Gianoncelli, A., & Gaggiotti, G. (2021). Managing for Impact: The Role of Investors' Networks in Enhancing Appropriate Impact Measurement and Management Practices. In A. Richards & J. Nicholls (Eds.), *Generation Impact: International Perspectives on Impact Accounting* (pp. 219–233). Emerald Publishing Limited. https://doi.org/10.1108/978-1-78973-929-920200022

Gianoncelli, A., Gaggiotti, G., Boiardi, P., & Martínez, A. P. (2019). *EVPA REPORT 15 YEARS OF IMPACT TAKING STOCK AND LOOKING AHEAD*. European Venture Philanthropy Association. https://www.impacteurope.net/sites/www.evpa.ngo/files/publications/15_Years_of_Impact-Taking_Stock_and_Looking_ahead_2019.pdf

GIIN. 2025a. ***What you need to know about impact investing.*** The Global Impact Investing Network (GIIN). Retrieved April 17, 2025, from https://thegiin.org/publication/post/about-impact-investing/#who-is-making-impact-investments.

GIIN. (2025a). *What you need to know about impact investing*. The Global Impact Investing Network (GIIN). https://thegiin.org/publication/post/about-impact-investing/#who-is-making-impact-investments

GIIN. (2025b). *Welcome to IRIS+ System | the generally accepted system for impact investors to measure, manage, and optimize their impact*. Thegiin.org; The Global Impact Investing Network (GIIN). https://iris.thegiin.org/

Griffiths, R., Thomas, A., & Pemberton, A. (2016). *Qualitative evaluation of the DWP Innovation Fund: Final report*. Department for Work and Pensions. https://assets.publishing.service.gov.uk/media/5a819f76e5274a2e8ab54fc9/rr922-qualitative-evaluation-of-the-dwp-innovation-fund-final-report.pdf

Grünhaus, C., & Rauscher, O. (2021). *Impact und Wirkungsanalyse in Nonprofit Organisationen, Unternehmen und Organisationen mit gesellschaftlichem Mehrwert*.

Kompetenzzentrum für Nonprofit-Organisationen und Social Entrepreneurship. https://www.wu.ac.at/fileadmin/wu/d/cc/npocompetence/12_Publikationen_NPO_SE/Gr%C3%BCnhaus_Rauscher_Impact_Wirkungsanalyse_gesellMehrwert_Apr2021.pdf

Grünhaus, C., & Rauscher, O. (2022). Evaluation und Wirkungsmessung. In M. Meyer, R. Simsa, & C. Badelt (Eds.), *Handbuch der Nonprofit-Organisation Strukturen und Management.* (pp. 507–528). Schäffer-Poeschel.

Hand, D., Ulanow, M., Pan, H., & Xiao, K. (2024). *SIZING THE IMPACT INVESTING MARKET 2024*. The Global Impact Investing Network (GIIN). https://s3.amazonaws.com/giin-web-assets/giin/assets/publication/giin-sizingtheimpactinvestingmarket-2024.pdf

Harris, M., & Raviv, A. (1978). Some results on incentive contracts with applications to education and employment, health insurance, and law enforcement. *The American economic review*, *68*(1), 20-30.

Hevenstone, D., Fraser, A., Lukas Hobi, Wojtek Przepiorka, & Geuke, G. (2023). The impact of social impact bond financing. *Public Administration Review*, *83*(4), 930–946. https://doi.org/10.1111/puar.13631

Hill, C. W. L., & Jones, T. M. (1992). STAKEHOLDER-AGENCY THEORY. *Journal of Management Studies*, *29*(2), 131–154. https://doi.org/10.1111/j.1467-6486.1992.tb00657.x

Höchstädter, A. K., & Scheck, B. (2015). What's in a Name: An Analysis of Impact Investing Understandings by Academics and Practitioners. *Journal of Business Ethics*, *132*(2), 449–475. https://doi.org/10.1007/s10551-014-2327-0

INDIGO. (2025). *Impact Bond Dataset*. The Government Outcomes Lab. https://golab.bsg.ox.ac.uk/knowledge-bank/indigo/impact-bond-dataset-v2/

Jensen, M. C. (1983). Organization theory and methodology. *The Accounting review, 58*(2) 319-339.

Jensen, M. C., & Meckling, W. H. (1976). Theory of the Firm: Managerial behavior, Agency Costs and Ownership Structure. *Journal of Financial Economics*, *3*(4), 305–360. https://doi.org/10.1016/0304-405X(76)90026-X

Lake, R. W. (2015). The financialization of urban policy in the age of Obama. *Journal of Urban Affairs*, *37*(1), 75-78.

Liebman, J. B. (2011). Social Impact Bonds: A promising new financing model to accelerate social innovation and improve government performance. *Center for American Progress*.

Liebman, J., & Sellman, A. (2013). *SOCIAL IMPACT BONDS A Guide for State and Local Governments*. Harvard Kennedy School Social Impact Bond Technical Assistance Lab. https://portal.ct.gov/-/media/DCF/SocialImpactBonds/pdf/socialimpactbondsaguideforstateandlocalgovernments1pdf.pdf

Logue, D., Höllerer, M. A., Clegg, S., Millner, R., & Jebabli, J. (2024). Managing Social Impact Bonds: Intermediary Work and Designing Institutional Infrastructure. *Journal of Management Inquiry*, 1-15. https://doi.org/10.1177/10564926241274328

Maier, F., Barbetta, G. P., & Godina, F. (2018). Paradoxes of Social Impact Bonds. *Social Policy & Administration*, *52*(7), 1332–1353. https://doi.org/10.1111/spol.12343

Maier, F., & Meyer, M. (2017). Social Impact Bonds and the Perils of Aligned Interests. *Administrative Sciences*, *7*(3), 24. https://doi.org/10.3390/admsci7030024

Maier, F., Schober, C., Simsa, R., & Millner, R. (2015). SROI as a Method for Evaluation Research: Understanding Merits and Limitations. *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, *26*(5), 1805–1830. https://doi.org/10.1007/s11266-014-9490-x

Mayring, P., & Brunner, E. (2006). Qualitative Textanalyse - Qualitative Inhaltsanalyse. In V. Flaker & T. Schmid (Eds.), *Von der Idee zur Forschungsarbeit* (pp. 453–462). Böhlau.

Mayring, P., & Fenzl, T. (2019). Qualitative Inhaltsanalyse. In N. Baur & J. Blasius (Eds.), *Handbuch Methoden der empirischen Sozialforschung* (Vol. 2, pp. 633–648). Springer VS. https://doi.org/10.1007/978-3-658-21308-4_42

McHugh, N., Sinclair, S., Roy, M., Huckfield, L., & Donaldson, C. (2013). Social impact bonds: a wolf in sheep's clothing? *Journal of Poverty and Social Justice*, *21*(3), 247–257. https://doi.org/10.1332/204674313x13812372137921

McKay, K. (2013). Debunking the Myths Behind Social Impact Bond Speculation. *Stanford Social Innovation Review*. https://doi.org/10.48558/8Y02-0834

Millar, R., & Hall, K. (2013). Social Return on Investment (SROI) and Performance Measurement. *Public Management Review*, *15*(6), 923–941. https://doi.org/10.1080/14719037.2012.698857

Millner, R., & Meyer, M. (2022). Collaborative governance in Social Impact Bonds: aligning interests within divergent accountabilities? *Public Management Review, 24*(5), 720–742. https://doi.org/10.1080/14719037.2021.2000253

Mitchell, R., & Meacheam, D. (2011). Knowledge worker control: understanding via principal and agency theory. *The Learning Organization, 18*(2), 149–160. https://doi.org/10.1108/09696471111103740

Morley, J. (2019). The ethical status of social impact bonds. *Journal of Economic Policy Reform*, *24*(1), 44–60. https://doi.org/10.1080/17487870.2019.1573681

Mulgan, G., Reeder, N., Aylott, M., & Bo'Sher, L. (2011). Social impact investment: the challenge and opportunity of social impact bonds. *London: The Young Foundation*, 1-38.

Nazari Chamaki, F., Jenkins, G. P., & Hashemi, M. (2018). Social Impact Bonds: Implementation, Evaluation, and Monitoring. *International Journal of Public Administration*, *42*(4), 289–297. https://doi.org/10.1080/01900692.2018.1433206

New Philanthropy Capital. (2010). *Social Return on Investment Position Paper*. https://npproduction.wpenginepowered.com/wp-content/uploads/2018/07/SROI-position-paper.pdf?_gl=1

O'Flynn, P., & Barnett, C. (2017). *Evaluation and Impact Investing: A Review of Methodologies to Assess Social Impact.* Institute of Development Studies. https://opendocs.ids.ac.uk/articles/report/Evaluation_and_Impact_Investing_A_Review_of_Methodologies_to_Assess_Social_Impact/26479189/1/files/48252709.pdf

Ormiston, J., Charlton, K., Donald, M. S., & Seymour, R. G. (2015). Overcoming the Challenges of Impact Investing: Insights from Leading Investors. *Journal of Social Entrepreneurship*, *6*(3), 352–378. https://doi.org/10.1080/19420676.2015.1049285

Parsons, J., Weiss, C., & Wei, Q. (2016). *Impact Evaluation of the Adolescent Behavioral Learning Experience (ABLE) Program*. Vera Institute of Justice. https://s3.eu-west-2.amazonaws.com/golab.prod/documents/PFS_evaluation_report_2016_0se77e1.pdf

Pauly, M., & Swanson, A. (2013). Social Impact Bonds in Nonprofit Health Care: New Product or New Package? *NBER Working paper No. 18991*. https://doi.org/10.3386/w18991

Pequeneza, N. (2019). The Downside of Social Impact Bonds. *Stanford Social Innovation Review*. https://doi.org/10.48558/BGGV-YP16

Rizzello, A., & Kabli, A. (2020). Sustainable Financial Partnerships for the SDGs: The Case of Social Impact Bonds. *Sustainability*, *12*(13), 5362. https://doi.org/10.3390/su12135362

Roche, C. (2000). Impact assessment: Seeing the wood and the trees. *Development in Practice*, *10*(3-4), 543–555. https://doi.org/10.1080/09614520050116703

Ronicle, J., & Smith, K. (2020). *Youth Engagement Fund Evaluation Final Report*. https://assets.publishing.service.gov.uk/media/5ec66044d3bf7f45fa0989d3/YEF_Evaluation_Report_.pdf

Ryan, P. W., & Lyne, I. (2008). Social enterprise and the measurement of social value: methodological issues with the calculation and application of the social return on investment. *Education, Knowledge and Economy*, *2*(3), 223–237. https://doi.org/10.1080/17496890802426253

Salis, S., Wishart, R., & McKay, S. (2018). *Evaluation of the Innovation Fund pilot Quantitative assessment of impact and social return on investment*. Department for Work and Pensions. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/737021/evaluation-of-the-innovation-fund-pilot-quantitative-assessment-of-impact-and-social-return-on-investment.pdf

Sandberg, J., Juravle, C., Hedesström, T. M., & Hamilton, I. (2009). The Heterogeneity of Socially Responsible Investment. *Journal of Business Ethics*, *87*(4), 519–533. https://doi.org/10.1007/s10551-008-9956-0

Santos, F., Pache, A. C., & Birkholz, C. (2015). Making hybrids work: Aligning business models and organizational design for social enterprises. *California management review*, *57*(3), 36-58.

Sinclair, S., McHugh, N., Huckfield, L., Roy, M., & Donaldson, C. (2014). Social impact bonds: Shifting the boundaries of citizenship. In K. Farnsworth & Z. Irving (Eds.), *Social Policy Review 26* (pp. 119-136). Bristol University Press.

Social Finance. (2013). *Guide to Social Impact Bond Development*. Social Finance Ltd. https://www.socialfinance.org.uk/assets/documents/technical-guide-to-developing-social-impact-bonds.pdf

Social Finance. (2016). *Evaluating Impact Bonds - Balancing Evidence and Risk*. Social Finance Ltd. https://www.socialfinance.org.uk/assets/documents/balancing-evidence-and-risk.pdf

Social Finance. (2021). *ESSEX EDGE OF CARE SIB FINAL REPORT CONTENTS*. Social Finance Ltd. https://www.socialfinance.org.uk/assets/documents/Essex-Edge-of-Care-final-report.pdf

Spiess-Knafl, W., & Scheck, B. (2023). Impact Investing. In *Palgrave studies in impact finance*. Springer International Publishing. https://doi.org/10.1007/978-3-031-32183-2

Stanworth, N., & ATQ Consultants. (2023). *Mental Health and Employment Partnership (MHEP)*. https://www.tnlcommunityfund.org.uk/media/research-documents/social-investment/MHEP-InDepth-Review-3rd-report.pdf?mtime=20231201095343&focal=none

Tan, S., Fraser, A., Giacomantonio, C., Kruithof, K., Sim, M., Lagarde, M., Disley, E., Rubin, J., Mays, N., & RAND Europe. (2015). *An evaluation of Social Impact Bonds in Health and Social Care Interim Report*. Policy Innovation Research Unit. https://golab.bsg.ox.ac.uk/documents/100/Trailblazer_SIBs_interim_report_March_2015_for_publication_on_PIRU_siteapril_a_V6EBXNS.pdf

The Bridgespan Group. (2018). *What Is Impact Investing and Why Should You Care?* Bridgespan. https://www.bridgespan.org/insights/what-is-impact-investing

Wang, H., Xiong, W., Wu, G., & Zhu, D. (2018). Public–private partnership in Public Administration discipline: a literature review. *Public Management Review*, *20*(2), 293–316. https://doi.org/10.1080/14719037.2017.1313445

Warner, M. E. (2013). Private finance for public goods: social impact bonds. *Journal of Economic Policy Reform*, *16*(4), 303–319. https://doi.org/10.1080/17487870.2013.835727

Warner, M. E. (2015). Profiting from Public Value? The Case of Social Impact Bonds. In J. M. Bryson, B. C. Crosby, & L. Bloomberg (Eds.), *Creating public value in practice : advancing the common good in a Multi-Sector, Shared-Power, No-One-Wholly-in-Charge world* (pp. 143–160). CRC Press.

Weiss, C. H. (1995). Nothing as Practical as Good Theory : Exploring Theory-Based Evaluation for Comprehensive Community Initiatives for Children and Families. In J. Connel, A. Kubisch, L. Schorr, & C. Weiss (Eds.), *New Approaches to Evaluating Community Initiatives: Concepts, Methods, and Contexts* (pp. 65–92). The Aspen Institute.

Whitworth, A., & Cullen, A. M. (2023). *Evaluation of the WDP IPS Into Work service Final Report June 2023*. https://golab.bsg.ox.ac.uk/documents/Evaluation_of_the_WDP_into_Work_Service.pdf

Williams, J. W. (2021). "Let's Not Have the Perfect Be the Enemy of the Good": Social Impact Bonds, Randomized Controlled Trials, and the Valuation of Social Programs. *Science, Technology, & Human Values*, *48*(1), 91-114. https://doi.org/10.1177/01622439211042083

Wilson, K. E. (2014). Social Investment: New Investment Approaches for Addressing Social and Economic Challenges. *OECD Science, Technology and Industry Policy Paper No. 15*. https://doi.org/10.2139/ssrn.2501247

10.Appendix

Appendix A: Variable overview from the SIB database

Variables from the "Completed SIBs" sheet:

Variable Name	Description	Source
SIB_ID	Unique identifier for each Social Impact Bond (used for linking across sheets).	INDIGO
Name	Official name of the SIB project according to the INDIGO database.	INDIGO
Country	Country where the SIB was implemented.	INDIGO
Bond Type	Bond type (Social Impact Bond or Development Impact Bond)	INDIGO
Stage of Development	Development stage of the SIB project (Contracted, Implemented, or Completed).	INDIGO
Problem	Social issue or challenge the SIB aims to address.	INDIGO
Policy Area	Policy sector of the intervention (e.g., Employment, Health).	INDIGO
Start Date	Date when the SIB project started implementation.	INDIGO
Currency	Currency of the SIB.	INDIGO
Capital raised	Total investment capital secured for delivering the intervention in the original currency.	INDIGO
Target Population	Description of the main beneficiary group the SIB aimed to serve.	INDIGO
Cohort size	Number of individuals actually engaged in the intervention.	INDIGO
Fund Affiliation	Larger fund or initiative associated with the SIB (if any).	INDIGO
Name of Fund	Name of the fund	INDIGO
# of COMs	Number of distinct contractual outcome metrics (COMs) defined in the project.	INDIGO
Payment Structure	Description of how payments were tied to outcomes (all-or-noting, percentage-based, pay-per-outcome)	Manual
Intermediary	Name of the intermediary of the SIB.	INDIGO, Evaluation report
Qualitative Evaluation Methodologies	Methods used for qualitative evaluation of the SIB (e.g., interviews)	INDIGO, Evaluation report
Quantitative Evaluation Methodologies	Methods used for quantitative evaluation of the SIB (e.g., SROI)	INDIGO, Evaluation report
Success	Reported outcome success based on outcome payment (Yes, No, Partially, or Unclear).	Manual
Success Info (# of COMs achieved)	Additional info on number or share of COMs achieved.	Manual
Comment	Additional comments.	Manual

Variables from the “Deep-dive on COMs” sheet:

Variable Name	Description	Source
SIB_ID	Unique identifier for each Social Impact Bond (used for linking across sheets).	INDIGO
Stage of Development	Development stage of the SIB project (Contracted, Implemented, or Completed).	INDIGO
Bond Type	Bond type (Social Impact Bond or Development Impact Bond)	INDIGO
Country	Country where the SIB was implemented.	INDIGO
Policy Area	Policy sector of the intervention (e.g., Employment, Health).	INDIGO
Name of Fund	Name of the fund	INDIGO
COM ID	Unique identifier for each COM.	Manual
COM Group	Overarching group of the COM.	Manual
Description	Narrative explanation of the outcome metric.	INDIGO
Validation Method Group	Overarching group of the validation method.	Manual
Validation Method/Impact Measurement	Method used to validate outcomes related to this metric.	INDIGO, Evaluation report
Currency	Currency of the SIB.	INDIGO
Payment for COM	Amount of money paid for achieving this COM in the original currency.	INDIGO, Evaluation report
Payment Structure	Description of how payments were tied to outcomes (all-or-nothing, percentage-based, pay-per-outcome)	Manual
Outcome Payment Cap	Payment ceiling associated with the COM across all cases or per individual. Only relevant for pay-per-outcome SIBs.	Evaluation report
COM achieved (Yes/No)	Whether the COM was reported as achieved (Yes/No).	Evaluation report
COM achieved (Text)	Comment on the achievement of the COM in text form.	Manual based on evaluation report
Comment	Additional comments.	Manual

Variables from the “Report Overview” sheet:

Variable Name	Description	Source
SIB_ID	Unique identifier for each Social Impact Bond (used for linking across sheets).	INDIGO
Name	Official name of the SIB project according to the INDIGO database.	INDIGO
Country	Country where the SIB was implemented.	INDIGO
Stage of Development	Development stage of the SIB project (Contracted, Implemented, or Completed).	INDIGO
Bond Type	Bond type (Social Impact Bond or Development Impact Bond)	INDIGO
Policy Area	Policy sector of the intervention (e.g., Employment, Health).	INDIGO
Report Type	Type of report (e.g., final report, interim report, factsheet).	Evaluation report
Report Language	Language of the report.	Evaluation report
Link	Link to the report.	Internet website
File Name	File name of the downloaded report.	Manual
Comment	Additional comments.	Manual

Appendix B: Full List of COM Grouping

Note: The definitions of the COMs as per the SIB database were shortened and summarized where possible to reduce the length of the table.

The Qualifications and Credit Framework (QCF) replaced the National Qualifications Framework (NQF).

COM Group	Defined Contractual Outcome Metrics (shortened descriptions and summarized where possible)
Basic Skills	Basic skills
Change in Healthcare Costs	Change in forecast healthcare usage costs
Engagement with Service/Project	Employment Specialist completion of three appointments with the Service User or completed vocational profile
	Individual engages with IPS programme and completes vocational profile
	Successful engagement with IPS services.
	Enrollment onto the DFN-MoveForward programme
Entry into Education or Training	Entry into Education or Training
Entry into Employment	Entry into full-time Employment
	Entry into part-time employment
	Individual gains competitive employment and starts Job
	Individual spends at least one full day (or 4 hours for part-time work) in paid competitive employment
Entry into Volunteering	Entry into volunteering
Language Qualification	Completion of English as a Second Language Course with a minimum of a pass grade
QCF Qualification	QCF-accredited entry-level qualification
	Achievement of a first QCF-accredited entry-level qualification with a literacy and numeracy focus
	First Level 1 NQF/QCF qualification
	First Level 2 NQF/QCF qualification
	First Level 3 NQF/QCF qualification
	First Level 4 NQF/QCF qualification or above
	1st Entry level 1, 2 and/or 3 qualification (incl. Maths and English)
	NVQ level 1 or equivalent (incl. Maths and English if not held)
	NVQ level 2 and above or equivalent or apprenticeship

	Attainment of work skills qualification
School behavior, attendance, and attitude	Improved attitude to school/education (min. 13 weeks; only for participants aged 14/15 at the start of the program)
	Improved attitude to school (min. 13 weeks; only for participants aged 16 at the start of the program)
	Improved attitude to school/education
	Improved attendance at School (min. 13 weeks; only for participants aged 14/15 at the start of the program)
	Improved school attendance (min. 13 weeks)
	Improved behaviour at school (min. 13 weeks; only for participants aged 14/15 at the start of the program)
	Improved school behaviour. (min. 13 weeks)
Sustained Employment	Entry into First Employment (16+ hours per week) with training element (e.g., an Apprenticeship, or work-based learning) (min. 13 weeks)
	Entry into sustained employment – 26 weeks
	13 weeks apprenticeship sustainment
	13 weeks full-time Employment
	26 weeks full-time Employment
	13 weeks part-time Employment
	26 weeks part-time Employment
	Job sustainment (min. 13 weeks at either under 16 hours or over 16 hours)
	Job sustainment (< 16 hours per week)
	Job sustainment (6 weeks)
	Job sustainment (min. 13 weeks at < 16 hours per week)
	Entry Into Work (min. 3 months at > 16 hours per week)
	Job sustainment (> 16 hours per week)
	Job sustainment (6 months)
	Job sustainment (min. 13 weeks at >16 hours per week)
Sustained Volunteering	13 weeks volunteering
	26 weeks volunteering
Wellbeing & Health	Initial well-being assessment
	Second well-being assessment
	Third wellbeing assessment
	Fourth wellbeing assessment

	Fifth wellbeing assessment
	Health improvements
Workplace Activities Undertaken	4+ workplace activities undertaken

Source: Author's elaboration based on INDIGO (2025).

Appendix C: Full List of Validation Method Grouping

Note: The validation method groups shown below do not match perfectly with the column of the SIB database. In the database, there is one row for every COM, but oftentimes more than one associated validation method. In the table below there is one row for every validation method.

Validation Method Group	Validation Method
Confirmation Letter	Confirmation from university/college
	Confirmation letter from employer
	Confirmation letter from employer, supported by evidence
	Confirmation letter from school/teacher/home tutor
Confirmation Letter + Business Plan	Letter from a recognised business start up organisation including a business plan or evidence of trading that is clearly linked to the company and be proportionate to the business
Copy of Attendance Record	Copy of Attendance Record
Copy of Certificate	Copy of Certificate
	Copy of certificate showing achievement and/or course completion
Copy of Registration Form	Copy of registration Form
Exam Results Slips	Exam results slips or notifications from educational establishments
Ex-post Verificaton	Simple Ex-post verification
Monthly Monitoring Data	Monthly monitoring data
Patient Questionnaire	Addictions patient reported outcomes questionnaire
Payslip	1st payslip
Self-Declaration	Self-declarations
Self-Employment Verification Template + Trading Evidence	Self Employment Verification Template must be supported by trading evidence
Timeslip	Time slips

Source: Author's elaboration based on evaluation reports.

Appendix D: List of evaluation reports used for the content analysis

Reference	File Name	Report Type	Associated SIB_ID	Comment	URL
Griffiths et al. (2016)	002_rr922-qualitative-evaluation-of-the-dwp-innovation-fund-final-report	Final	INDIGO-POJ-0138; INDIGO-POJ-0139; INDIGO-POJ-0140; INDIGO-POJ-0141; INDIGO-POJ-0142; INDIGO-POJ-0143; INDIGO-POJ-0144; INDIGO-POJ-0145; INDIGO-POJ-0146; INDIGO-POJ-0147	Qualitative Report from the Innovation Fund Pilot containing 10 SIBs in total	https://assets.publishing.service.gov.uk/media/5a819f76e5274a2e8ab54fc9/rr922-qualitative-evaluation-of-the-dwp-innovation-fund-final-report.pdf
Salis et al. (2018)	006_evaluation-of-the-innovation-fund-pilot-quantitative-assessment-of-impact-and-social-return-on-investment	Final	INDIGO-POJ-0138; INDIGO-POJ-0139; INDIGO-POJ-0140; INDIGO-POJ-0141; INDIGO-POJ-0142; INDIGO-POJ-0143; INDIGO-POJ-0144; INDIGO-POJ-0145; INDIGO-POJ-0146; INDIGO-POJ-0147	Quantitative Report from the Innovation Fund Pilot containing 10 SIBs in total	https://assets.publishing.service.gov.uk/media/5b87ce4440f0b63cacd32743/evaluation-of-the-innovation-fund-pilot-quantitative-assessment-of-impact-and-social-return-on-investment.pdf
Ronicle & Smith (2020)	016_YEF_Evaluation_Report_	Final	INDIGO-POJ-0163; INDIGO-POJ-0164; INDIGO-POJ-0165; INDIGO-POJ-0166	Final evaluation report of the Youth Engagement Fund comprised of four SIBs in total	https://assets.publishing.service.gov.uk/media/5ec66044d3bf7f45fa0989d3/YEF_Evaluation_Report_.pdf
ConnectMore Solutions &	039_FutureImpact_LCF_-_Final_report	Final	INDIGO-POJ-0172		https://golab.bsg.ox.ac.uk/documents/FutureImpact_LCF_-_Final_report.pdf

Richmond Baxter Ltd (2023)					
Whitworth & Cullen (2023)	043_IPS_Evaluation_of_the_WDP_into_Work_Service	Final	INDIGO-POJ-0190		https://golab.bsg.ox.ac.uk/documents/Evaluation_of_the_WDP_into_Work_Service.pdf
Stanworth & ATQ Consultants (2023)	MHEP-InDepth-Review-3rd-report	Final	INDIGO-POJ-0118; INDIGO-POJ-0309; INDIGO-POJ-0316	Final in-depth review of the MHEP 1 SIBs	https://www.tnlcommunityfund.org.uk/media/research-documents/social-investment/MHEP-InDepth-Review-3rd-report.pdf?mtime=20231201095343&focal=none
Gadenne et al. (2020)	019_Social_Impact_Bond_IPS_Research_Report_Final	Interim	INDIGO-POJ-0118; INDIGO-POJ-0309; INDIGO-POJ-0316	Process Evaluation of the MHEP 1 and MHEP 2 SIBs, but used only as a complementary report to the final in-depth review of the MHEP 1 SIBs	https://golab.bsg.ox.ac.uk/documents/Social_Impact_Bond_IPS_Research_Report_Final.pdf

Source: Author's elaboration.

Appendix E: Content Analysis Codebook

First-level Codes	Second-Level Codes	Definition	Decision Rules	Exemplary Quote
SIB Design, Construction & Delivery	SIB Rationale & Framing	Justification and narrative for launching the SIB; how the SIB is framed as a policy or innovation tool	**Included:** Motivations, expected benefits, policy context, framing language. **Excluded:** Technical contract details, payment mechanics, or operational implementation.	"The aim was to use these innovative funding mechanisms to test the effectiveness of intervening early in tackling the complex issues faced by multiply disadvantaged teenagers, which can result in higher levels of unemployment as adults." (Griffiths et al., 2016, p. 26)
	Stakeholder Collaboration	How stakeholders (commissioners, intermediaries, SSPs, investors, etc.) negotiate, align interests, and co-design the SIB, also reflecting collaboration among stakeholders during implementation and learnings	**Included:** Negotiation of roles, governance arrangements, co-design practices, communication structures, collaborative problem-solving, and trust-building. **Excluded:** Technical contract issues.	"The Central Government stakeholders reported that the SIB enabled good collaboration at the design stage, as it enabled departments to coalesce around a shared interest in the same outcomes." (Ronicle & Smith, 2020, p. 54)
	Payment Structure	Covers how payments are designed and triggered in SIBs, including outcome-based payments, block payments, rate cards, staged return mechanisms, and the distribution of financial risk among stakeholders.	**Included:** PbR models, rate cards, block vs. outcome payments, sliding scales, caps, staged returns, and rationale for design choices. **Excluded:** Operational or contextual challenges that are unrelated to the structure of payment mechanisms.	"The payment of these intermediate outcomes made the business proposition more attractive to social investors by enabling cash flow to be generated quickly, acknowledging the fact that employment was likely to be a long way off for many IF participants, so too distant and risky a prospect for many investors." (Griffiths et al., 2016, p. 26)

	COM Definition & Appropriateness	The process and rationale for selecting and defining Contractual Outcome Metrics (COMs), including debates over metric types (also post-implementation).	**Included:** Justifications for metric choice, reflections on achievability, target setting and realism of targets, and revisions based on delivery experience. **Excluded:** Simple listing of COMs without rationale for choice.	"The one way that we think we help young people, that isn't represented on the outcomes, is by that mind-set change … the skills and capabilities internal to young people, that coaches … spend lots of their time working on and are proudest of improving – but of course it's difficult to include in a PbR contract.'" (Griffiths et al., 2016, p. 46)
	Delivery & Adaptations	Covers how interventions were delivered in practice and adjustments made in response to emerging challenges or learning	**Included:** Key delivery features (e.g., mentoring, in-school provision), adaptations to targets, staffing, service focus, and contractual terms; tailoring to local contexts or client needs. **Excluded:** Initial programme design choices not revised during implementation.	"Projects typically responded by recruiting additional schools in which to run provision and by refocusing support and interventions onto young people who were still in compulsory education." (Griffiths et al., 2016, p. 40)
Measurement & Evaluation	Evaluation Methodology	Covers the methodologies used and the methodological robustness of evaluations	**Included:** Use of RCTs or quasi-experiments, attribution, data quality, counterfactual logic. **Excluded:** Stakeholder perceptions about results.	"The comparison group was obtained by implementing the Propensity Score Matching (PSM) methodology, which entailed pairing (or 'matching') each IF participant with the most similar non-participant young person. Similarity was assessed across a range of personal characteristics captured by means of administrative and survey data." (Salis et al., 2018, p. 15)
	Limitations & Constraints	Challenges related to data access, methodological	**Included:** Data gaps, measurement mismatch, limited	"Unfortunately, the second wave of the survey achieved sample sizes

		limitations, attribution issues, and contextual or administrative burdens affecting the robustness of evaluation	sample sizes, time constraints, administrative burden, and attribution uncertainty. **Excluded:** Normative or ethical critique unrelated to practical or methodological issues.	which were too small to provide reliable findings and therefore exploring longer-term outcomes was not possible using information from respondents." (Salis et al., 2018, p. 14)
Perception of Success	Financial/Outcome Achievement	Mentions of measurable, contractually agreed outcomes (e.g. employment, qualifications, education attendance) achieved or missed by participants, and their implications for financial performance (e.g. outcome payments, investor returns)	**Included:** Achievement or failure of targets such as NQF qualifications, job starts, EET status; impact assessments; outcome caps reached or missed; SROI/benefit-cost metrics; return to investors linked to hard outcome achievement.; discussion of financial success **Excluded:** Qualitative success.	"All IF projects successfully operated for the full three years and expected, as a minimum, to repay investments to their social investors. Several generated a clear financial return for investors." (Griffiths et al., 2016, p. 92)
	Qualitative Outcomes	Stakeholder and beneficiary perceptions of qualitative outcomes, wellbeing, or social change. Outcomes that are not directly tied to formal qualifications or job outcomes, but seen as meaningful indicators of progress	**Included:** Increased self-esteem, resilience, emotional regulation, improved school engagement, better family relations, raised aspirations, social inclusion, general well-being, and similar qualitative outcomes. **Excluded:** Isolated quantitative metrics or payment data.	"Improved social interaction and relationships were also a common theme in the research. Interviewees frequently talked about how they were much better at communication, particularly when in situations with strangers." (Ronicle & Smith, 2020, p. 43)
	Contextual & Critical Reflections	Analytical observations that question, nuance, or contextualise the meaning of outcomes and success	**Included:** Reflections on whether recorded outcomes reflect real or perceived success, discussion of how systemic factors mediate impact, tensions	"No projects were allowed to fail and investors went to considerable lengths to support and capacity build providers that were struggling to generate sufficient outcomes for

			between outcome-based metrics and holistic understandings of change, critical insights into scalability and sustainability, and doubts about attribution or counterfactual reasoning in assessing success **Excluded:** Explicit forward-looking recommendations.	financial viability, although the risk of incurring losses was clearly high." (Griffiths et al., 2016, p. 28)
	Unintended Effects & Risks	Negative or unforeseen consequences of programme design, delivery, or incentive structures that may undermine equity, effectiveness, or mission integrity	**Included:** Evidence or discussion of behavior caused by SIB design or payment structure; cherry-picking and parking behaviours, incentives skewing delivery towards easier outcomes, sidelining of the most disadvantaged youth, caseload pressures, diluted support quality, and risks from financial viability concerns. **Excluded:** Technical/data problems without systemic/ethical implications.	"You always have to bear in mind that you're doing this for an outcome, so you're limited in what you can do. So the support you want to give, you have to say no because you don't have time to do that." (Ronicle & Smith, 2020, p. 61)
	Stakeholder Development	Changes in the capacity, systems, understanding, or future readiness of organisations involved in SIBs, resulting from their participation	**Include:** Learning effects, organisational development, improved data/reporting systems, upskilling of providers and commissioners, shifts in attitudes or readiness for future outcomes-based contracting.	"The involvement in the SIB, coupled with the support from the investors, developed the capacity of the VCSEs concerned to deliver future government contracts, including PbR and SIB contracts. Two of the service providers were involved in future SIBs, and others

			Exclude: Learning limited to individual delivery adaptations or internal project management	were applying to be involved." (Ronicle & Smith, 2020, p. 9)
Design Recommendations	Partnership & Process Recommendations	Lessons and suggestions for improving stakeholder collaboration, process management, or implementation partnerships	**Included:** Partnerships with schools and local actors, formal agreements (e.g., SLAs), process flexibility, coordination improvements, and other operational adjustments. **Excluded:** Outcome metric or contract structure tweaks.	"An early lesson learned by providers was therefore the need for flexibility in terms of how they delivered the intervention in schools; if they were too rigid or prescriptive, the school would decline to take part." (Griffiths et al., 2016, p. 63)
	Metric & Contractual Recommendations	Lessons and suggestions for improving outcome metrics and contractual/payment structure	**Included:** Revisions to outcome targets, risk-sharing mechanisms, flexible payment structures, eligibility criteria, and contract design. **Excluded:** Partnership or implementation process improvements.	"There must be a strong link between the intervention provided by the service provider and the outcomes that payments are attached to. The intervention must be the main influencer on those outcomes." (Ronicle & Smith, 2020, p. 39)
Basic Information		Basic background information about the projects	N/A	"Six projects were awarded IF contracts by the DWP in spring 2012 and commenced operation in April of that year. A further four projects were funded from a second round of bidding in late summer 2012, commencing operation in November and December 2012." (Griffiths et al., 2016, p. 23)

Source: Author's elaboration.

Appendix F: Structural Data of Relevant SIBs

SIB_ID	Name	Start Date	Fund Name	Intermediary	Payment Structure	Qual. Evaluation Methods	Quant. Evaluation Methods	# of COMs	# of achieved COMs	Success	Comment
INDIGO-POJ-0118	MHEP Staffordshire (MHEP 1)	2016	Commissioning Better Outcomes Fund (CBO), Social Outcomes Fund (SOF)	Social Finance; MHEP Ltd	Pay-per-outcome	Interviews, case studies, IPS Fidelity reviews	Monthly monitoring data	4	1	No	Project was terminated early and no COMs were achieved --> therefore, no success
INDIGO-POJ-0124	MHEP North London – Camden (MHEP 2)	2018	CBO	Social Finance; MHEP Ltd	Pay-per-outcome	N/A	N/A	4	N/A	N/A	Final report not available
INDIGO-POJ-0138	The Advance Programme	2012	Innovation Fund (Round One)	Advance Personnel Management UK Ltd	Pay-per-outcome	Interviews	Survey (participants + non-participants; Administrative data; Propensity Score Matching (PSM) for identifying comparison groups; Social Return on Investment (SROI)	9	N/A	Yes	
INDIGO-POJ-0139	3SC Capitalise	2012	Innovation Fund (Round Two)	3SC	Pay-per-outcome	Interviews	Survey (participants + non-participants; Administrative data; PSM for identifying comparison groups; SROI	10	N/A	Yes	
INDIGO-POJ-0140	Energise	2012	Innovation Fund (Round Two)	Social Finance	Pay-per-outcome	Interviews	Survey (participants + non-participants; Administrative data; PSM for identifying comparison groups; SROI	10	N/A	Yes	
INDIGO-POJ-0141	Links for Life	2012	Innovation Fund (Round One)	Bridges Outcomes Partnerships	Pay-per-outcome	Interviews	Survey (participants + non-participants; Administrative data; PSM for identifying comparison groups; SROI	9	N/A	Yes	

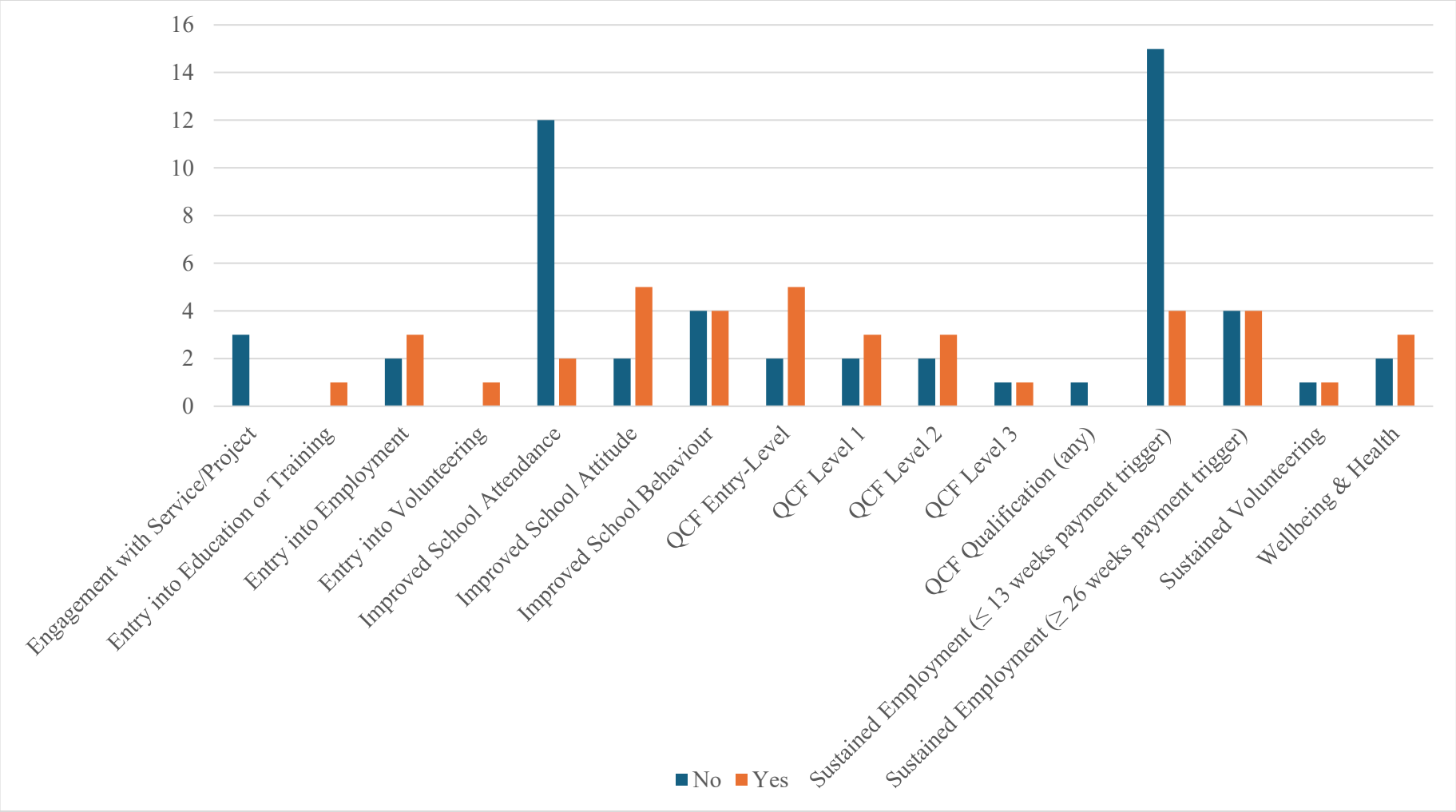
INDIGO-POJ-0142	Living Balance	2012	Innovation Fund (Round One)	Indigo Project Solutions	Pay-per-outcome	Interviews	Survey (participants + non-participants; Administrative data; PSM for identifying comparison groups; SROI	9	N/A	Yes	
INDIGO-POJ-0143	New Horizons (Career Connect)	2012	Innovation Fund (Round One)	Bridges Outcomes Partnerships; Triodos Bank UK	Pay-per-outcome	Interviews	Survey (participants + non-participants; Administrative data; PSM for identifying comparison groups; SROI	9	N/A	Yes	
INDIGO-POJ-0144	Nottingham Futures	2012	Innovation Fund (Round One)	Nottingham City Council	Pay-per-outcome	Interviews	Survey (participants + non-participants; Administrative data; PSM for identifying comparison groups; SROI	9	N/A	Yes	
INDIGO-POJ-0145	Prevista	2012	Innovation Fund (Round Two)	Prevista	Pay-per-outcome	Interviews	Survey (participants + non-participants; Administrative data; PSM for identifying comparison groups; SROI	10	N/A	Yes	
INDIGO-POJ-0146	Teens and Toddlers	2012	Innovation Fund (Round Two)	Social Finance; Bridges Outcomes Partnerships	Pay-per-outcome	Interviews	Survey (participants + non-participants; Administrative data; PSM for identifying comparison groups; SROI	11	N/A	Yes	
INDIGO-POJ-0147	Think Forward (Tomorrow's People)	2012	Innovation Fund (Round One)	Impetus-PEF	Pay-per-outcome	Interviews	Survey (participants + non-participants; Administrative data; PSM for identifying comparison groups; SROI	9	N/A	Yes	
INDIGO-POJ-0163	Futureshapers Sheffield	2015	Youth Engagement Fund	Triodos Bank	Pay-per-outcome	Interviews, focus group, case study visits, Theory of Change	Quantitative evaluation did not take place due to issues with data access and changes in data-sharing regulations	11	7	Yes	Since >50% of COMs were achieved, it is regarded as successful
INDIGO-POJ-0164	Prevista	2015	Youth Engagement Fund	No intermediary	Pay-per-outcome	Interviews, focus group, case study visits, Theory of Change	Quantitative evaluation did not take place due to issues with data access and changes in data-sharing regulations	11	4	No	Since <50% of COMs were achieved, it is regarded as unsuccessful; One COM not applicable as

											no target number has been defined
INDIGO-POJ-0165	Teens and Toddlers	2015	Youth Engagement Fund	Social Finance	Pay-per-outcome	Interviews, focus group, case study visits, Theory of Change	Quantitative evaluation did not take place due to issues with data access and changes in data-sharing regulations	11	4	Yes	Five COMs not applicable as no target numbers have been defined; since >50% of the remaining COMs were achieved, it is regarded as successful
INDIGO-POJ-0166	Unlocking Potential (Career Connect)	2015	Youth Engagement Fund	Triodos Bank	Pay-per-outcome	Interviews, focus group, case study visits, Theory of Change	Quantitative evaluation did not take place due to issues with data access and changes in data-sharing regulations	11	10	Yes	Since >50% of COMs were achieved, it is regarded as successful
INDIGO-POJ-0169	DFN-MoveForward	2018	Life Chances Fund	N/A	N/A	N/A	N/A	4	N/A	N/A	No report available
INDIGO-POJ-0172	FutureImpact	2018	Life Chances Fund	Triodos Bank	Pay-per-outcome	Interviews, focus groups, document reviews, thematic analysis	Self-reported well-being assessments, outcomes database, finance system	19	14	Yes	Since >50% of COMs were achieved, it is regarded as successful
INDIGO-POJ-0176	MHEP Haringey & Barnet (MHEP 3)	2019	Life Chances Fund	Social Finance	Pay-per-outcome	Interviews	Aggregated performance data	4	N/A	N/A	Only interim report available
INDIGO-POJ-0188	MHEP Enfield (MHEP 3)	2020	Life Chances Fund	Social Finance	Pay-per-outcome	Interviews	Aggregated performance data	4	N/A	N/A	Only interim report available
INDIGO-POJ-0189	MHEP Shropshire (MHEP 3)	2020	Life Chances Fund	Social Finance	Pay-per-outcome	Interviews	Aggregated performance data	4	N/A	N/A	Only interim report available

INDIGO-POJ-0190	IPS employment support for people with drug and alcohol addictions	2019	Life Chances Fund	Social Finance	Pay-per-outcome	Interviews, document review, thematic coding of interviews	IPS Fidelity review, data from WDP's service user database, comparison of actual service data to key performance targets, multiple regression analysis, cluster analysis	5	N/A	N/A	Payment structure and information on COMs not specifically mentioned in the report
INDIGO-POJ-0192	MHEP Tower Hamlets Learning Disabilities (MHEP 3)	2020	Life Chances Fund	Social Finance	Pay-per-outcome	Interviews	Aggregated performance data	4	N/A	N/A	Only interim report available
INDIGO-POJ-0193	MHEP Tower Hamlets Mental Health (MHEP 3)	2020	Life Chances Fund	Social Finance	Pay-per-outcome	Interviews	Aggregated performance data	4	N/A	N/A	Only interim report available
INDIGO-POJ-0305	MHEP North London – Barnet (MHEP 2)	2017	CBO, SOF	Social Finance; MHEP Ltd	Pay-per-outcome	N/A	N/A	4	N/A	N/A	Final report not available
INDIGO-POJ-0306	MHEP North London – Enfield (MHEP 2)	2017	CBO	Social Finance; MHEP Ltd	Pay-per-outcome	N/A	N/A	4	N/A	N/A	Final report not available
INDIGO-POJ-0309	MHEP Haringey (MHEP 1)	2017	CBO, SOF	Social Finance; MHEP Ltd	Pay-per-outcome	Interviews, case studies, IPS Fidelity reviews	Monthly monitoring data	4	0	No	Since <50% of COMs were achieved, it is regarded as unsuccessful
INDIGO-POJ-0316	MHEP Tower Hamlets (MHEP 1)	2016	CBO, SOF	Social Finance; MHEP Ltd	Pay-per-outcome	Interviews, case studies, IPS Fidelity reviews	Monthly monitoring data	4	0	No	Since <50% of COMs were achieved, it is regarded as unsuccessful

Source: Author's elaboration based on INDIGO (2025) and evaluation reports.

Appendix G: COM Subgroups and COM Achievement



Source: Author's elaboration.